Artificial Intelligence

NVIDIA Fundamentals of Deep Learning III: Pre-trained Models; Natural Language Processing

1141AI07
MBA, IM, NTPU (M5276) (Fall 2025)
Tue 2, 3, 4 (9:10-12:00) (B3F17)

https://meet.google.com/paj-zhhj-mya

Min-Yuh Day, Ph.D,
Professor and Director

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

2025-11-11

Syllabus

Week	Date	Subject/Topics
1	2025/09/09	Introduction to Artificial Intelligence
2	2025/09/16	Artificial Intelligence and Intelligent Agents; Problem Solving
3	2025/09/23	Knowledge, Reasoning and Knowledge Representation; Uncertain Knowledge and Reasoning
4	2025/09/30	Case Study on Artificial Intelligence I
5	2025/10/07	Machine Learning: Supervised and Unsupervised Learning; The Theory of Learning and Ensemble Learning

Syllabus

Week	Date	Subject/Topics
6	2025/10/14	NVIDIA Fundamentals of Deep Learning I: Deep Learning; Neural Networks
7	2025/10/21	NVIDIA Fundamentals of Deep Learning II: Convolutional Neural Networks; Data Augmentation and Deployment
8	2025/10/28	Self-Learning
9	2025/11/04	Midterm Project Report
10	2025/11/11	NVIDIA Fundamentals of Deep Learning III: Pre-trained Models; Natural Language Processing

Syllabus

Week	Date	Subject/Topics
11	2025/11/18	Case Study on Artificial Intelligence II
12	2025/11/25	Computer Vision and Robotics
13	2025/12/02	Generative AI, Agentic AI, and Physical AI
14	2025/12/09	Philosophy and Ethics of AI and the Future of AI
15	2025/12/16	Final Project Report I
16	2025/12/23	Final Project Report II

NVIDIA Getting Started with Deep Learning

1: An Introduction to Deep Learning

2: The Theory Behind Neural Networks

3: Convolutional Neural Networks

4: Data Augmentation and Deployment

5: Pre-trained Models

6: Natural Language Processing

7: Wrap-up and Assessment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fall 2025

Deep Learning

Spring 2025

Generative AI Innovative Applications

NVIDIA

University Ambassador

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver NVIDIA instructor-led workshop for academia

Greg Estes
Vice President, NVIDIA

Issue Date: : March 7, 2025
Ambassador Certification ID: cCFh1ZWWTvqKTq7dcKkEWw

https://learn.nvidia.com/certificates?id=cCFh1ZWWTvqKTq7dcKkEWw

NVIDIA

Certified Instructor

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver the instructor-led workshop

Building RAG Agents with LLMs

Greg Estes

Vice President, NVIDIA

Issue Date: : March 7, 2025

Certification ID: OVmqY4cSSya0BdMQBWHxzw

https://learn.nvidia.com/certificates?id=OVmqY4cSSya0BdMQBWHxzw

NVIDIA

Certified Instructor

This certificate acknowledges that

Min-Yuh Day

has been certified to deliver the instructor-led workshop

Fundamentals of Deep Learning

Greg Estes

Vice President, NVIDIA

Issue Date: : August 1, 2025

Certification ID: XhLz9YFzRPyew64462TNjQ

https://learn.nvidia.com/certificates?id=ny50l3TXT5W8Q_6kENkIaQ

NVIDIA

Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Getting Started with Deep Learning

Greg Estes

Vice President, NVIDIA

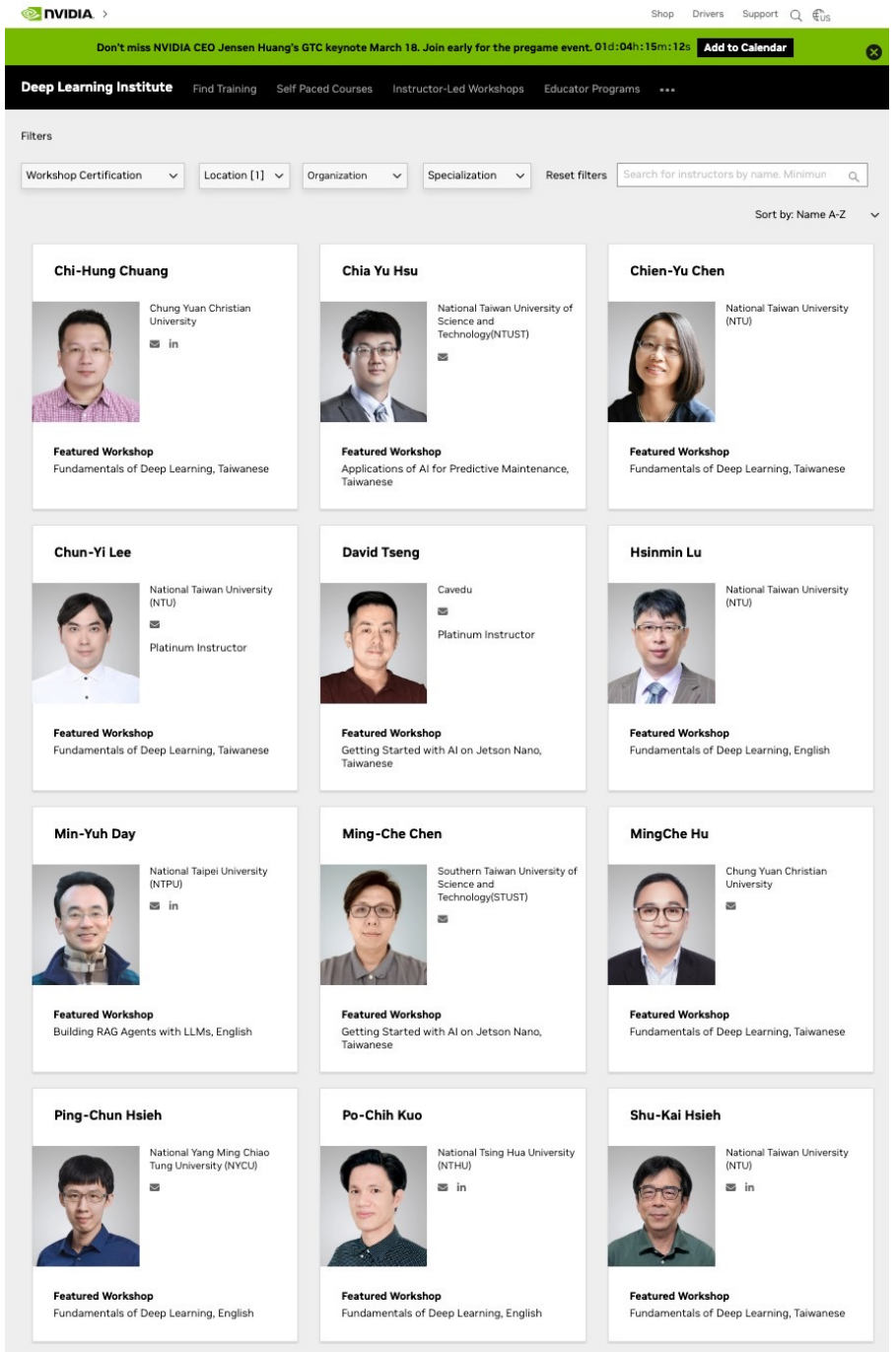
Issue Date: : April 13, 2025

Certification ID: ny50l3TXT5W8Q_6kENklaQ

https://learn.nvidia.com/certificates?id=XhLz9YFzRPyew64462TNjQ

NVIDIA Certified Instructors

(12 from Taiwan)

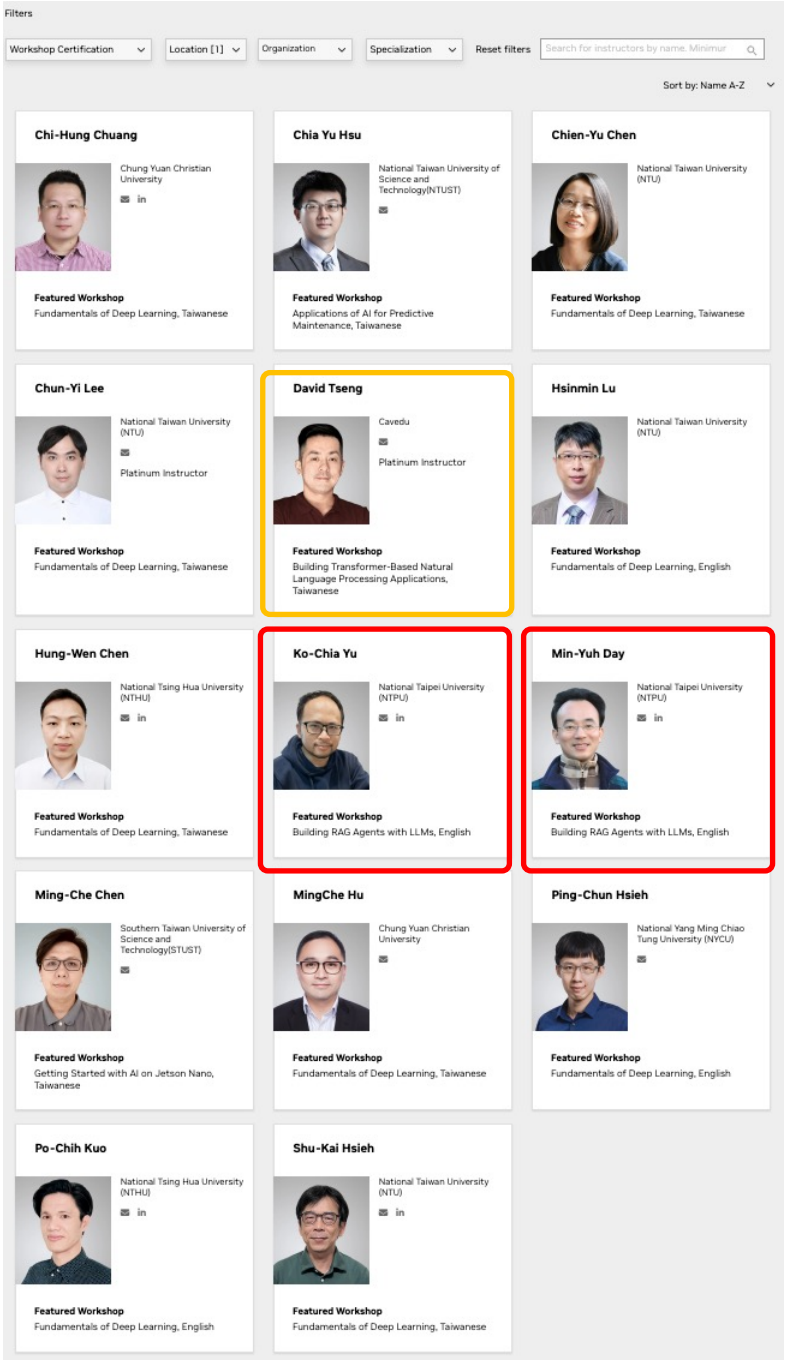


https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Certified Instructors

(14 from Taiwan)

(2 from NTPU)(April 2025)



https://www.nvidia.com/en-us/training/instructor-directory/search/

NVIDIA Developer Program

https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA
Deep Learning Institute (DLI)

https://learn.nvidia.com/

Get NVIDIA DLI Certificate Before the NVIDIA Workshop

- **Step 1. Join NVIDIA Developer Program (Free)**
  https://developer.nvidia.com/join-nvidia-developer-program
- **Step 2. Visit NVIDIA Deep Learning Institute (DLI)**
  https://learn.nvidia.com/
- **Step 3. Enroll "Building RAG Agents with LLMs" Self-Paced Course (Free)**
  https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Get NVIDIA DLI Certificate

- Welcome to NVIDIA DLI "Building RAG Agents with LLMs" workshop.
- Step 1. Visit https://learn.nvidia.com/dli-event
- Step 2. Enter the event code: NTPU_RAG_AMBASSADOR_OC25
- Step 3. Complete the NVIDIA DLI course and review the course datasheet at https://developer.nvidia.com/dli/getready

NVIDIA Deep Learning Institute (DLI)

Building RAG Agents with LLMs workshop

Self-Paced Course	Instructor-Led Workshop
Building RAG Agents With LLMs	**Building RAG Agents With LLMs**
Certificate available Free 8 hours	Certificate available $500 8 hours

Step 1.
Visit
https://learn.nvidia.com/dli-event

Step 2.
Enter the event code:
NTPU_RAG_AMBASSADOR_OC25

https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

Get NVIDIA Certificate

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

NVIDIA

Certificate of Competency

This certificate is awarded to

Min-Yuh Day

for demonstrating competence in the completion of

Building RAG Agents with LLMs

Greg Estes
Vice President, NVIDIA

Issue Date: : December 8, 2024
Certification ID: ed-qOCIMQaatzU8SNUNxgw

Download Certificate

Add to Profile

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)

Generative AI and LLMs | Graphics and Simulation | Accelerated Computing | Data Science | Deep Learning

8 hours

Getting Started With Deep Learning

Explore the fundamentals of deep learning by training neural networks and using results to improve performance and capabilities.

2 hours

Modeling Time-Series Data With Recurrent Neural Networks in Keras

Explore how to classify and forecast time-series data using recurrent neural networks (RNNs), such as modeling a patient's health over time.

4 hours

Deploying a Model for Inference at Production Scale

Learn how to deploy your own machine learning models on a GPU server.

8 hours

Building Real-Time Video AI Applications

Gain the knowledge and skills needed to enable the real-time transformation of raw video data from widely deployed camera sensors into deep learning-based insights.

2 hours

Introduction to Graph Neural Networks

Learn the basic concepts, models, and applications of graph neural networks.

4 hours

Introduction to Physics-Informed Machine Learning With Modulus

Learn the various building blocks of NVIDIA Modulus, which turbocharges use cases by building physics-based deep learning models that are 100,000X faster than traditional methods and offers high-fidelity simulation results.

2 hours

Get Started With Highly Accurate Custom ASR for Speech AI

Learn to build, train, fine-tune, and deploy a GPU-accelerated automatic speech recognition (ASR) service with NVIDIA® Riva that includes customized features.

2 hours

Integrating Sensors With NVIDIA DRIVE

Find out how to integrate automotive sensors into your applications using NVIDIA DRIVE®.

https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA Deep Learning Institute (DLI)

Self-Paced Course

Generative AI Explained

Free
2 hours

Self-Paced Course

Getting Started With Deep Learning

Certificate available
$90
8 hours

Instructor-Led Workshop

Fundamentals of Deep Learning

Certificate available
$500
8 hours

Self-Paced Course

Introduction to Transformer-Based Natural Language Processing

Certificate available
$30
6 hours

Self-Paced Course

Building RAG Agents With LLMs

Certificate available
Free
8 hours

Instructor-Led Workshop

Building RAG Agents With LLMs

Certificate available
$500
8 hours

Self-Paced Course

Generative AI with Diffusion Models

Certificate available
$90
8 hours

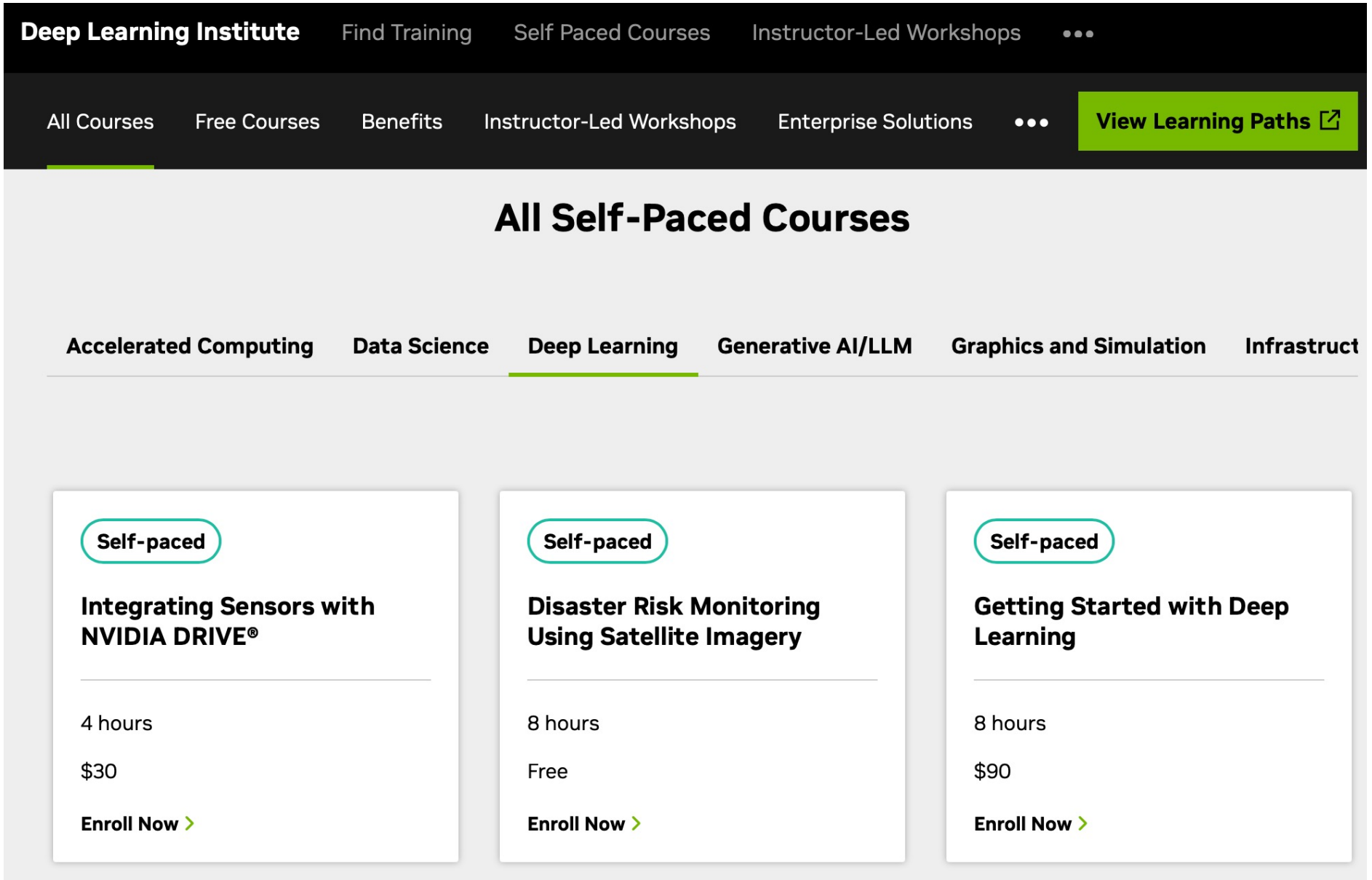
Instructor-Led Workshop

Generative AI with Diffusion Models

Certificate available
$500
8 hours

https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

NVIDIA Getting Started with Deep Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Getting Started with Deep Learning

Learn how deep learning works through hands-on exercises in computer vision and natural language processing.

Online purchases of self-paced courses is supported in select locations worldwide. **Learn more here.**
Free courses are available for enrollment worldwide. Our apologies for the inconvenience.

About Course Objectives Topics Covered Course Outline Stay Informed Contact Us **Continue Learning**

About this Course

Businesses worldwide are using artificial intelligence (AI) to solve their greatest challenges. Healthcare professionals use AI to enable more accurate, faster diagnoses in patients. Retail businesses use it to offer personalized customer shopping experiences. Automakers use it to make personal vehicles, shared mobility, and delivery services safer and more efficient. Deep learning is a powerful AI approach that uses multi-layered artificial neural networks to deliver state-of-the-art accuracy in tasks such as object detection, speech recognition, and language translation. Using deep learning, computers can learn and recognize patterns from data that are considered too complex or subtle for expert-written software.

Course Details

Duration: 08:00

Price: $90

Level: Introductory

Subject: Deep Learning

Language: English

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Deep Learning Institute (DLI)

NVIDIA Products Solutions Industries For You Shop Drivers Support Register Login

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

All Training Course Catalog Stay Informed

What do you want to learn today?

Filters

Level +

Format +

Topics −

- Deep Learning
- Accelerated Computing
- Generative AI/LLM
- Graphics and Simulation
- OpenUSD
- Data Science
- NIMS
- NIM
- RAPIDS

Free / Paid +

Language +

Generative AI

Sort by: --

Showing 19 results

Generative AI ×

Generative AI

All Courses

Self-paced
Generative AI Explained
Free
02:00

Self-paced
Generative AI with Diffusion Models
$90
08:00

Instructor-Led
Generative AI with Diffusion Models
08:00

Self-paced
Augment your LLM Using

Self-paced
Introduction to Transformer-

Instructor-Led
Rapid Application

https://learn.nvidia.com/en-us/training/find-training?q=Generative+AI

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Building RAG Agents with LLMs

Agents powered by large language models (LLMs) have shown great retrieval capability for using tools, looking at documents, and plan their approaches. This course will show you how to deploy an agent system in practice with the flexibility to scale up your system to meet the demands of users and customers.

About Course | Objectives | Topics Covered | Course Outline | Stay Informed | Contact Us | Continue Learning

About this Course

This course is free for a limited time.

The evolution and adoption of large language models (LLMs) have been nothing short of revolutionary, with retrieval-based systems at the forefront of this technological leap. These models are not just tools for automation; they are partners in enhancing productivity, capable of holding informed conversations by interacting with a vast array of tools and documents. This course is designed for those eager to explore the potential of these systems, focusing on practical deployment and the efficient implementation required to manage the considerable demands of both users and deep learning models. As we delve into the intricacies of LLMs, participants will gain insights into advanced orchestration techniques that include internal reasoning, dialog management, and effective tooling strategies.

Course Details

Duration: 08:00

Price: Free

Level: Technical - Intermediate

Subject: Generative AI/LLM

Language: English

Course Prerequisites:
Introductory deep learning knowledge, with comfort

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Generative AI Explained

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Generative AI Explained

In this no-coding course, learn Generative AI concepts and applications, as well as the challenges and opportunities in this exciting field.

About Course | Objectives | Topics Covered | Course Outline | Stay Informed | Contact Us | Continue Learning

About this Course

Generative AI describes technologies that are used to generate new content based on a variety of inputs. In recent time, Generative AI involves the use of neural networks to identify patterns and structures within existing data to generate new content. In this course, you will learn Generative AI concepts, applications, as well as the challenges and opportunities in this exciting field.

Learning Objectives

Upon completion, you will have a basic understanding of Generative AI and be able to more effectively use the various tools built on this

Course Details

Duration: 02:00

Price: Free

Level: Technical - Beginner

Subject: Generative AI/LLM

Language: English

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Introduction to Transformer-Based Natural Language Processing

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Introduction to Transformer-Based Natural Language Processing

Learn how Transformers are used as the building blocks of modern large language models (LLMs). You'll then use these models for various NLP tasks, including text classification, named-entity recognition (NER), author attribution, and question answering.

About Course | Objectives | Topics Covered | Course Outline | Stay Informed | Contact Us | Continue Learning

About this Course

Large Language Models (LLMs), or Transformers, have revolutionized the field of natural language processing (NLP). Driven by recent advancements, applications of NLP and generative AI have exploded in the past decade. With the proliferation of applications like chatbots and intelligent virtual assistants, organizations are infusing their businesses with more interactive human-machine experiences. Understanding how Transformer-based large language models (LLMs) can be used to manipulate, analyze, and generate text-based data is essential. Modern pre-trained LLMs can encapsulate the nuance, context, and sophistication of language, just as humans do. When fine-tuned and deployed correctly, developers can use these LLMs to build powerful NLP applications that provide natural and seamless human-computer interactions within chatbots, AI voice agents, and more. In this course, you'll learn how Transformers are used as the building blocks of modern large language models (LLMs). You'll then use these models for various NLP

Course Details

Duration: 06:00

Price: $30

Level: Technical - Beginner

Subject: Generative AI/LLM

Language: English

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-08+V1

Generative AI with Diffusion Models

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Generative AI with Diffusion Models

Take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines, with applications in creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more.

About Course | Objectives | Topics Covered | Course Outline | Stay Informed | Contact Us | Continue Learning

About this Course

Thanks to improvements in computing power and scientific theory, generative AI is more accessible than ever before. Generative AI plays a significant role across industries due to its numerous applications, such as creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more. In this course, learners will take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines.

Learning Objectives

Course Details

Duration: 08:00

Price: $90

Subject: Generative AI/LLM

Language: English

Course Prerequisites:
A basic understanding of Deep Learning Concepts.

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)

NVIDIA DEVELOPER Home Blog Forums Docs Downloads Training Join

Topics Platforms Industries Resources

Choose a Technical Training Course. It's On Us.

Join the NVIDIA Developer Program and take one of the complimentary technical self-paced courses below (worth up to $90).

Claim Now

Generative AI and LLMs | Graphics and Simulation | Accelerated Computing | Data Science | Deep Learning

6 hours

Introduction to Transformer-Based Natural Language Processing

Learn how transformers are used as the building blocks of modern large language models (LLMs). Then use these models for various natural language processing (NLP) tasks, including text classification, named-entity recognition (NER), author attribution, and question answering.

3 hours

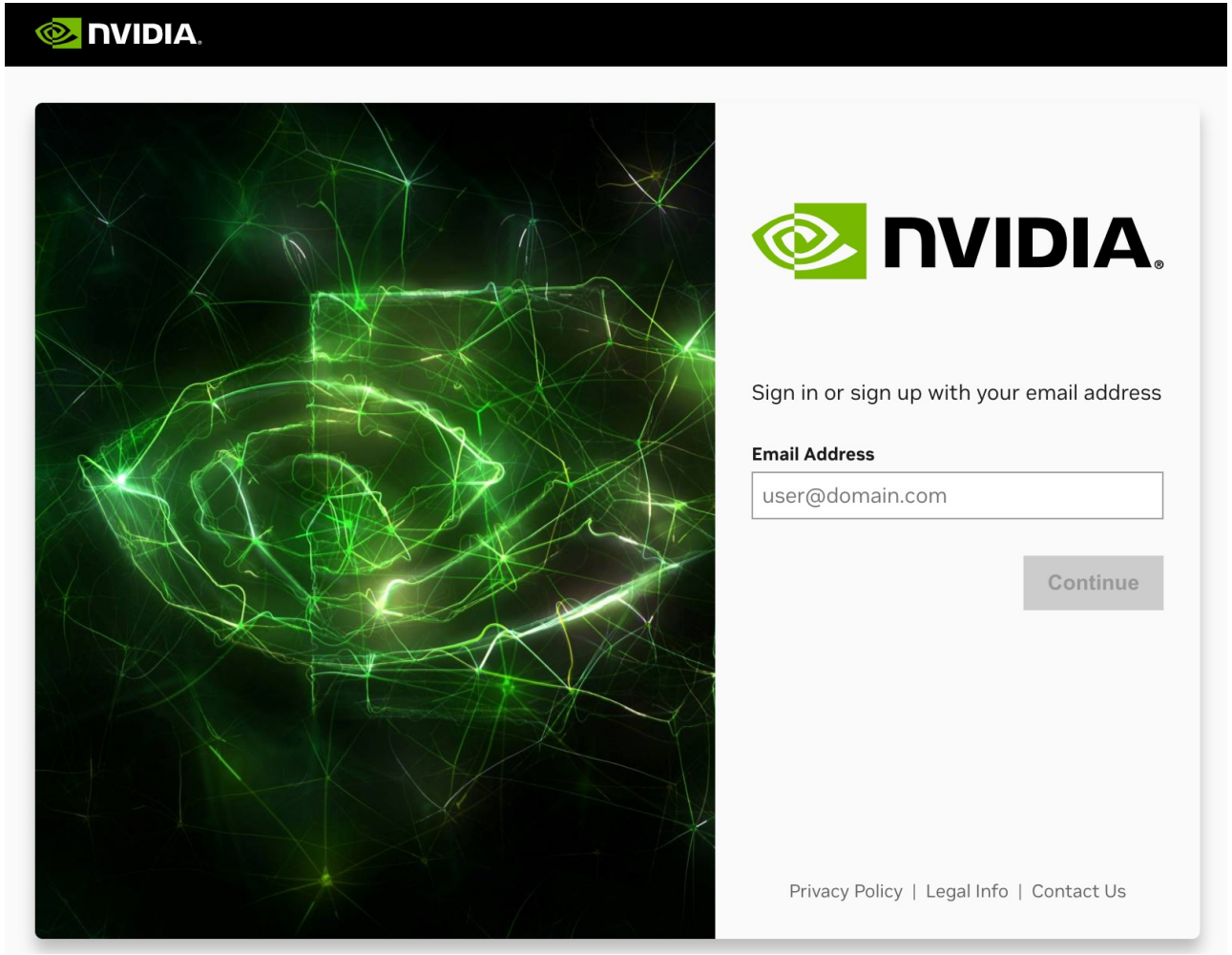
Prompt Engineering With Llama 2

Interact with and prompt engineer Llama 2 models to analyze documents, generate text, and be an AI assistant.

https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

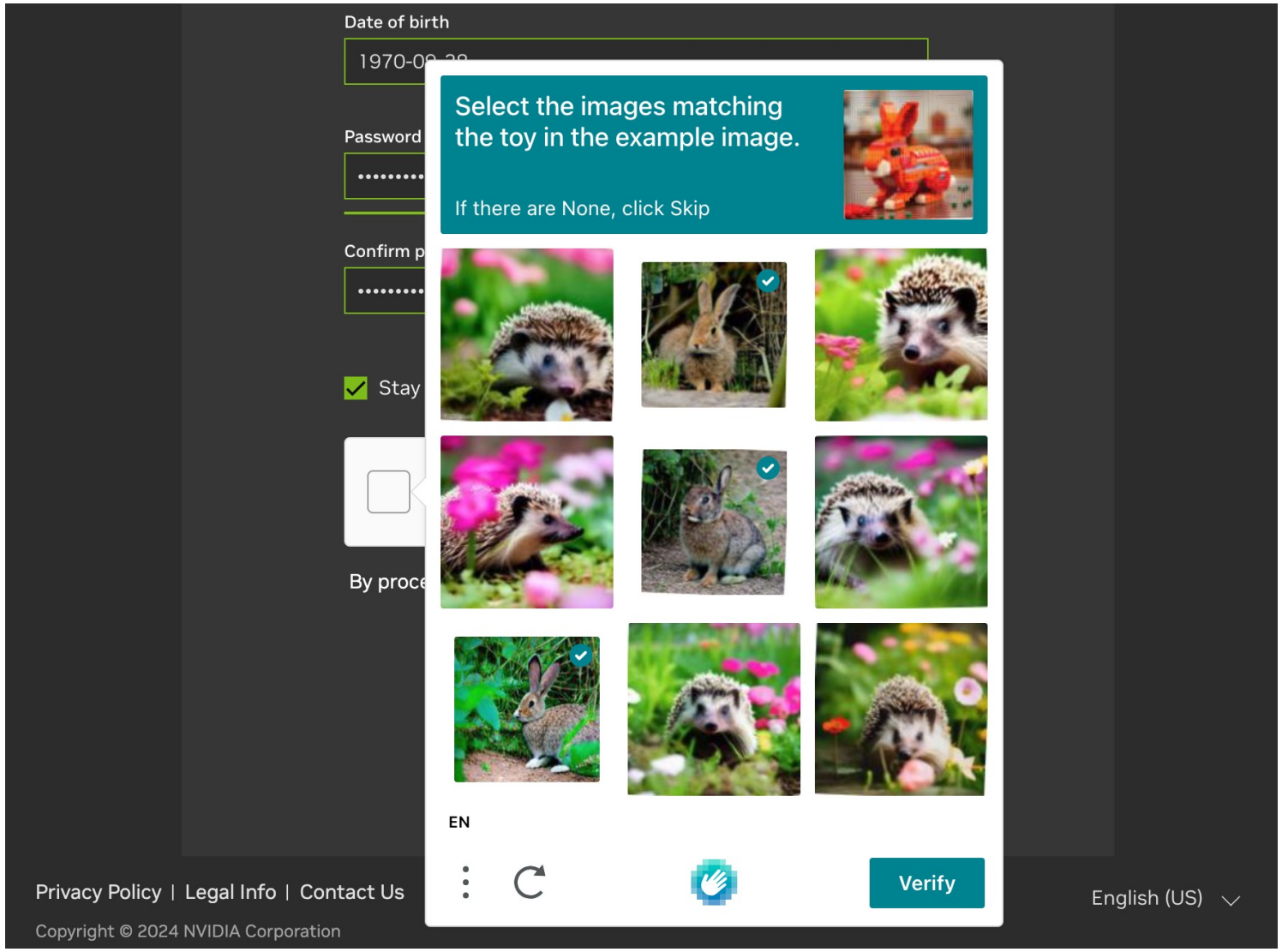
take one of the complimentary technical self-paced courses (worth up to $90)



https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)



https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)

NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

SELECT YOUR FREE COURSE.

Thank you for your participation in the NVIDIA Developer Program. Please select your free DLI course below.

English

- Integrating Sensors with NVIDIA DRIVE®
- Getting Started with Deep Learning
- Deploying a Model for Inference at Production Scale
- Get Started with Highly Accurate Custom ASR for Speech AI
- Introduction to Graph Neural Networks
- Introduction to Transformer-Based Natural Language Processing
- Prompt Engineering with LLaMA-2 (Access Expires Dec. 5th 2025)
- Generative AI with Diffusion Models
- Building Real-Time Video AI Applications
- Introduction to Robotic Simulations in Isaac Sim

Generative AI with Diffusion Models

Take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines, with applications in creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more.

Certificate Available

Duration: 08:00

Continue >

https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)

NVIDIA DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

SELECT YOUR FREE COURSE.

Thank you for your participation in the NVIDIA Developer Program. Please select your free DLI course below.

English

- Modeling Time Series Data with Recurrent Neural Networks in Keras (Access ends 10/16/2024)
- Optimizing CUDA Machine Learning Codes With Nsight Profiling Tools
- Getting Started with Accelerated Computing in CUDA C/C++
- Fundamentals of Accelerated Computing with CUDA Python
- Fundamentals of Accelerated Computing with OpenACC
- Integrating Sensors with NVIDIA DRIVE®
- Getting Started with Deep Learning
- Deploying a Model for Inference at Production Scale
- Get Started with Highly Accurate Custom ASR for Speech AI

Getting Started with Deep Learning

Learn how deep learning works through hands-on exercises in computer vision and natural language processing.

Certificate Available

Duration: 08:00

Continue >

https://developer.nvidia.com/join-nvidia-developer-program

Join the NVIDIA Developer Program

take one of the complimentary technical self-paced courses (worth up to $90)

NVIDIA. DEVELOPER DEEP LEARNING INSTITUTE PROGRAM BENEFITS

SELECT YOUR FREE COURSE.

Thank you for your participation in the NVIDIA Developer Program. Please select your free DLI course below.

English

- Get Started with Highly Accurate Custom ASR for Speech AI
- Introduction to Graph Neural Networks
- Introduction to Transformer-Based Natural Language Processing
- Prompt Engineering with LLaMA-2 (Access Expires Dec. 5th 2025)
- Generative AI with Diffusion Models
- Building Real-Time Video AI Applications
- Introduction to Robotic Simulations in Isaac Sim
- Introduction to Physics-informed Machine Learning with Modulus
- Essentials of USD in Omniverse: Access Expires 09/18/2025
- Synthetic Data Generation for Training Computer Vision Models

Generative AI with Diffusion Models

Take a deeper dive into denoising diffusion models, which are a popular choice for text-to-image pipelines, with applications in creative content generation, data augmentation, simulation and planning, anomaly detection, drug discovery, personalized recommendations, and more.

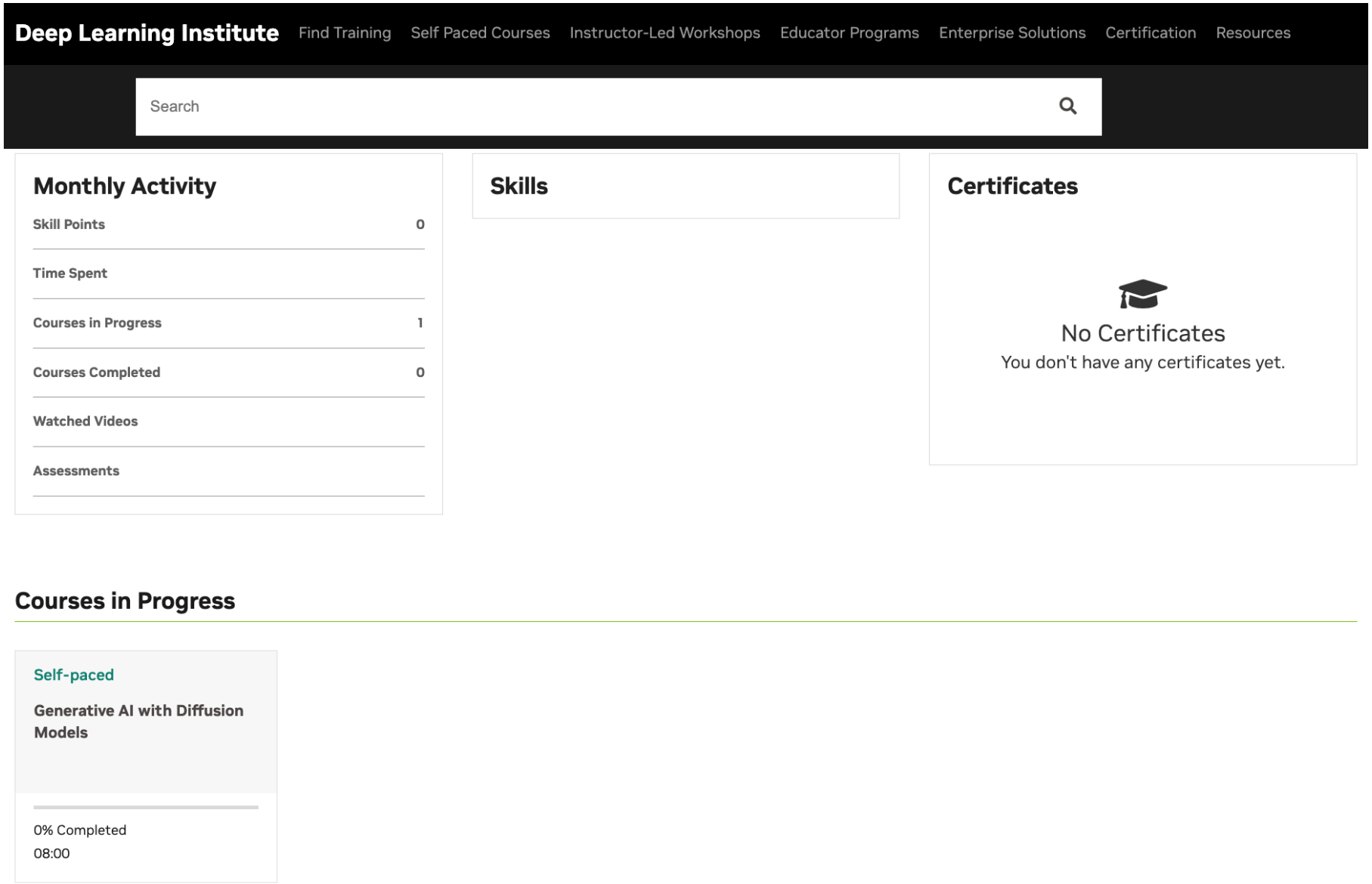
Certificate Available

Duration: 08:00

Continue >

https://developer.nvidia.com/join-nvidia-developer-program

NVIDIA Deep Learning Institute (DLI)



https://learn.nvidia.com/my-learning

NVIDIA Deep Learning Institute (DLI)

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Generative AI with Diffusion Models

Course Progress Bookmarks Updates

Generative AI with Diffusion Models > Start Here > 0: Server Access

- Generative AI with Diffusion Models
- Start Here
- Next Steps
- Feedback

Previous Next

0: Server Access

Bookmark this page

Welcome to Generative AI with Diffusion Models. Please click "Next" below to get started.

Underneath each video is a link to start your own private server for hands-on coding practice. Click the "Start" button to boot up the server. In a few minutes after the server is done loading, click "Launch" to access the code labs.

1: From U-Nets to Diffusion

Bookmark this page

Theory

https://learn.nvidia.com/my-learning

Deep Learning Institute (DLI)

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Search

Monthly Activity

Skill Points	0
Time Spent	
Courses in Progress	16
Courses Completed	12
Watched Videos	
Assessments	

Skills

Certificates

- Introduction to Transformer-Based Natural Language Processing
- Building RAG Agents with LLMs
- Building RAG Agents with LLMs
- Accelerating End-to-End Data Science Workflows
- Generative AI with Diffusion Models
- Building Agentic AI Applications with LLMs

Completed Courses

View more

Self-paced	Self-paced	Self-paced	Self-paced	Self-paced
Sizing LLM Inference Systems	Augment your LLM Using Retrieval Augmented Generation	Building RAG Agents with LLMs	Generative AI Explained	Introduction to Transform Based Natural Language Processing
100% Completed	100% Completed	100% Completed	100% Completed	100% Completed
03:00	01:00	08:00	02:00	06:00

- Introduction to Transformer-Based Natural Language Processing
- Building RAG Agents with LLMs
- Building RAG Agents with LLMs
- Accelerating End-to-End Data Science Workflows
- Generative AI with Diffusion Models
- Building Agentic AI Applications with LLMs
- Building Agentic AI Applications with LLMs
- Rapid Application Development with Large Language Models (LLMs)
- Getting Started with Deep Learning
- Generative AI with Diffusion Models
- Building LLM Applications With Prompt Engineering
- 基於擴散模型的生成式人工智慧
- Fundamentals of Deep Learning
- Domain-Adaptive Pre-Training: Tailoring LLMs for Specialized Applications

https://learn.nvidia.com/my-learning

NVIDIA

Certificate of Completion

This certificate is awarded to

Min-Yuh Day

for successfully completing

Building RAG Agents with LLMs

Greg Estes

Vice President, NVIDIA

Issue Date: : December 8, 2024

Certification ID: ed-qOCIMQaatzU8SNUNxgw |

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw/courses/course?course_id=course-v1:DLI+S-FX-15+V1

https://learn.nvidia.com/certificates?id=ed-qOCIMQaatzU8SNUNxgw

NVIDIA Deep Learning Institute (DLI)

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-Paced Courses All Courses Free Courses Benefits Partners Resources

All Self-Paced Courses

Accelerated Computing | Data Science | Deep Learning | Generative AI/LLM | Graphics and Simulation | Infrastructure

Share Generative AI/LLM Courses

Self-paced Generative AI Explained Free 02:00	Self-paced Introduction to NVIDIA NIM™ Microservices Free 02:00	Self-paced Introduction to Deploying RAG Pipelines for Production at Scale $90 03:00	Self-paced Generative AI with Diffusion Models $90 08:00
Self-paced Techniques for Improving the Effectiveness of RAG Systems $30 03:00	Self-paced Introduction to Transformer-Based Natural Language Processing $30 06:00	Self-paced Building LLM Applications With Prompt Engineering $90 08:00	Self-paced Synthetic Tabular Data Generation Using Transformers $30 04:00
Self-paced Sizing LLM Inference Systems Free 03:00	Self-paced Building RAG Agents with LLMs Free 08:00	Self-paced Augment your LLM Using Retrieval Augmented Generation Free 01:00	

https://learn.nvidia.com/en-us/training/self-paced-courses

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Building RAG Agents with LLMs

Agents powered by large language models (LLMs) have shown great retrieval capability for using tools, looking at documents, and plan their approaches. This course will show you how to deploy an agent system in practice with the flexibility to scale up your system to meet the demands of users and customers.

About Course | Objectives | Topics Covered | Course Outline | Stay Informed | Contact Us | Continue Learning

About this Course

This course is free for a limited time.

The evolution and adoption of large language models (LLMs) have been nothing short of revolutionary, with retrieval-based systems at the forefront of this technological leap. These models are not just tools for automation; they are partners in enhancing productivity, capable of holding informed conversations by interacting with a vast array of tools and documents. This course is designed for those eager to explore the potential of these systems, focusing on practical deployment and the efficient implementation required to manage the considerable demands of both users and deep learning models. As we delve into the intricacies of LLMs, participants will gain insights into advanced orchestration techniques that include internal reasoning, dialog management, and effective tooling strategies.

Course Details

Duration: 08:00

Price: Free

Level: Technical - Intermediate

Subject: Generative AI/LLM

Language: English

Course Prerequisites:
Introductory deep learning knowledge, with comfort

https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

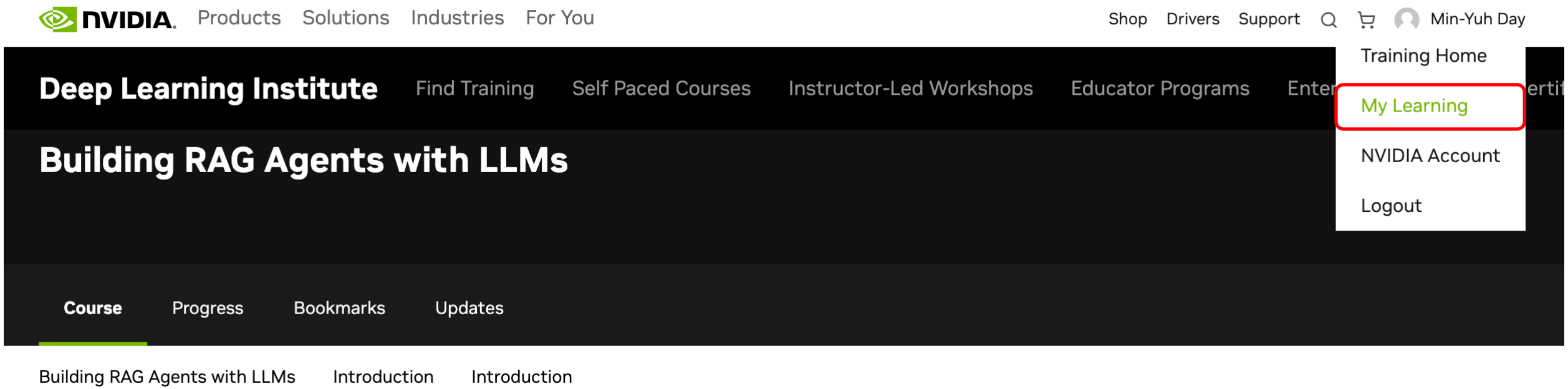
Building RAG Agents with LLMs Introduction Introduction

- Building RAG Agents with LLMs
  - Introduction
  - Environment and LLMs
  - LangChain
  - Documents and Embeddings
  - Retrieval-Augmented Generation
  - Next Steps
  - Feedback

Previous Next

Source: NVIDIA DLI (2024), Building RAG Agents with LLMs, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs



https://learn.nvidia.com/my-learning

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Environment and LLMs Environment [0, 1, 2]

- Building RAG Agents with LLMs
  - Introduction
  - Environment and LLMs
    - Environment [0, 1, 2]
    - Part 1: Course Environment
    - Part 2: LLM Services
  - LangChain
    - Environment [3, 4]
    - Part 3: LangChain
    - Part 4: Running States
  - Documents and Embeddings
    - Environment [5, 6]

Previous Next

Bookmark this page

Welcome to **Building RAG Agents with LLMs**. In this first section, we will get introduced to the overall course environment, LLM services, and recommended workflows!

This tab contains the course environment for this section, which will contain the notebooks for the next two videos! Please click through the videos in the remaining tabs to watch the material and work through the exercises!

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.

NVIDIA DEEP LEARNING INSTITUTE

START

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Environment and LLMs Environment [0, 1, 2]

- Building RAG Agents with LLMs
  - Introduction
    - Introduction
    - Course Slides
  - Environment and LLMs
    - Environment [0, 1, 2]
    - Part 1: Course Environment
    - Part 2: LLM Services
  - LangChain
    - Environment [3, 4]
    - Part 3: LangChain

Previous Next

Welcome to **Building RAG Agents with LLMs**. In this first section, we will get introduced to the overall course environment, LLM services, and recommended workflows!

This tab contains the course environment for this section, which will contain the notebooks for the next two videos! Please click through the videos in the remaining tabs to watch the material and work through the exercises!

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.

NVIDIA DEEP LEARNING INSTITUTE

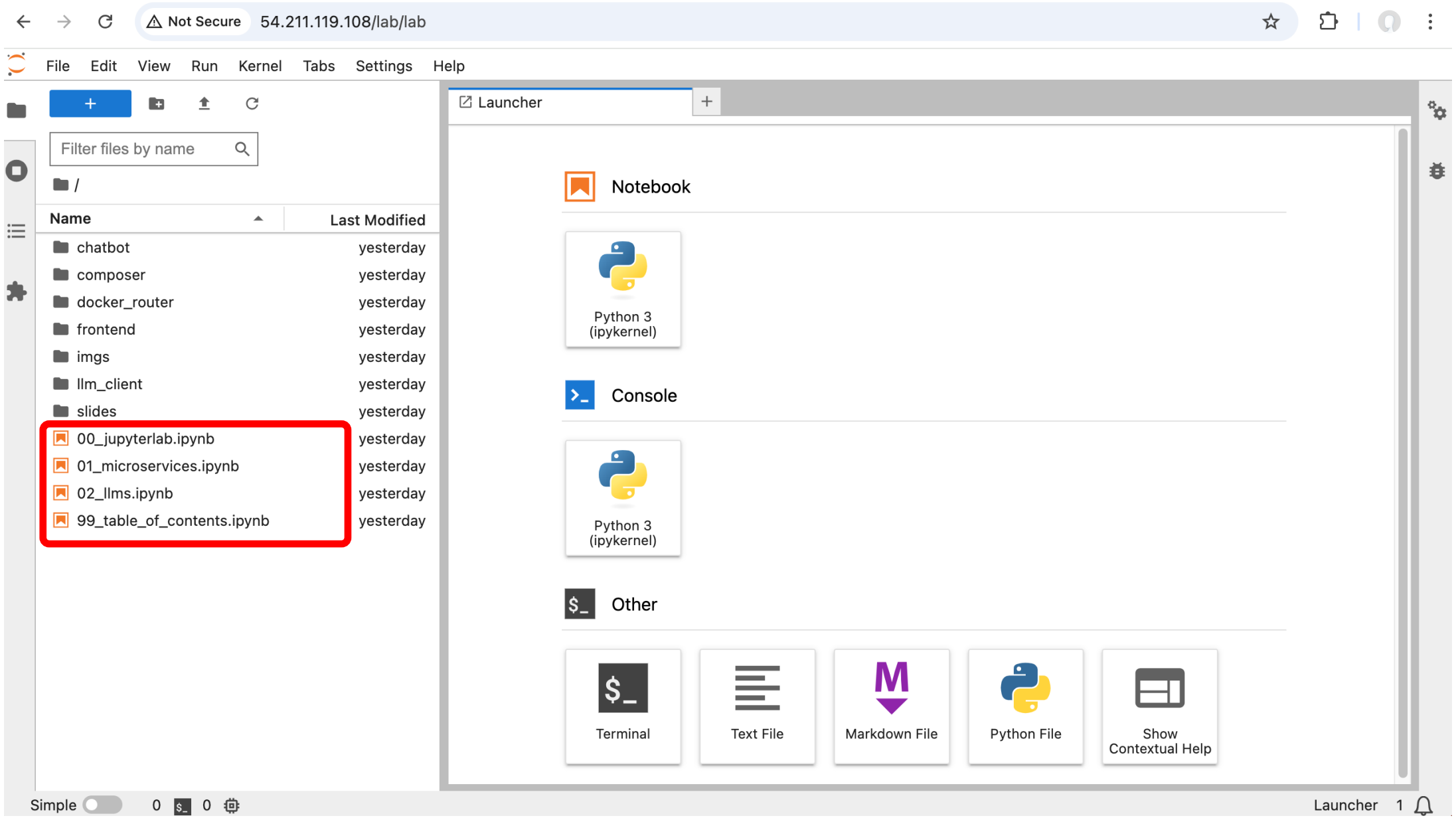
This Lab 0 : 01 : 06 / 2 : 00 : 00

Course 13 : 45 : 51 / 32 : 00 : 00

LAUNCH STOP TASK

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs



https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs

Not Secure 54.211.119.108/lab/lab/tree/00_jupyterlab.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name	Last Modified
chatbot	yesterday
composer	yesterday
docker_router	yesterday
frontend	yesterday
imgs	yesterday
llm_client	yesterday
slides	yesterday
00_jupyterlab.ipynb	yesterday
01_microservices.ipynb	yesterday
02_llms.ipynb	yesterday
99_table_of_contents.ipynb	yesterday

Launcher | 00_jupyterlab.ipynb | 01_microservices.ipynb | 02_llms.ipynb

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

Notebook 0: JupyterLab

We use JupyterLab to manage our environment for this hands-on lab. The JupyterLab Interface is a dashboard that provides access to interactive iPython notebooks, as well as the folder structure of our environment and a terminal window into the Ubuntu operating system. The view includes a **menu bar** at the top, a **file browser** in the **left sidebar**, and a **main work area** initially open to the "Launcher" page.

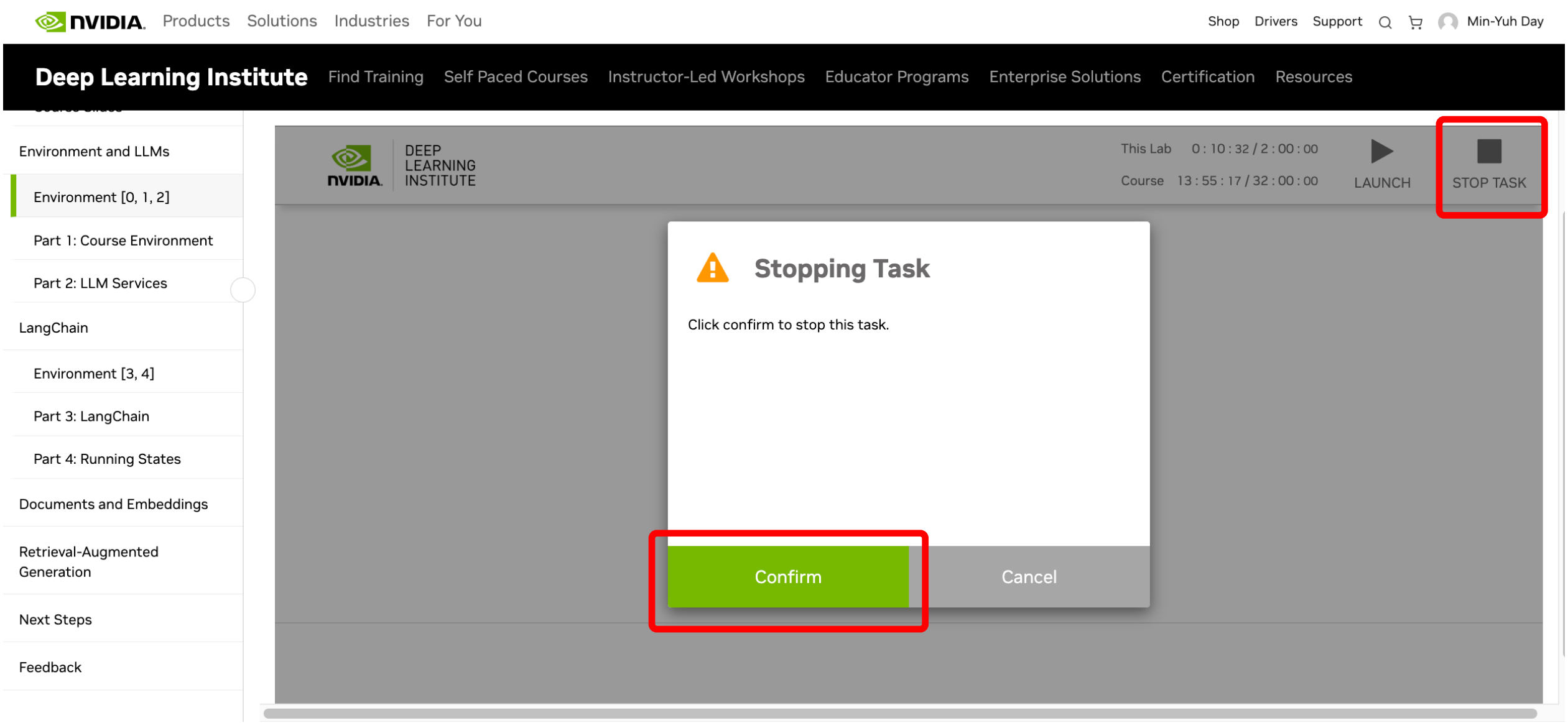
>

Menu bar

Simple 0 3 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 00_jupyterlab.ipynb 1

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs



https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Retrieval-Augmented Generation Environment [7, 8, Assessment]

Previous Next

Bookmark this page

In this section, we will combine all of our prior efforts to integrate and evaluate retrieval-augmented generation pipelines! Along the way, you will also get the opportunity to work through the assessment, which will involve Gradio, LangServe, FAISS, RAG, and Evaluation! ***Good Luck!***

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.

NVIDIA DEEP LEARNING INSTITUTE

This Lab 0 : 15 : 39 / 4 : 00 : 00
Course 14 : 12 : 18 / 32 : 00 : 00
LAUNCH STOP TASK ASSESS TASK

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

Building RAG Agents with LLMs

Not Secure 34.227.20.149/lab/lab/tree/08_evaluation.ipynb

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name	Last Modified
chatbot	yesterday
composer	yesterday
docker_router	yesterday
frontend	yesterday
imgs	yesterday
llm_client	yesterday
slides	yesterday
solutions	yesterday
00_jupyterlab.ipynb	yesterday
01_microservices.ipynb	yesterday
02_llms.ipynb	yesterday
03_langchain_intro.ipynb	yesterday
04_running_state.ipynb	yesterday
05_documents.ipynb	yesterday
06_embeddings.ipynb	yesterday
07_vectorstores.ipynb	yesterday
08_evaluation.ipynb	yesterday
09_langserve.ipynb	yesterday
64_guardrails.ipynb	yesterday
99_table_of_contents.ipynb	yesterday

Simple 0 12 Python 3 (ipykernel) | Idle

Markdown Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

Notebook 8 [Assessment]: RAG Evaluation

Welcome to the last notebook of the course! In the previous notebook, you integrated a vector store solution into a RAG pipeline! In this notebook, you will take that same pipeline and evaluate it using numerical RAG evaluation techniques incorporating LLM-as-a-Judge metrics!

Learning Objectives:

- Learn how to integrate the techniques from prior notebooks to numerically approximate the goodness of your RAG pipeline.

Mode: Command Ln 1, Col 1 08_evaluation.ipynb 1

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+S-FX-15+V1

NVIDIA
Fundamentals of Deep Learning

NVIDIA Deep Learning Institute (DLI)

Fundamentals of Deep Learning Workshop

Self-Paced Course

Getting Started With Deep Learning

Certificate available
$90
8 hours

Instructor-Led Workshop

Fundamentals of Deep Learning

Certificate available
$500
8 hours

Step 1.
Visit
https://learn.nvidia.com/dli-event

Step 2.
Enter the event code:
NTPU_FDL_AMBASSADOR_DE25

https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

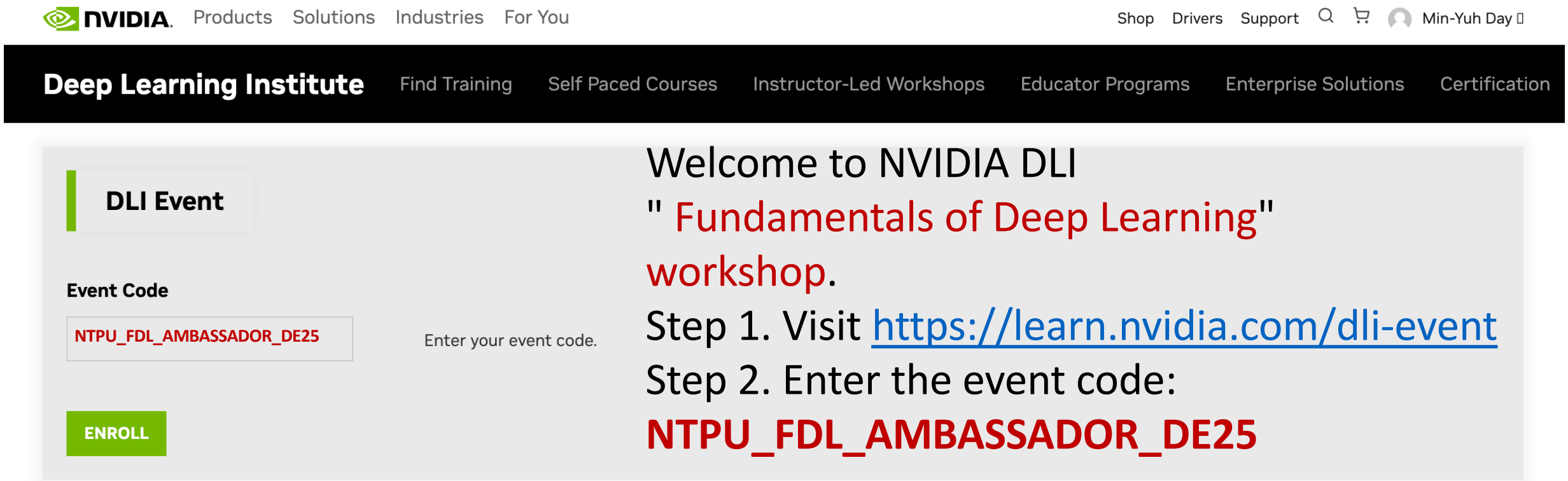
Get NVIDIA DLI Certificate

Fundamentals of Deep Learning

- Welcome to NVIDIA DLI "Fundamentals of Deep Learning" workshop.
- Step 1. Visit https://learn.nvidia.com/dli-event
- Step 2. Enter the event code: NTPU_FDL_AMBASSADOR_DE25
- Step 3. Complete the NVIDIA DLI course and review the course datasheet at https://developer.nvidia.com/dli/getready

Get NVIDIA DLI Certificate

Fundamentals of Deep Learning Workshop



Welcome to NVIDIA DLI

" Fundamentals of Deep Learning" workshop.

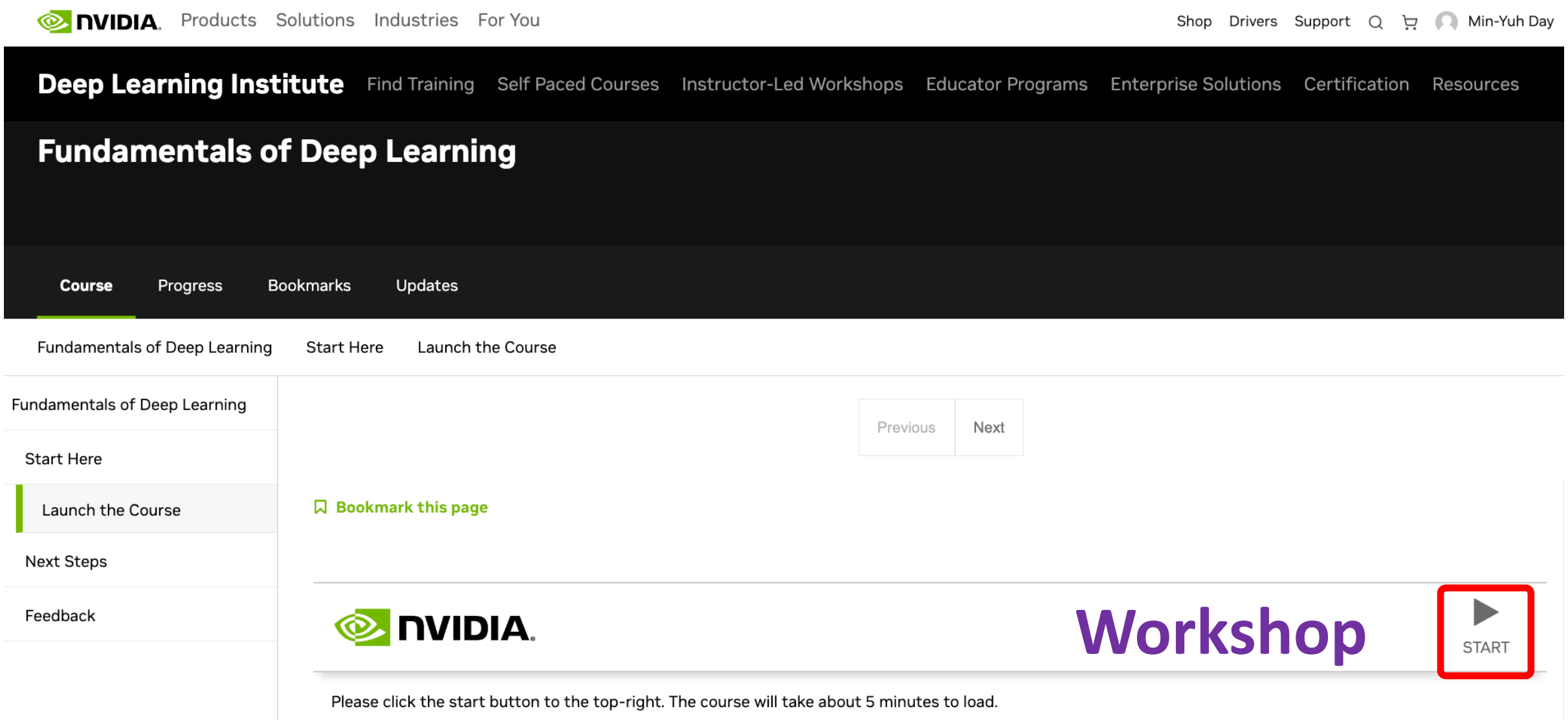
Step 1. Visit https://learn.nvidia.com/dli-event

Step 2. Enter the event code:

NTPU_FDL_AMBASSADOR_DE25

https://learn.nvidia.com/dli-event

NVIDIA Fundamentals of Deep Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+C-FX-01+V3

NVIDIA Deep Learning Institute (DLI)

Building RAG Agents with LLMs Workshop

Self-Paced Course	Instructor-Led Workshop
Building RAG Agents With LLMs	**Building RAG Agents With LLMs**
Certificate available Free 8 hours	Certificate available $500 8 hours

Step 1.
Visit
https://learn.nvidia.com/dli-event

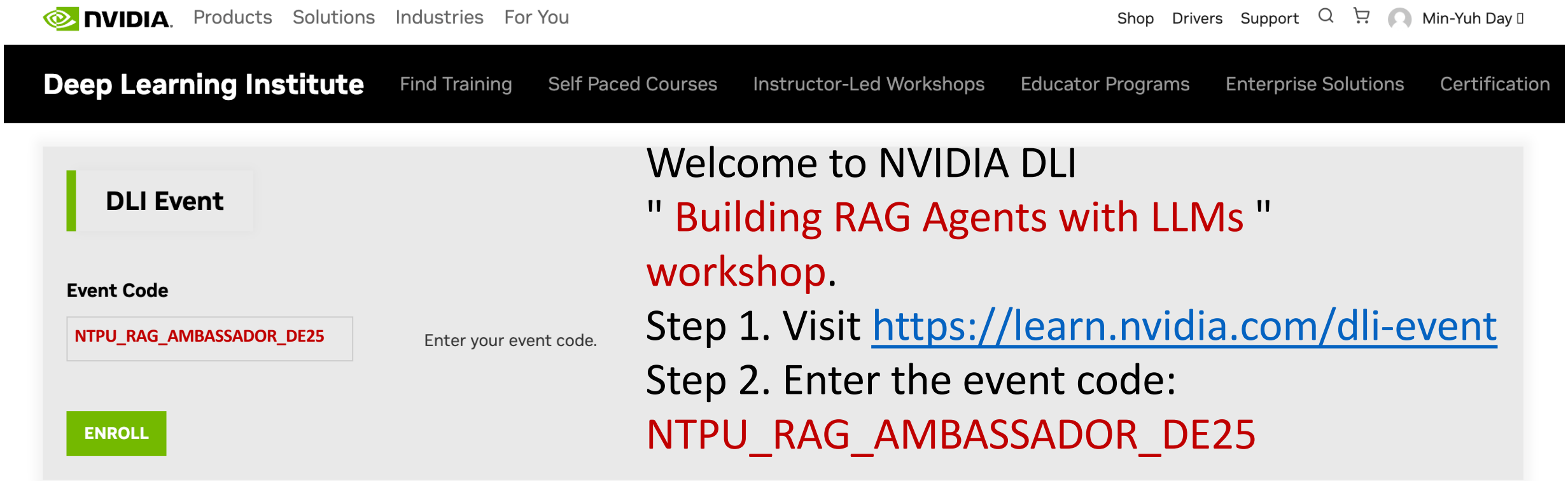
Step 2.
Enter the event code:
NTPU_RAG_AMBASSADOR_DE25

https://www.nvidia.com/en-us/learn/learning-path/generative-ai-llm/

Get NVIDIA DLI Certificate

Building RAG Agents with LLMs

Workshop



Welcome to NVIDIA DLI

" Building RAG Agents with LLMs " workshop.

Step 1. Visit https://learn.nvidia.com/dli-event

Step 2. Enter the event code:

NTPU_RAG_AMBASSADOR_DE25

https://learn.nvidia.com/dli-event

Building RAG Agents with LLMs

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Building RAG Agents with LLMs

Course Progress Bookmarks Updates

Building RAG Agents with LLMs Environment and LLMs Environment [0, 1, 2]

- Building RAG Agents with LLMs
  - Course Slides
    - Course Slides
  - Environment and LLMs
    - Environment [0, 1, 2]
  - LangChain
  - Documents and Embeddings
  - Retrieval-Augmented Generation
  - Next Steps

Previous Next

Welcome to **Building RAG Agents with LLMs**. In this first section, we will get introduced to the overall course environment, LLM services, and recommended workflows!

This tab contains the course environment for this section!

Please click the "Start" button to start up your own private server for hands-on coding practice. It will take a few minutes to start up, so go ahead and click it now and then proceed to the next video! After a few minutes when the server has loaded, click "Launch" to access the code labs.

NVIDIA **Workshop** START

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+C-FX-15+V1

NVIDIA Getting Started with Deep Learning

- **Learn how deep learning works through hands-on exercises in computer vision and natural language processing.**
- **Learning Objectives**
  - **Learn the fundamental techniques and tools required to train a deep learning model**
  - **Gain experience with common deep learning data types and model architectures**
  - **Enhance datasets through data augmentation to improve model accuracy**
  - **Leverage transfer learning between models to achieve efficient results with less data and computation**
  - **Build confidence to take on your own project with a modern deep learning framework**

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

1: An Introduction to Deep Learning

2: The Theory Behind Neural Networks

3: Convolutional Neural Networks

4: Data Augmentation and Deployment

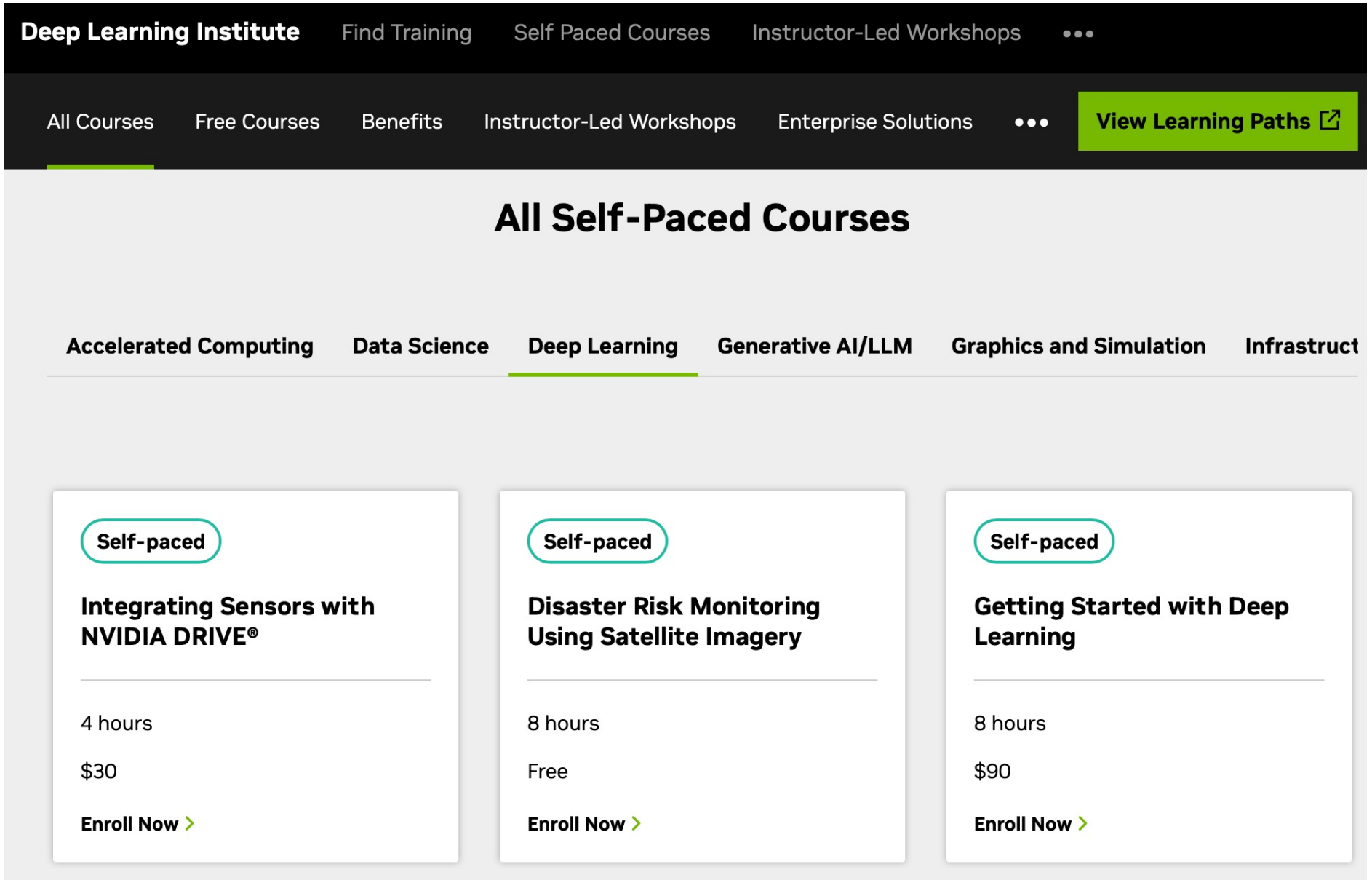
5: Pre-trained Models

6: Natural Language Processing

7: Wrap-up and Assessment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Self-paced Course

Getting Started with Deep Learning

Learn how deep learning works through hands-on exercises in computer vision and natural language processing.

Online purchases of self-paced courses is supported in select locations worldwide. **Learn more here.**
Free courses are available for enrollment worldwide. Our apologies for the inconvenience.

About Course Objectives Topics Covered Course Outline Stay Informed Contact Us **Continue Learning**

About this Course

Businesses worldwide are using artificial intelligence (AI) to solve their greatest challenges. Healthcare professionals use AI to enable more accurate, faster diagnoses in patients. Retail businesses use it to offer personalized customer shopping experiences. Automakers use it to make personal vehicles, shared mobility, and delivery services safer and more efficient. Deep learning is a powerful AI approach that uses multi-layered artificial neural networks to deliver state-of-the-art accuracy in tasks such as object detection, speech recognition, and language translation. Using deep learning, computers can learn and recognize patterns from data that are considered too complex or subtle for expert-written software.

Course Details

Duration: 08:00

Price: $90

Level: Introductory

Subject: Deep Learning

Language: English

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Getting Started with Deep Learning

Course Progress Bookmarks Updates

Getting Started with Deep Learning Start Here 1: An Introduction to Deep Learning

Getting Started with Deep Learning

Start Here

1: An Introduction to Deep Learning

2: The Theory Behind Neural Networks

3: Convolutional Neural Networks

4: Data Augmentation and Deployment

5: Pre-trained Models

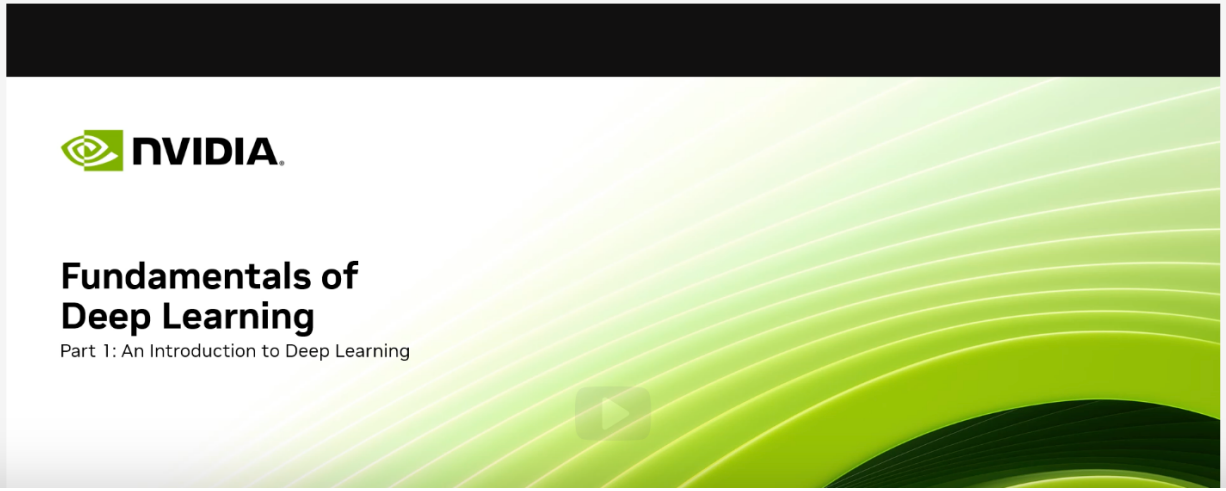
6: Natural Language Processing

7: Wrap-up and Assessment

Previous Next

Bookmark this page

Theory



Studying AI could be pretty challenging, but it's also very rewarding.

Not only can

we learn about how robots think, but we can also pick up the few tricks to help humans be more efficient learners too.

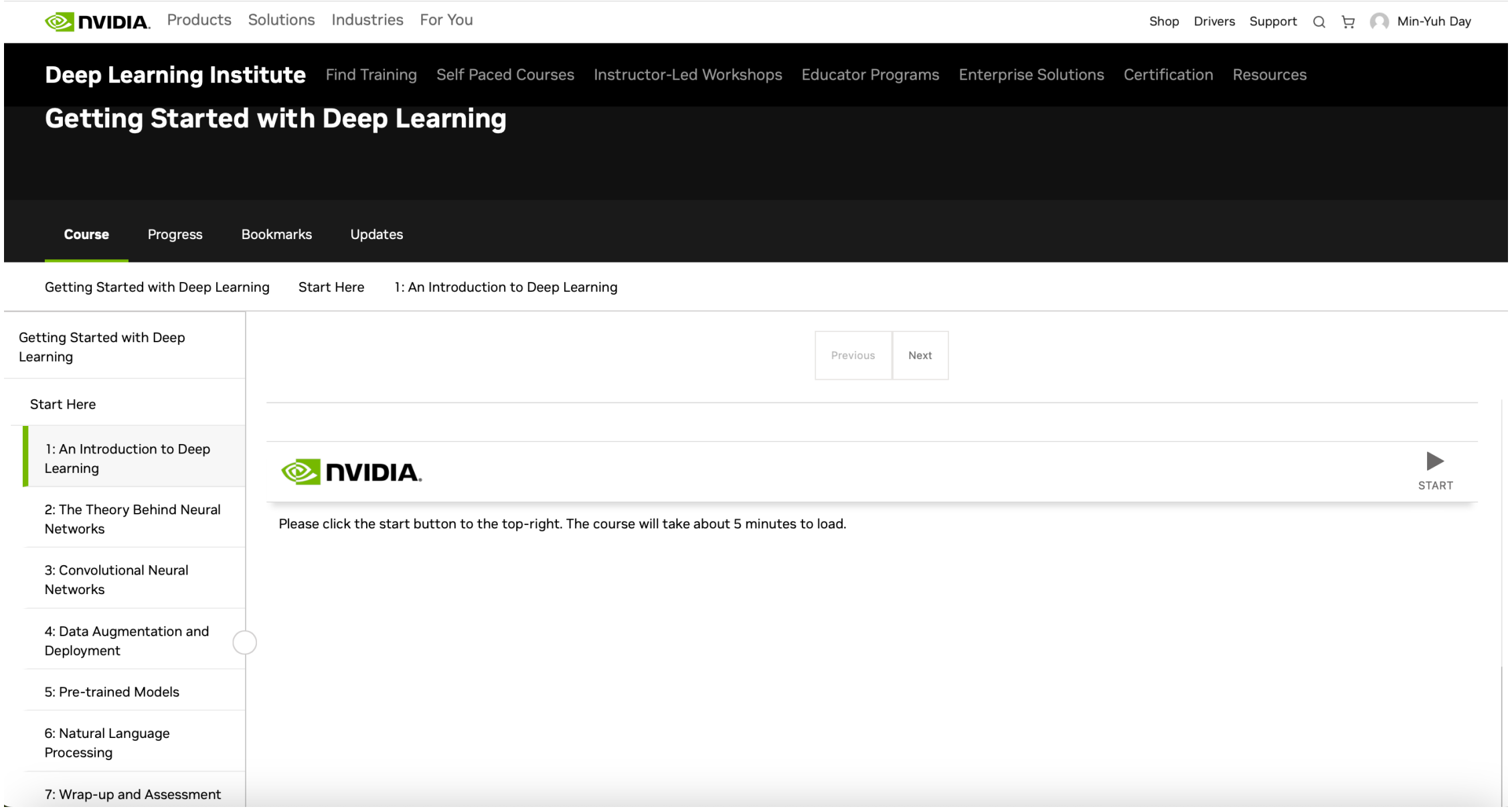
That's what this first lecture is going to be about.

Excited. Me too. Let's get going.

OK so I have my computer down here and the cameras up there and we're going to go over some technical things before we jump into the theory. So below this video you should hopefully see something like... this, that has the start button up in the top right. So go ahead and click it and that's going to wake up our GPU and when we do that, we can also see a copy of the slides

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Getting Started with Deep Learning Start Here 1: An Introduction to Deep Learning

- Getting Started with Deep Learning
- Start Here
  - 1: An Introduction to Deep Learning
  - 2: The Theory Behind Neural Networks
  - 3: Convolutional Neural Networks
  - 4: Data Augmentation and Deployment
  - 5: Pre-trained Models
  - 6: Natural Language Processing
  - 7: Wrap-up and Assessment
- Next Steps
- Feedback

Previous Next

Fundamentals of Deep Learning

Part 1: An Introduction to Deep Learning

0:01 / 13:21 Speed 1.0x

Studying AI could be pretty challenging, but it's also very rewarding.

Not only can

we learn about how robots think, but we can also pick up the few tricks to help humans be more efficient learners too.

That's what this first lecture is going to be about.

Excited. Me too. Let's get going.

OK so I have my computer down here and the cameras up there and we're going to go over some technical things before we jump into the theory. So below this video you should hopefully see something like... this, that has the start button up in the top right. So go ahead and click it and that's going to wake up our GPU and when we do that, we can also see a copy of the slides that we'll be going over.

And what's really cool about this, is down here

NVIDIA

This Lab 0 : 02 : 24 / 1 : 30 : 00

Course 3 : 35 : 51 / 24 : 00 : 00

LAUNCH STOP TASK

The course will take about 5 minutes to load. While it is loading, please feel free to review the slides below.

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Getting Started with Deep Learning

Course Progress Bookmarks Updates

Your progress

Course completion

This represents how much of the course content you have completed. Note that some content may not yet be released.

100% completed

Grades

This represents your weighted grade against the grade needed to pass this course.

Your current grade 100%

Passing grade 90%

You're currently passing this course

Grade summary

Assignment type	Weight	Grade	Weighted grade
Lab	100%	100%	**100%**
Your current weighted grade summary			**100%**

Related links

- Dates
  A schedule view of your course due dates and upcoming assignments.
- Course Outline
  A birds-eye view of your course content.

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Getting Started with Deep Learning

Course Progress Bookmarks Updates

Getting Started with Deep Learning

Pick up where you left off

Resume course

Agenda

Section	Title	Estimated Time to Complete
1	An Introduction to Deep Learning	1.5 Hours
2	The Theory Behind Neural Networks	1.5 Hours
3	Convolutional Neural Networks	1.5 Hours
4	Data Augmentation and Deployment	1.5 Hours
5	Pre-trained Models	1.5 Hours
6	Recurrent Neural Networks	1.5 Hours
7	Wrap-up and Assessment	2 Hours

Dismiss Show More

Expand all

Getting Started with Deep Learning

Start Here (7 Questions)

Next Steps

Feedback

Course Tools

Bookmarks

Updates

Launch tour

Course Handouts

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

Section	Title	Estimated Time to Complete
1	An Introduction to Deep Learning	1.5 Hours
2	The Theory Behind Neural Networks	1.5 Hours
3	Convolutional Neural Networks	1.5 Hours
4	Data Augmentation and Deployment	1.5 Hours
5	Pre-trained Models	1.5 Hours
6	Recurrent Neural Networks	1.5 Hours
7	Wrap-up and Assessment	2 Hours

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification

Fundamentals of Deep Learning

Course Progress Bookmarks Updates

Fundamentals of Deep Learning Start Here Launch the Course

- Fundamentals of Deep Learning
  - Start Here
  - Launch the Course
  - Next Steps
  - Feedback

Previous Next

Bookmark this page

NVIDIA LOADING STOP TASK

The course will take about 5 minutes to load. While it is loading, please feel free to review the slides below.

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

NVIDIA Products Solutions Industries For You Shop Drivers Support Min-Yuh Day

Deep Learning Institute Find Training Self Paced Courses Instructor-Led Workshops Educator Programs Enterprise Solutions Certification Resources

Fundamentals of Deep Learning

Course Progress Bookmarks Updates

Fundamentals of Deep Learning Start Here Launch the Course

- Fundamentals of Deep Learning
  - Start Here
  - Launch the Course
  - Next Steps
  - Feedback

Previous Next

Bookmark this page

NVIDIA

This Lab 0 : 01 : 51 / 9 : 00 : 00

Course 5 : 58 : 53 / 75 : 00 : 00

LAUNCH STOP TASK ASSESS TASK

The course will take about 5 minutes to load. While it is loading, please feel free to review the slides below.

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

PASSED

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

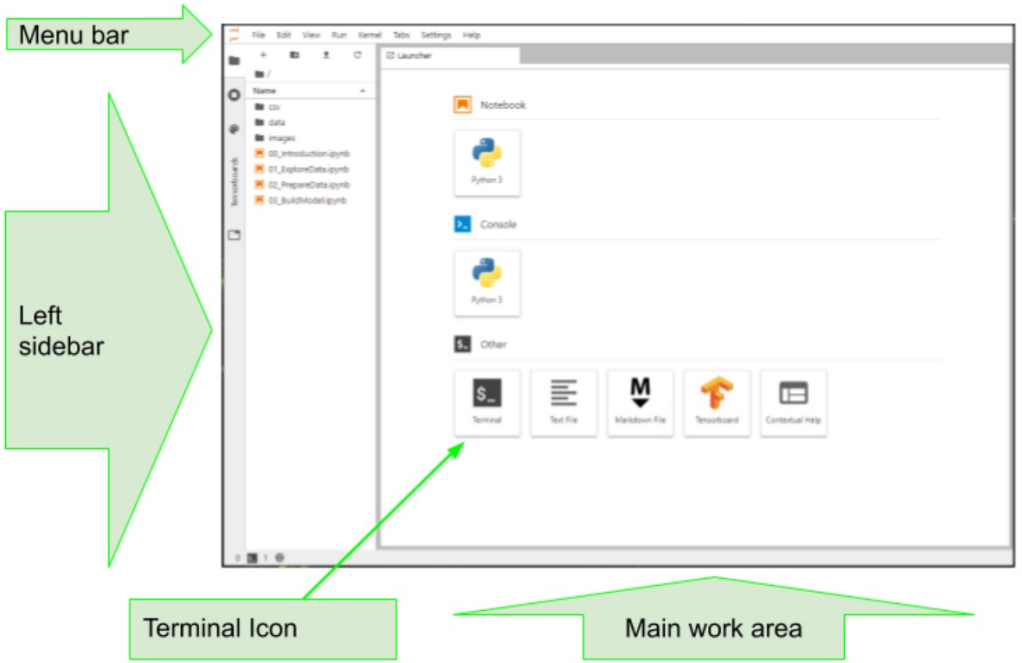
- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun 00_ 01_ 02_ 03_ 04a 04b 05a 05b 06_ 07_ utils

Markdown Python 3 (ipykernel)

JupyterLab

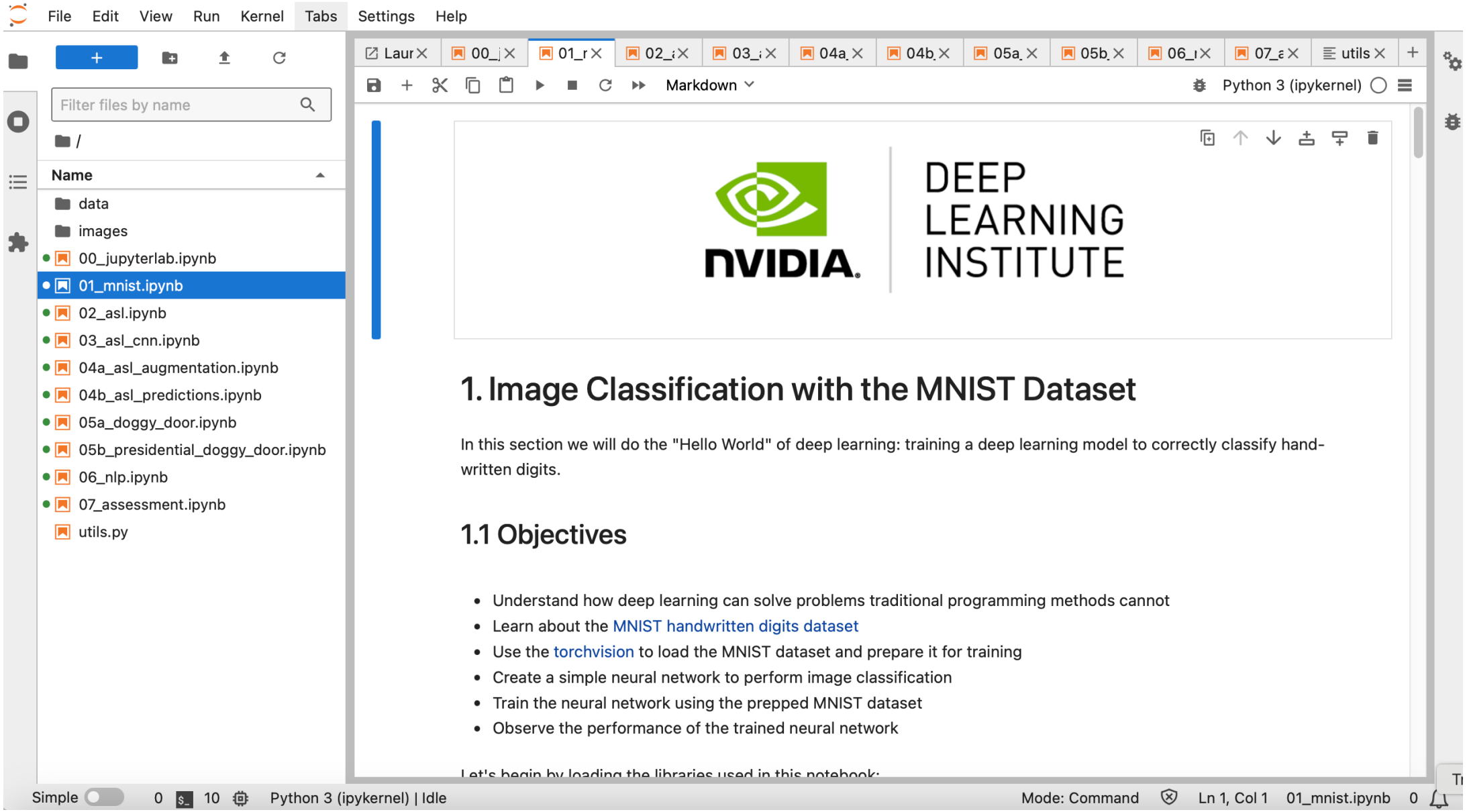
For this hands-on lab, we use JupyterLab to manage our environment. The JupyterLab Interface is a dashboard that provides access to interactive iPython notebooks, as well as the folder structure of our environment and a terminal window into the Ubuntu operating system. The first view you'll see includes a **menu bar** at the top, a **file browser** in the **left sidebar**, and a **main work area** that is initially open to the "Launcher" page.



Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 00_jupyterlab.ipynb

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

2. Image Classification of an American Sign Language Dataset

In this section, we will perform the data preparation, model creation, and model training steps we observed in the last section using a different dataset: images of hands making letters in American Sign Language.

2.1 Objectives

- Prepare image data for training
- Create and compile a simple model for image classification
- Train an image classification model and observe the results

[ ]:
```
import torch.nn as nn
import pandas as pd
import torch
from torch.optim import Adam
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 02_asl.ipynb 0 Trash

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

3. Convolutional Neural Networks

In the previous section, we built and trained a simple model to classify ASL images. The model was able to learn how to correctly classify the training dataset with very high accuracy, but, it did not perform nearly as well on validation dataset. This behavior of not generalizing well to non-training data is called overfitting, and in this section, we will introduce a popular kind of model called a convolutional neural network that is especially good for reading images and classifying them.

3.1 Objectives

- Prep data specifically for a CNN
- Create a more sophisticated CNN model, understanding a greater variety of model layers
- Train a CNN model and observe its performance

[ ]:
```
import torch.nn as nn
import pandas as pd
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 03_asl_cnn.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_n | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

4a. Data Augmentation

So far, we've selected a model architecture that vastly improves the model's performance, as it is designed to recognize important features in the images. The validation accuracy is still lagging behind the training accuracy, which is a sign of overfitting: the model is getting confused by things it has not seen before when it tests against the validation dataset.

In order to teach our model to be more robust when looking at new data, we're going to programmatically increase the size and variance in our dataset. This is known as *data augmentation*, a useful technique for many deep learning applications.

The increase in size gives the model more images to learn from while training. The increase in variance helps the model ignore unimportant features and select only the features that are truly important in classification, allowing it to generalize better.

4a.1 Objectives

- Augment the ASL dataset

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 04a_asl_augmentation.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_n | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

4b. Deploying Your Model

Now that we have a well trained model, it's time to use it. In this exercise, we'll expose new images to our model and detect the correct letters of the sign language alphabet. Let's get started!

4b.1 Objectives

- Load an already-trained model from disk
- Reformat images for a model trained on images of a different format
- Perform inference with new images, never seen by the trained model and evaluate its performance

[ ]:
```
import pandas as pd
import torch
import torch.nn as nn
from torch.optim import Adam
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 04b_asl_predictions.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

5a. Pre-Trained Models ¶

Though it is often necessary to have a large, well annotated dataset to solve a deep learning challenge, there are many freely available pre-trained models that we can use right out of the box. As you decide to take on your own deep learning project, it is a great idea to start by looking for existing models online that can help you achieve your goal. A great place to explore available models is NGC. There are also many models hosted on GitHub that you can find through searching on Google.

5a.1 Objectives

- Use TorchVision to load a very well-trained pretrained model
- Preprocess our own images to work with the pretrained model
- Use the pretrained model to perform accurate inference on your own images

[ ]:
```
import torch
import torchvision.transforms.v2 as transforms
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 05a_doggy_door.ipynb 0 Trash

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_ji | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

5b. Transfer Learning

So far, we have trained accurate models on large datasets, and also downloaded a pre-trained model that we used with no training necessary. But what if we cannot find a pre-trained model that does exactly what you need, and what if we do not have a sufficiently large dataset to train a model from scratch? In this case, there is a very helpful technique we can use called transfer learning.

With transfer learning, we take a pre-trained model and retrain it on a task that has some overlap with the original training task. A good analogy for this is an artist who is skilled in one medium, such as painting, who wants to learn to practice in another medium, such as charcoal drawing. We can imagine that the skills they learned while painting would be very valuable in learning how to draw with charcoal.

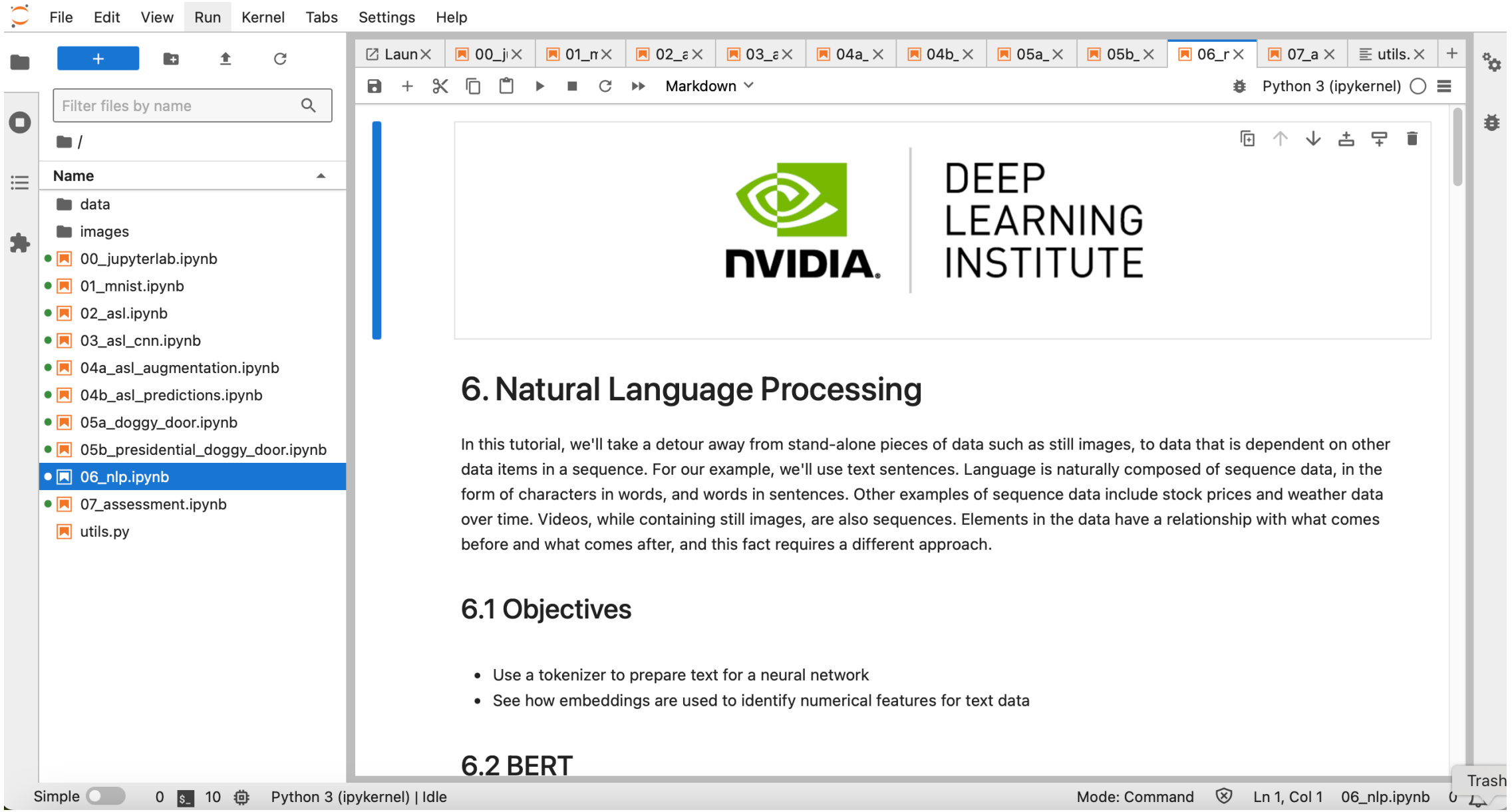
As an example in deep learning, say we have a pre-trained model that is very good at recognizing different types of cars, and we want to train a model to recognize types of motorcycles. A lot of the learnings of the car model would likely be very useful, for instance the ability to recognize headlights and wheels.

Transfer learning is especially powerful when we do not have a large and varied dataset. In this case, a model trained from

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 05b_presidential_doggy_door.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_n | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

7. Assessment

Congratulations on going through today's course! Hopefully, you've learned some valuable skills along the way and had fun doing it. Now it's time to put those skills to the test. In this assessment, you will train a new model that is able to recognize fresh and rotten fruit. You will need to get the model to a validation accuracy of `92%` in order to pass the assessment, though we challenge you to do even better if you can. You will have the use the skills that you learned in the previous exercises. Specifically, we suggest using some combination of transfer learning, data augmentation, and fine tuning. Once you have trained the model to be at least 92% accurate on the validation dataset, save your model, and then assess its accuracy. Let's get started!

[ ]:
```
import torch
import torch.nn as nn
from torch.optim import Adam
from torch.utils.data import Dataset, DataLoader
import torchvision.transforms.v2 as transforms
import torchvision.io as tv_io

import glob
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 07_assessment.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun × | 00_j × | 01_m × | 02_a × | 03_a × | 04a_ × | 04b_ × | 05a_ × | 05b_ × | 06_r × | 07_a × | utils. ×

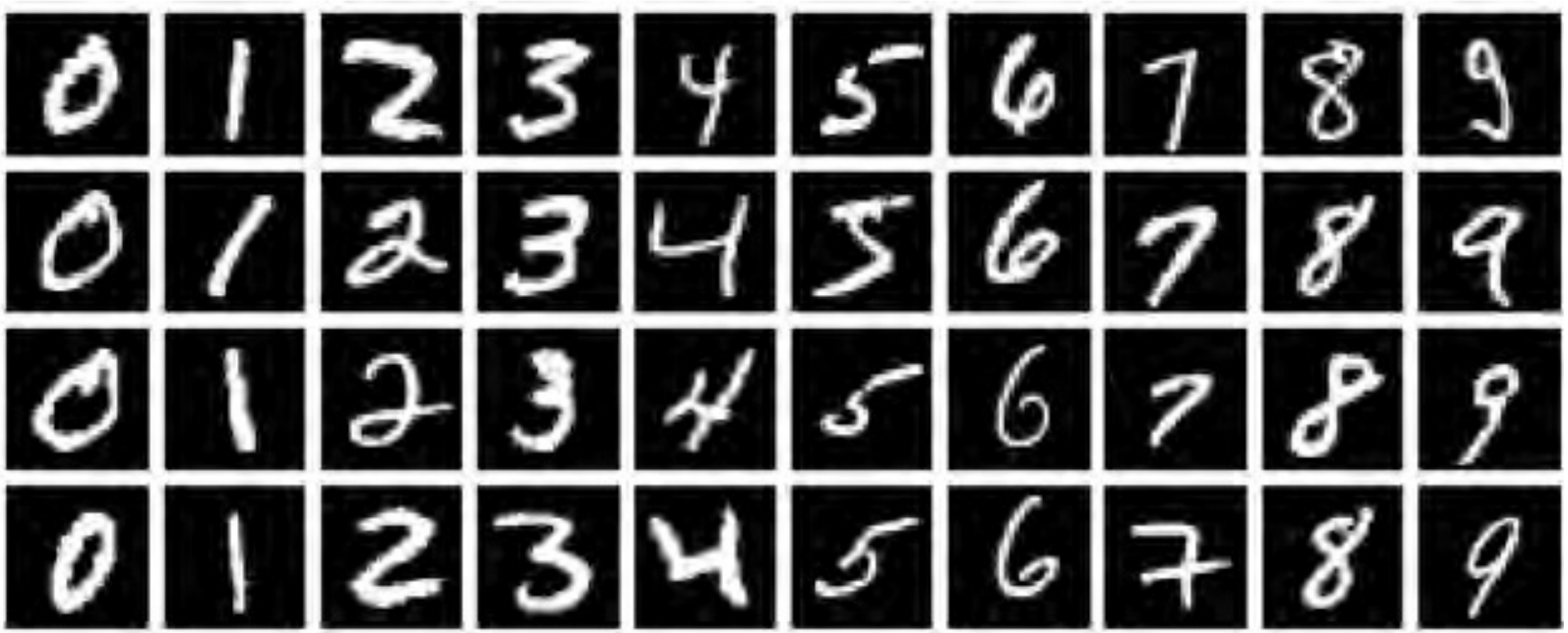
```python
import torch
import torch.nn as nn

class MyConvBlock(nn.Module):
    def __init__(self, in_ch, out_ch, dropout_p):
        kernel_size = 3
        super().__init__()

        self.model = nn.Sequential(
            nn.Conv2d(in_ch, out_ch, kernel_size, stride=1, padding=1),
            nn.BatchNorm2d(out_ch),
            nn.ReLU(),
            nn.Dropout(dropout_p),
            nn.MaxPool2d(2, stride=2)
        )

    def forward(self, x):
        return self.model(x)

def get_batch_accuracy(output, y, N):
    pred = output.argmax(dim=1, keepdim=True)
    correct = pred.eq(y.view_as(pred)).sum().item()
    return correct / N


def train(model, train_loader, train_N, random_trans, optimizer, loss_function):
    loss = 0
    accuracy = 0

    model.train()
    for x, y in train_loader:
        output = model(random_trans(x))
        optimizer.zero_grad()
        batch_loss = loss_function(output, y)
```

Simple 0 10 Python Ln 15, Col 10 Spaces: 4 utils.py Trash

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

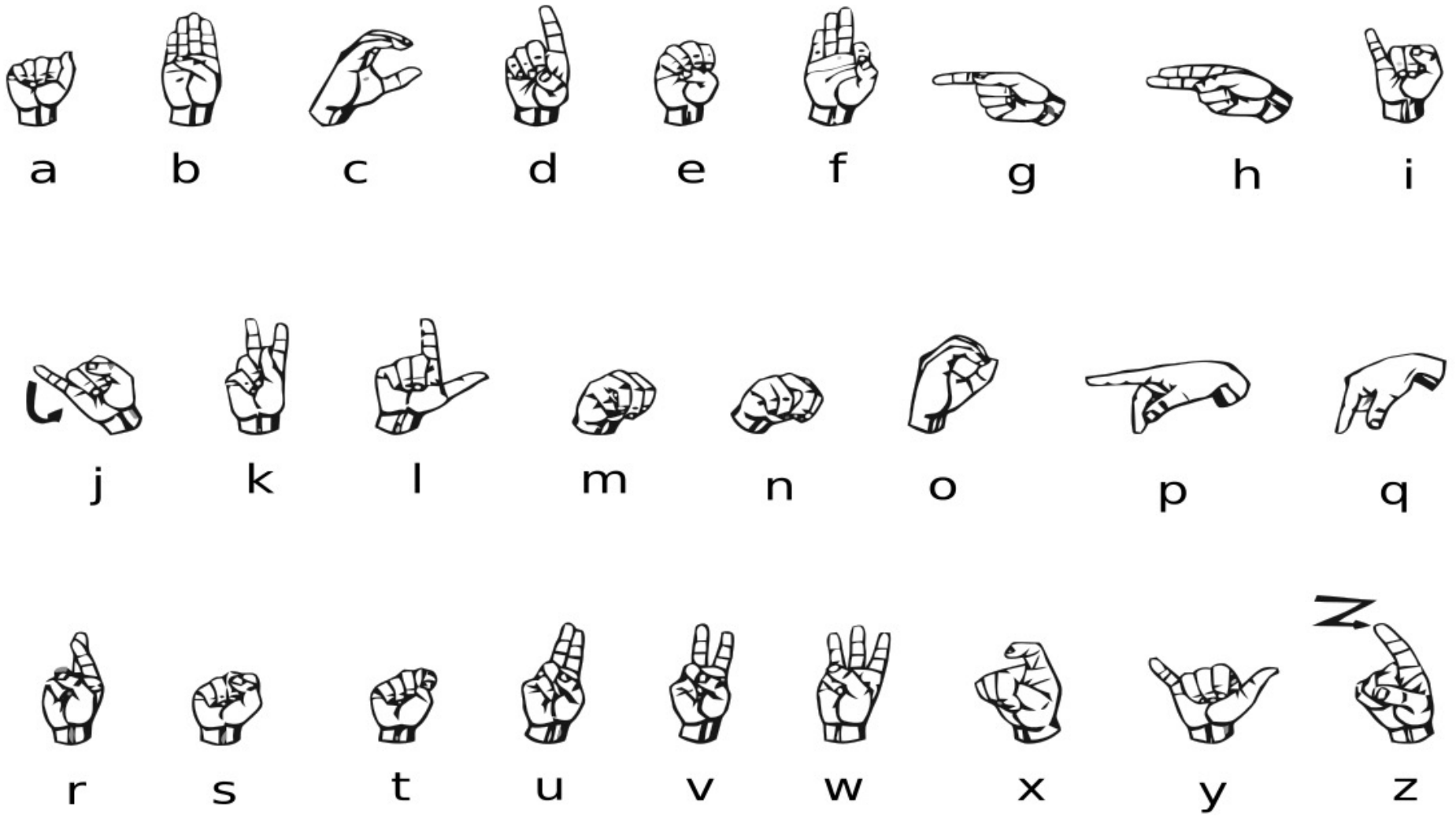
MNIST Handwritten Digits



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

American Sign Language (ASL) Alphabet



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

Fruits fresh and rotten for classification

Kaggle Datasets: https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Part 1: An Introduction to Deep Learning

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Agenda

- **Part 1: An Introduction to Deep Learning**
- **Part 2: How a Neural Network Trains**
- **Part 3: Convolutional Neural Networks**
- **Part 4: Data Augmentation and Deployment**
- **Part 5: Pre-Trained Models**
- **Part 6: Advanced Architectures**

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

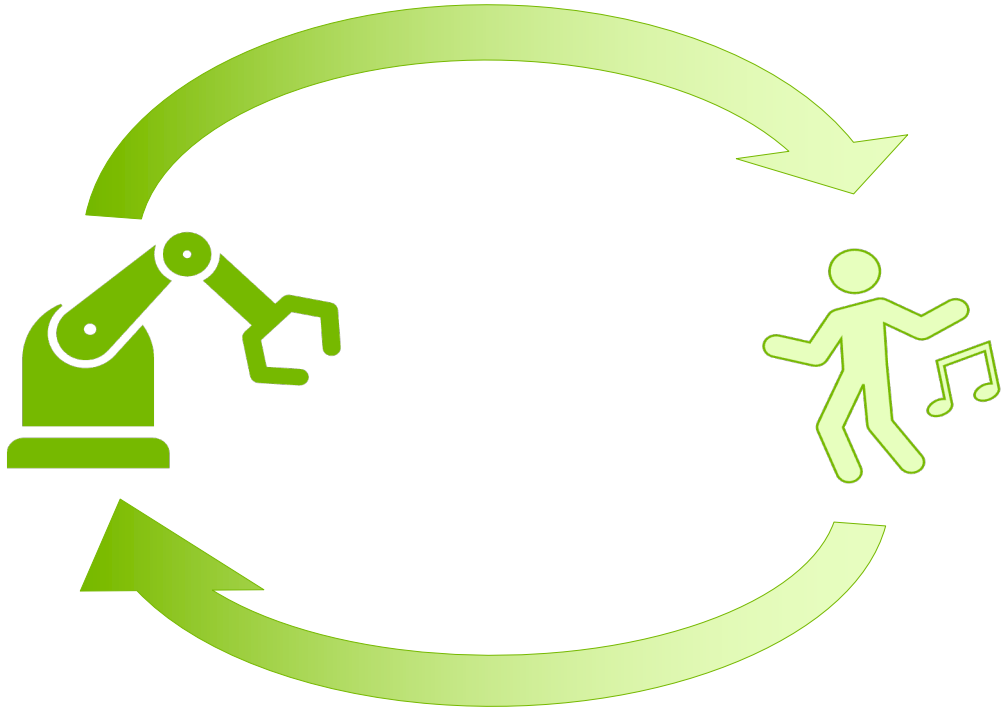
The Goals of This Course

- Get you up and on your feet quickly
- Build a foundation to tackle a deep learning project right away
- We won’t cover the whole field, but we’ll get a great head start
- Foundation from which to read articles, follow tutorials, take further classes

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Human Vs Machine Learning

Relaxed Alertness



Human	Machine
Rest and Digest	Training
Fight-or-flight	Prediction

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

History of AI

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Beginning of Artificial Intelligence

Computers are made in part to complete human tasks

→

Early on, generalized intelligence looked possible

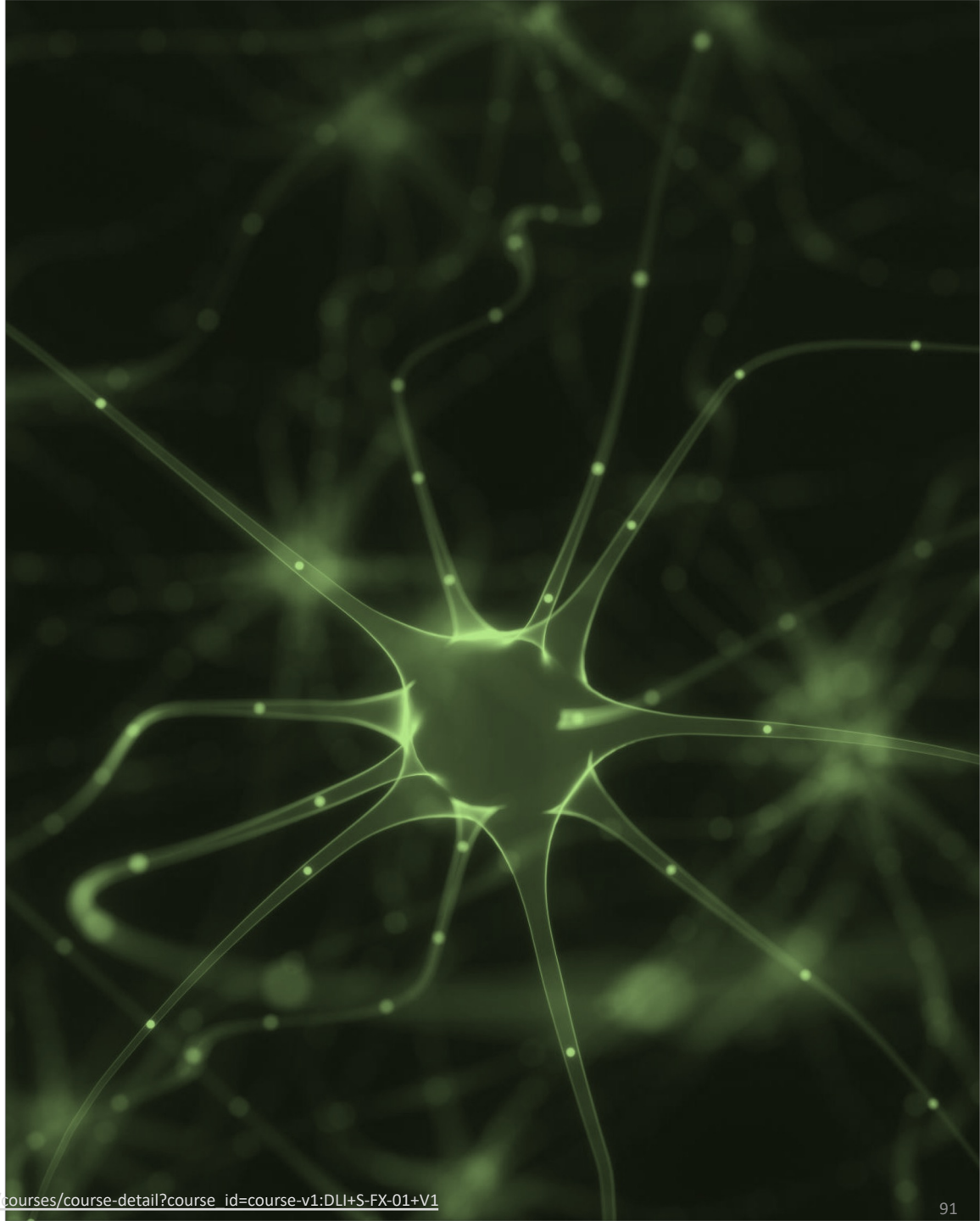
→

Turned out to be harder than expected

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Early Neural Networks

- Inspired by biology
- Created in the 1950's
- Outclassed by Von Neumann Architecture



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Expert Systems

Highly complex

Programmed by hundreds of engineers

Rigorous programming of many rules

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Expert Systems - Limitations

What are these three images?

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

How Do Children Learn?

- Expose them to lots of data
- Give them the “correct answer”
- They will pick up the important patterns on their own

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Deep Learning Revolution

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Data

- Networks need a lot of information to learn from
- The digital era and the internet has supplied that data

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Computing Power

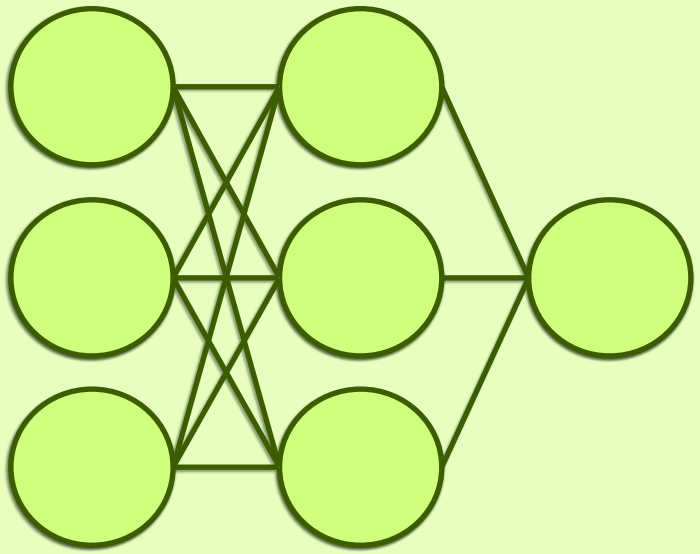
Need a way for our artificial “brain” to observe lots of data within a practical amount of time.

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Importance of the GPU

A Rendered Image

A Neural Network



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

What is Deep Learning?

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

“

Deep learning flips traditional programming on its head

”

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Traditional Programming

Building a Classifier

1. Define a set of rules for classification
2. Program those rules into the computer
3. Feed it examples, and the program uses the rules to classify

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Machine Learning

Building a Classifier

1. Show model the examples with the answer of how to classify

2. Model takes guesses, we tell it if it’s right or not

3. Model learns to correctly categorize as it’s training. The system learns the rules on its own

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

This is a Fundamental Shift

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

When to Choose Deep Learning

Classic Programming

If rules are clear and straightforward, often better to program it

Deep Learning

If rules are nuanced, complex, difficult to discern, use deep learning

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Deep Learning Compared to Other AI

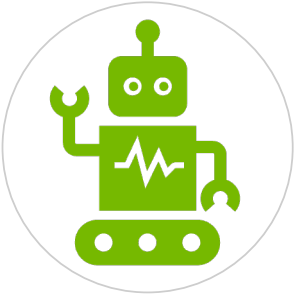
- Depth and complexity of networks
- Up to billions of parameters (and growing)
- Many layers in a model
- Important for learning complex rules

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

How Deep Learning is Transforming the World

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Computer Vision



ROBOTICS AND MANUFACTURING

OBJECT DETECTION

SELF-DRIVING CARS

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Natural Language Processing

REAL-TIME TRANSLATION

VOICE RECOGNITION

VIRTUAL ASSISTANTS

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Recommender Systems

CONTENT CURATION

TARGETED ADVERTISING

SHOPPING

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Reinforcement Learning

ALPHAGO BEATS WORLD CHAMPION IN GO

AI BOTS BEAT PROFESSIONAL VIDEOGAMERS

STOCK TRADING ROBOTS

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Overview of the Course

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Hands on Exercises

- Get comfortable with the process of deep learning
- Exposure to different models and datatypes
- Get a jump-start to tackle your own projects

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Structure of the Course

"Hello World" of Deep Learning

Train a more complicated model

New architectures and techniques to improve performance

Pre-trained models

Transfer learning

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Platform of the Course

GPU powered cloud server

JupyterLab platform

Jupyter notebooks for interactive coding

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Software of This Course

- Major deep learning platforms:
  - TensorFlow + Keras (Google)
  - PyTorch (Meta)
  - MXNet (Apache)
- We'll be using PyTorch
- Good idea to gain exposure to others moving forward

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

First Exercise:
Classify Handwritten Digits

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Hello Neural Networks

Train a network to correctly classify handwritten digits

Historically important and difficult task for computers

Try learning like a Neural Network

Get exposed to the example, and try to figure out the rules to how it works

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

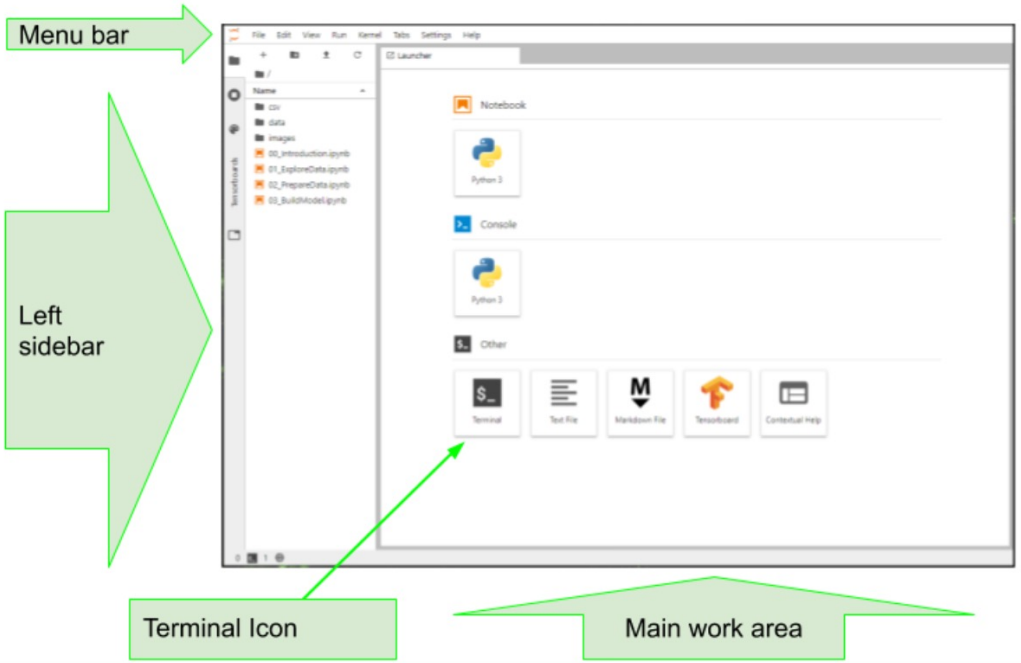
- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_ | 01_ | 02_ | 03_ | 04a | 04b | 05a | 05b | 06_ | 07_ | utils

Markdown | Python 3 (ipykernel)

JupyterLab

For this hands-on lab, we use JupyterLab to manage our environment. The JupyterLab Interface is a dashboard that provides access to interactive iPython notebooks, as well as the folder structure of our environment and a terminal window into the Ubuntu operating system. The first view you'll see includes a **menu bar** at the top, a **file browser** in the **left sidebar**, and a **main work area** that is initially open to the "Launcher" page.



Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 00_jupyterlab.ipynb

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

1. Image Classification with the MNIST Dataset

In this section we will do the "Hello World" of deep learning: training a deep learning model to correctly classify hand-written digits.

1.1 Objectives

- Understand how deep learning can solve problems traditional programming methods cannot
- Learn about the MNIST handwritten digits dataset
- Use the torchvision to load the MNIST dataset and prepare it for training
- Create a simple neural network to perform image classification
- Train the neural network using the prepped MNIST dataset
- Observe the performance of the trained neural network

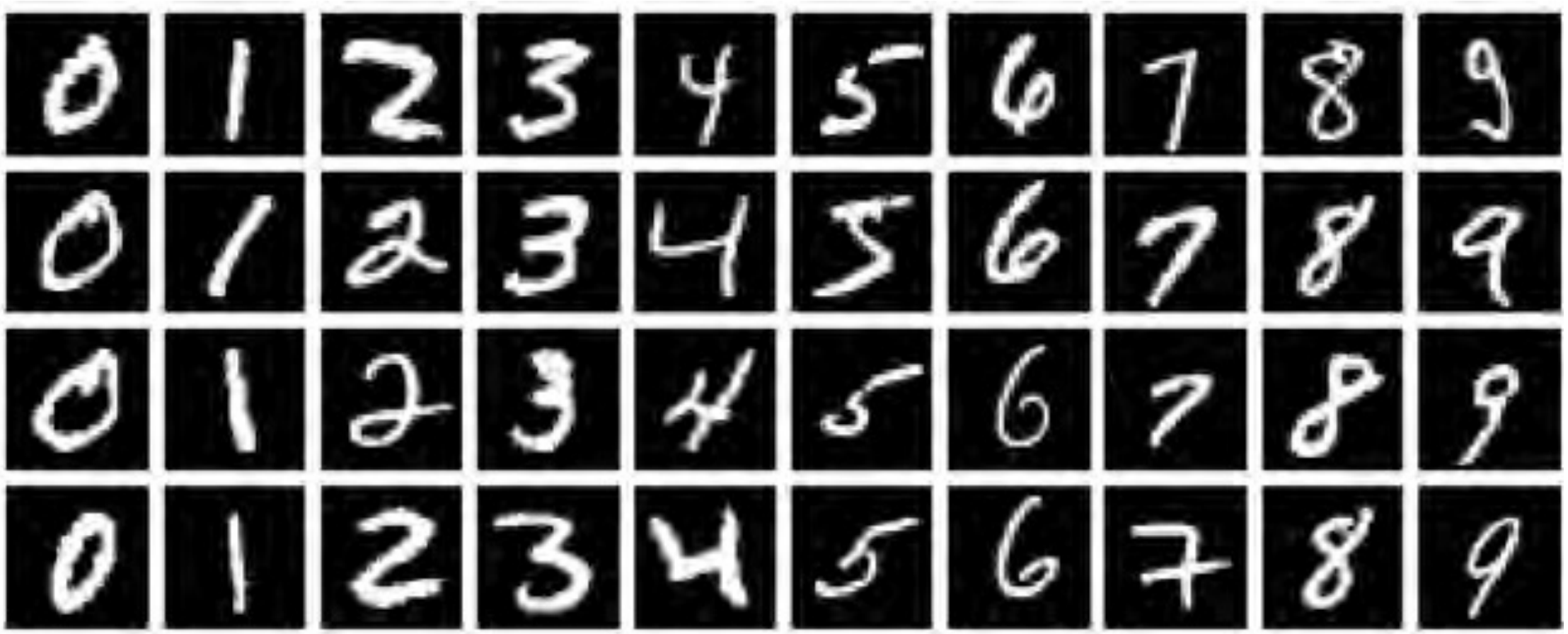
Let's begin by loading the libraries used in this notebook:

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 01_mnist.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

MNIST Handwritten Digits



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Part 2: How a Neural Network Trains

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Agenda

- Part 1: An Introduction to Deep Learning
- Part 2: How a Neural Network Trains
- Part 3: Convolutional Neural Networks
- Part 4: Data Augmentation and Deployment
- Part 5: Pre-Trained Models
- Part 6: Advanced Architectures

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Recap of the Exercise

What just happened?

Loaded and visualized our data

Edited our data (reshaped, normalized, to categorical)

Created our model

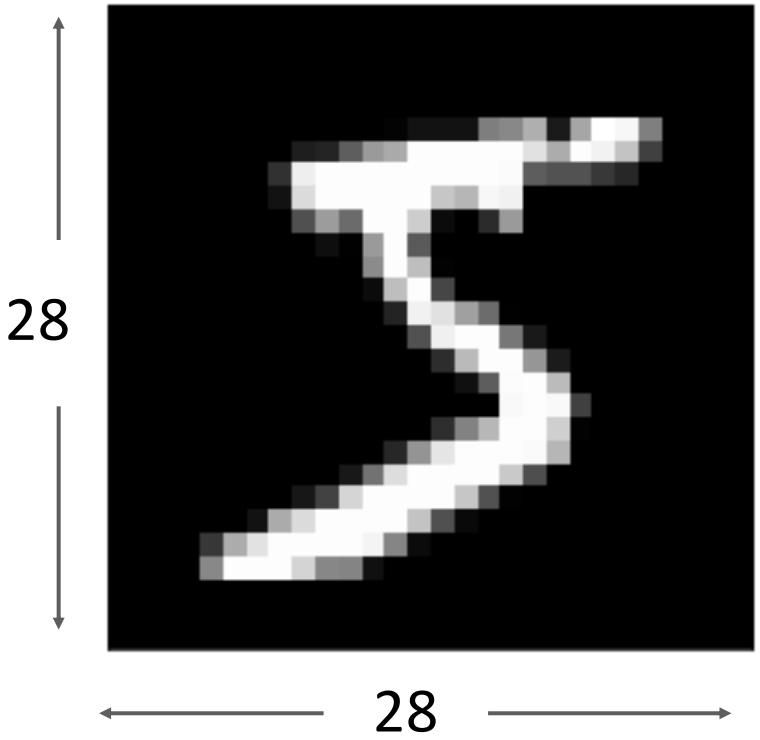
Compiled our model

Trained the model on our data

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Data Preparation

Input as an Array



[0,0,0,24,75,184,185,78,32,55,0,0,0…]

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Data Preparation

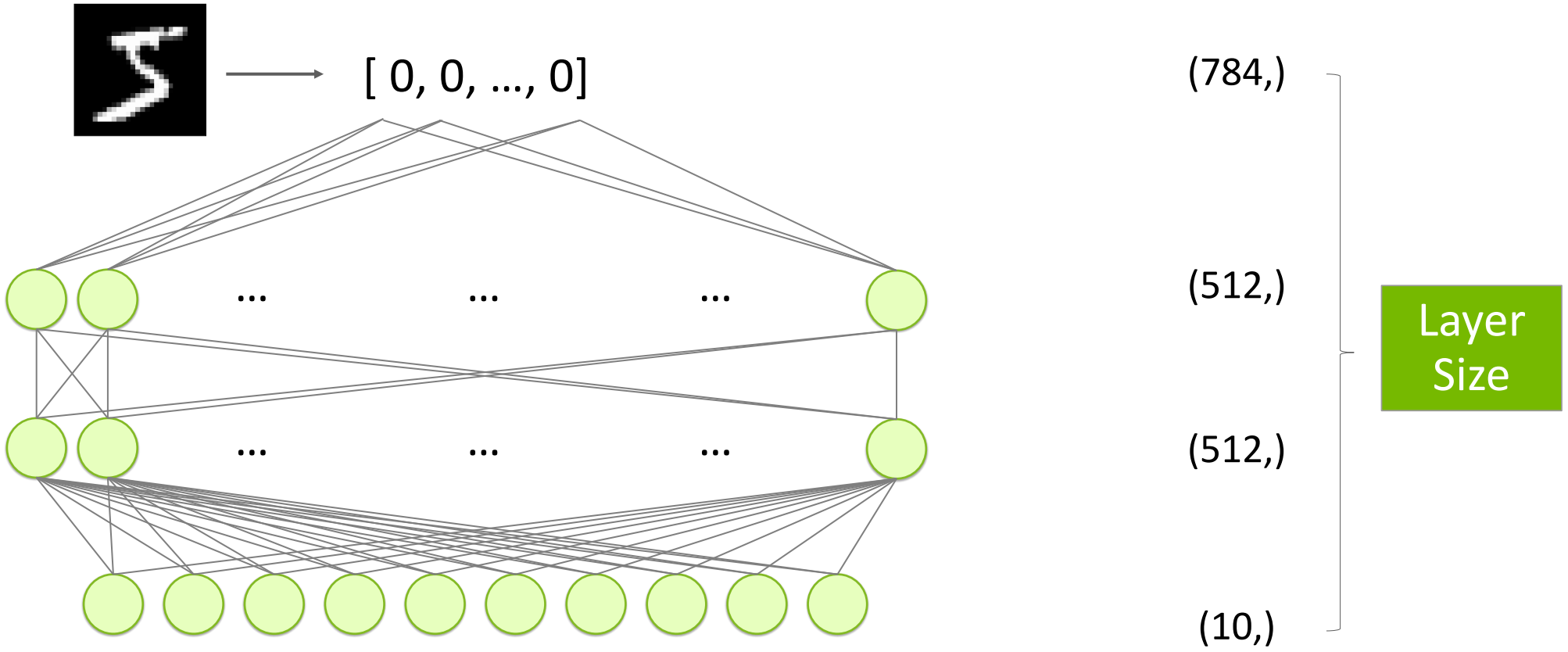
Targets as Categories

0 →	[1,0,0,0,0,0,0,0,0,0]
1 →	[0,1,0,0,0,0,0,0,0,0]
2 →	[0,0,1,0,0,0,0,0,0,0]
3 →	[0,0,0,1,0,0,0,0,0,0]

⋮

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

An Untrained Model



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

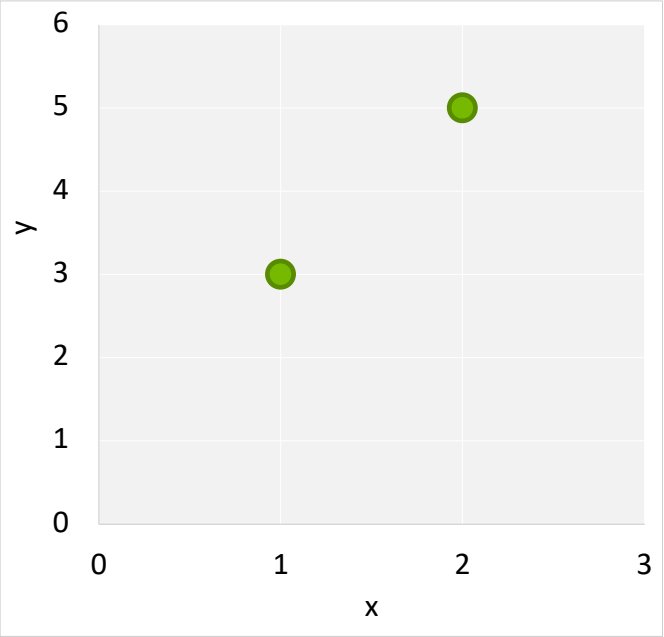
A Simpler Model

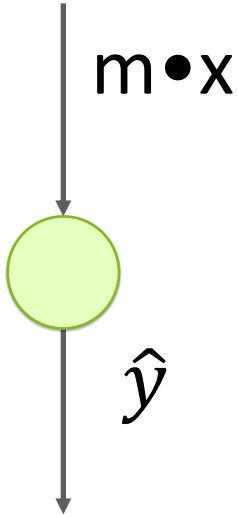
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

A Simpler Model

$$y = mx + b$$

x	y
1	3
2	5





$m = ?$

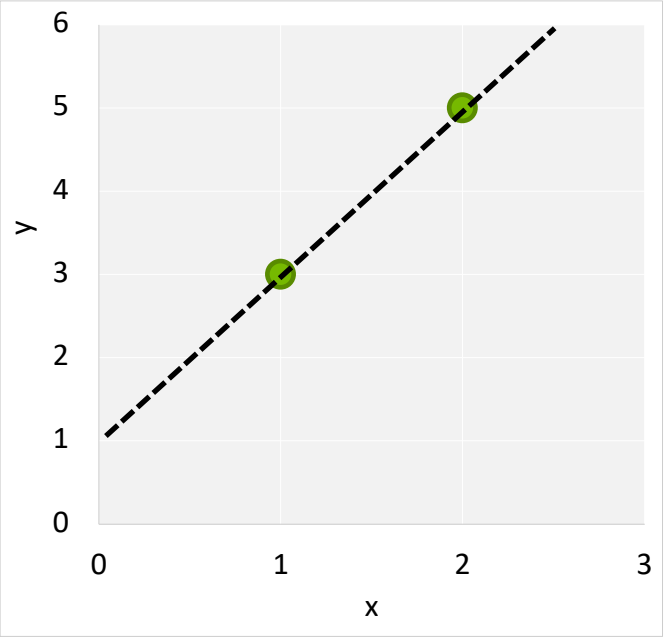
$b = ?$

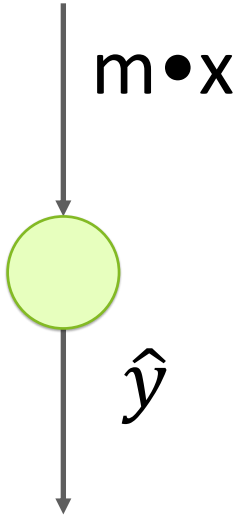
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

A Simpler Model

$$y = mx + b$$

x	y
1	3
2	5





$m = ?$

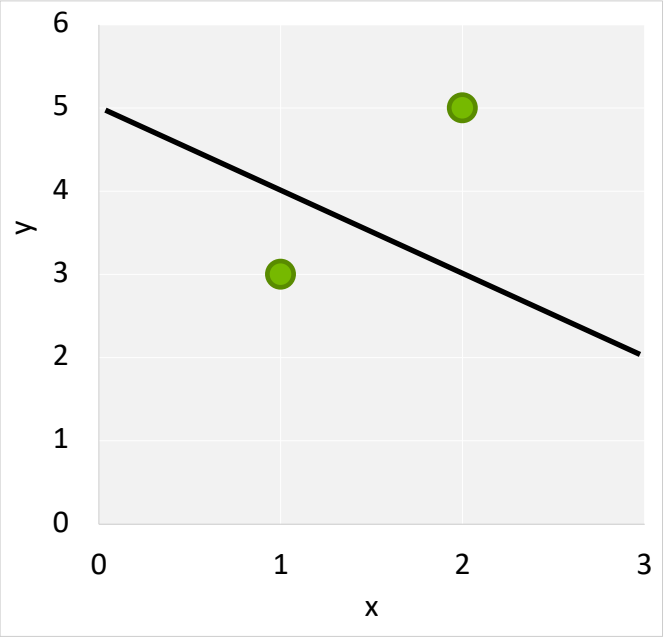
$b = ?$

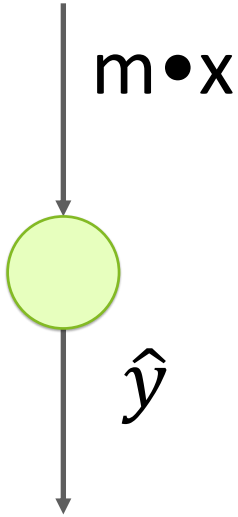
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

A Simpler Model

$$y = mx + b$$

x	y	$\hat{y}$
1	3	4
2	5	3





Start Random

$$m = -1$$

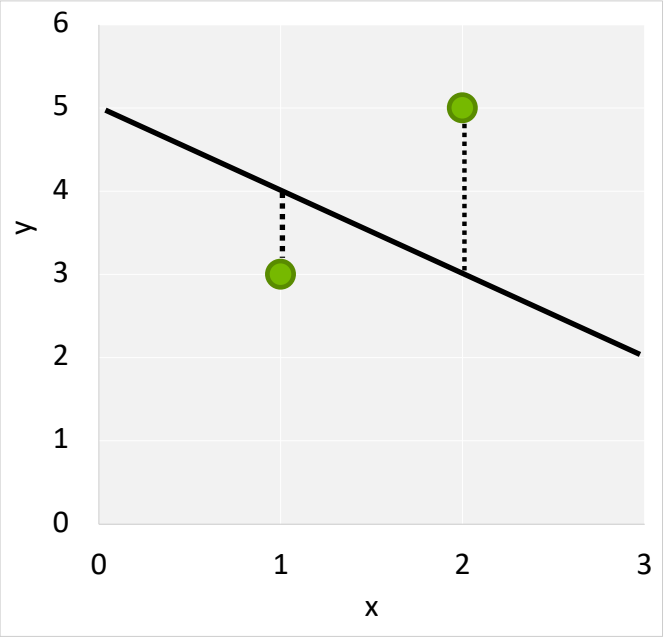
$$\mathrm{b} = 5$$

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

A Simpler Model

$$y = mx + b$$

x	y	$\hat{y}$	err^2
1	3	4	1
2	5	3	4
		MSE =	2.5
		RMSE =	1.6



$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

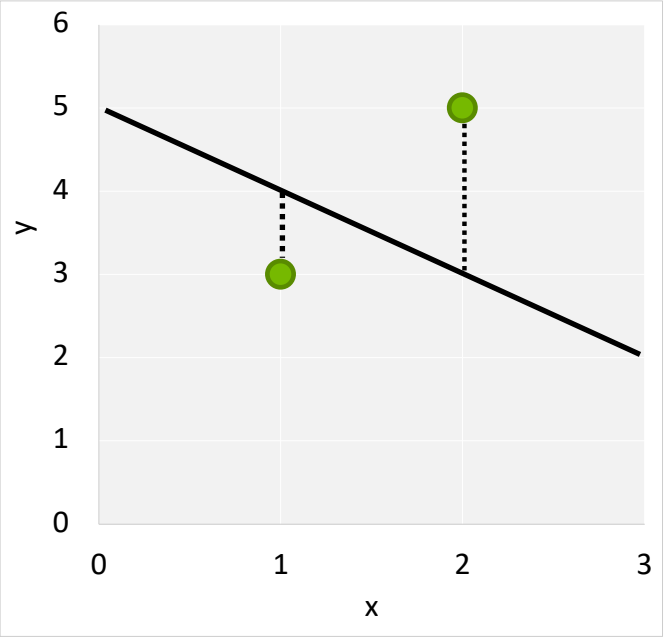
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

A Simpler Model

$$y = mx + b$$

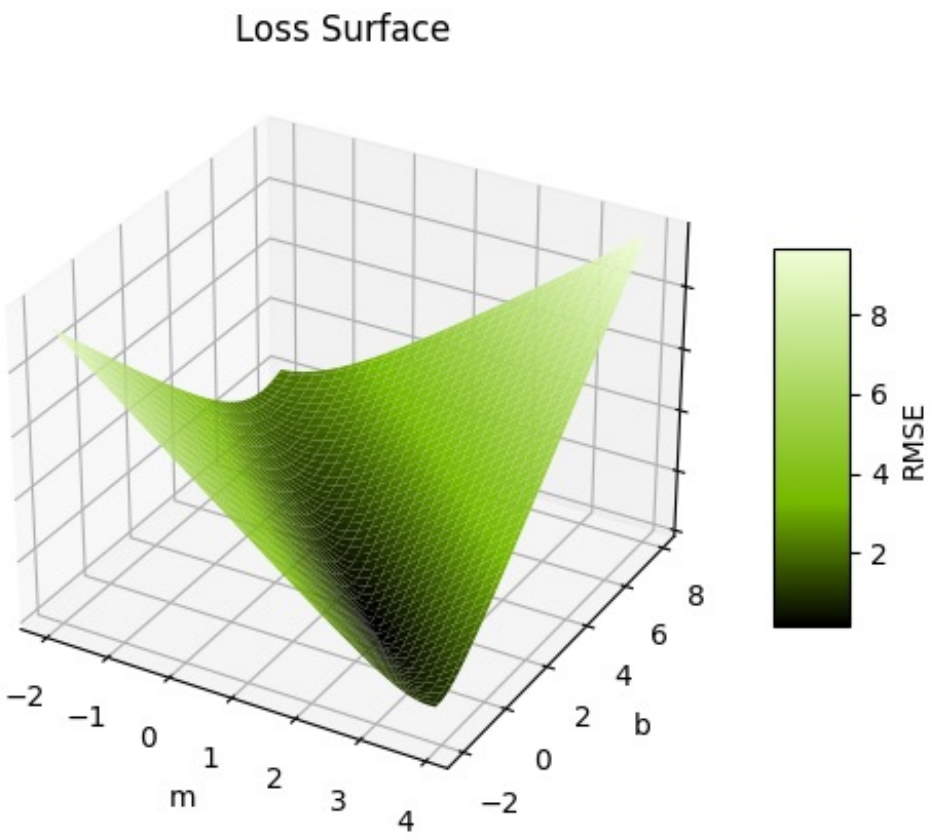
x	y	$\hat{y}$	err^2
1	3	4	1
2	5	3	4
		MSE =	2.5
		RMSE =	1.6

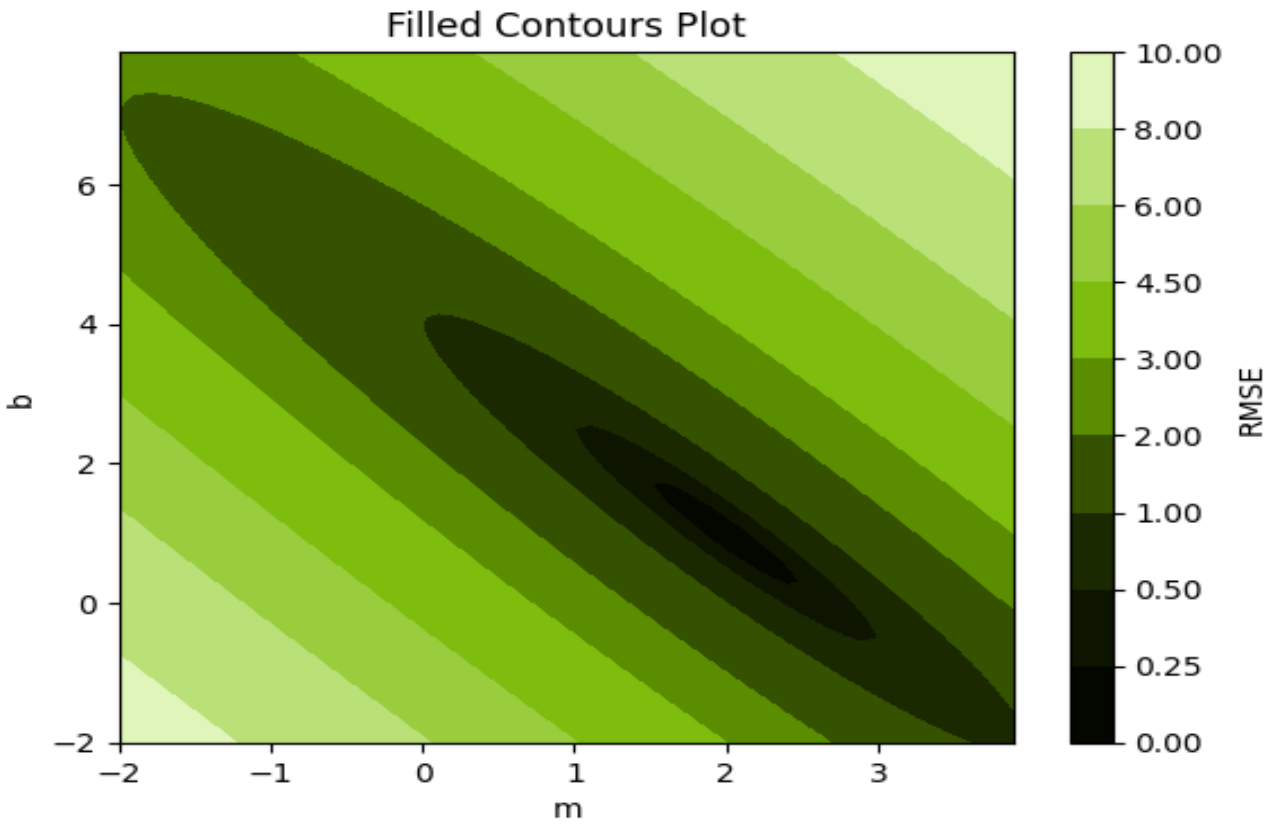



```python
data = [(1, 3), (2, 5)]
m = -1
b = 5


def get_rmse(data, m, b):
    """Calculates Mean Square Error"""
    n = len(data)
    squared_error = 0
    for x, y in data:
        # Find predicted y
        y_hat = m*x+b
        # Square difference between
        # prediction and true value
        squared_error += (
            y - y_hat)**2
    # Get average squared difference
    mse = squared_error / n
    # Square root for original units
    return mse ** .5
```

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

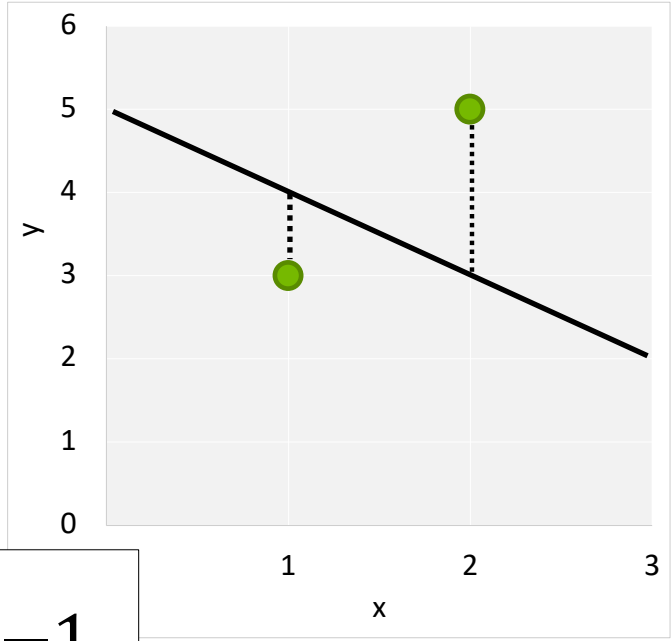
The Loss Curve





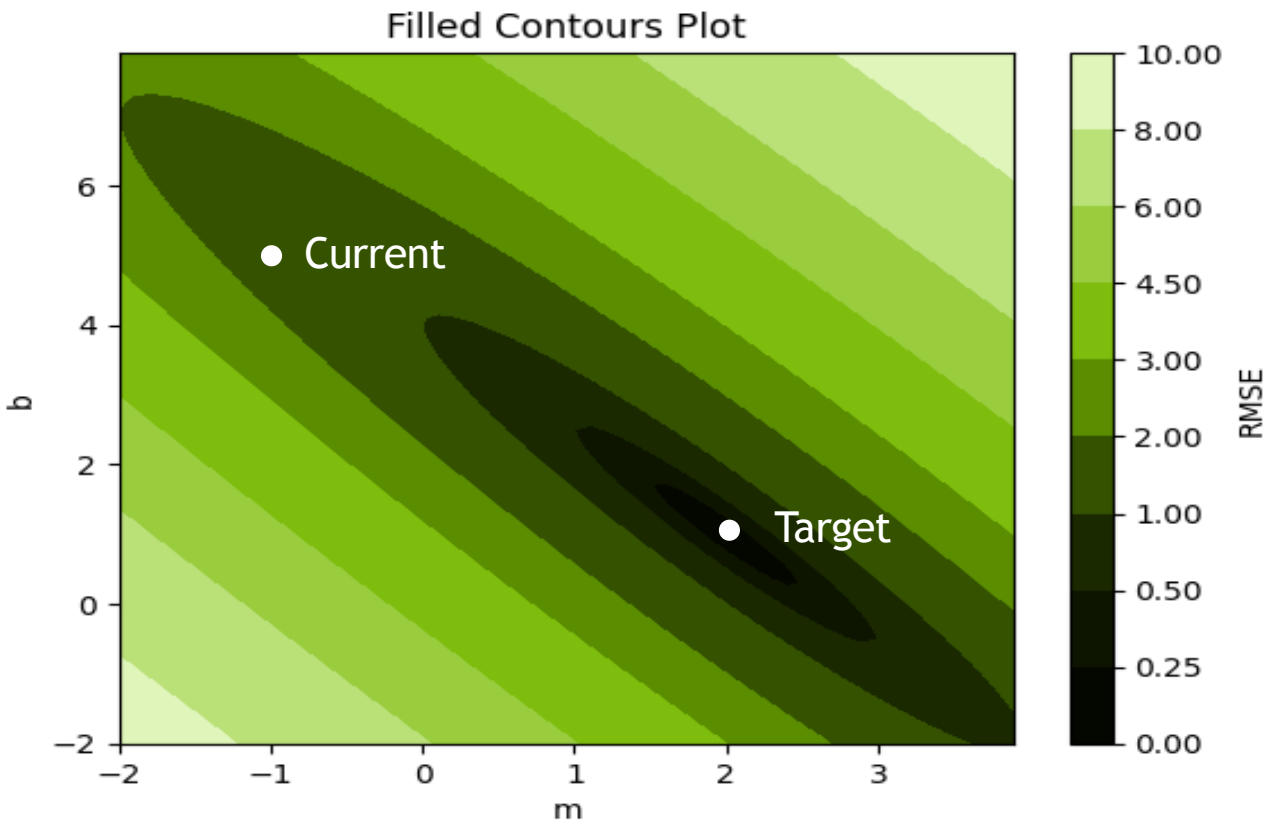
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve



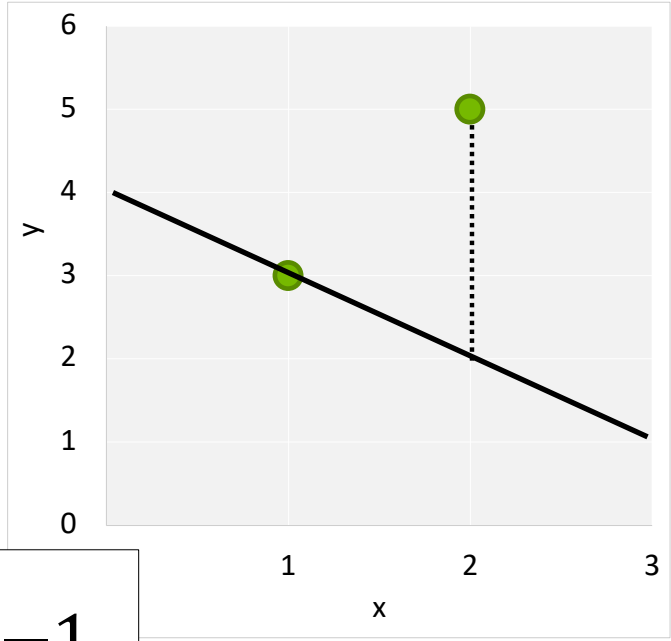
$m = -1$

$\mathrm{b} = 5$



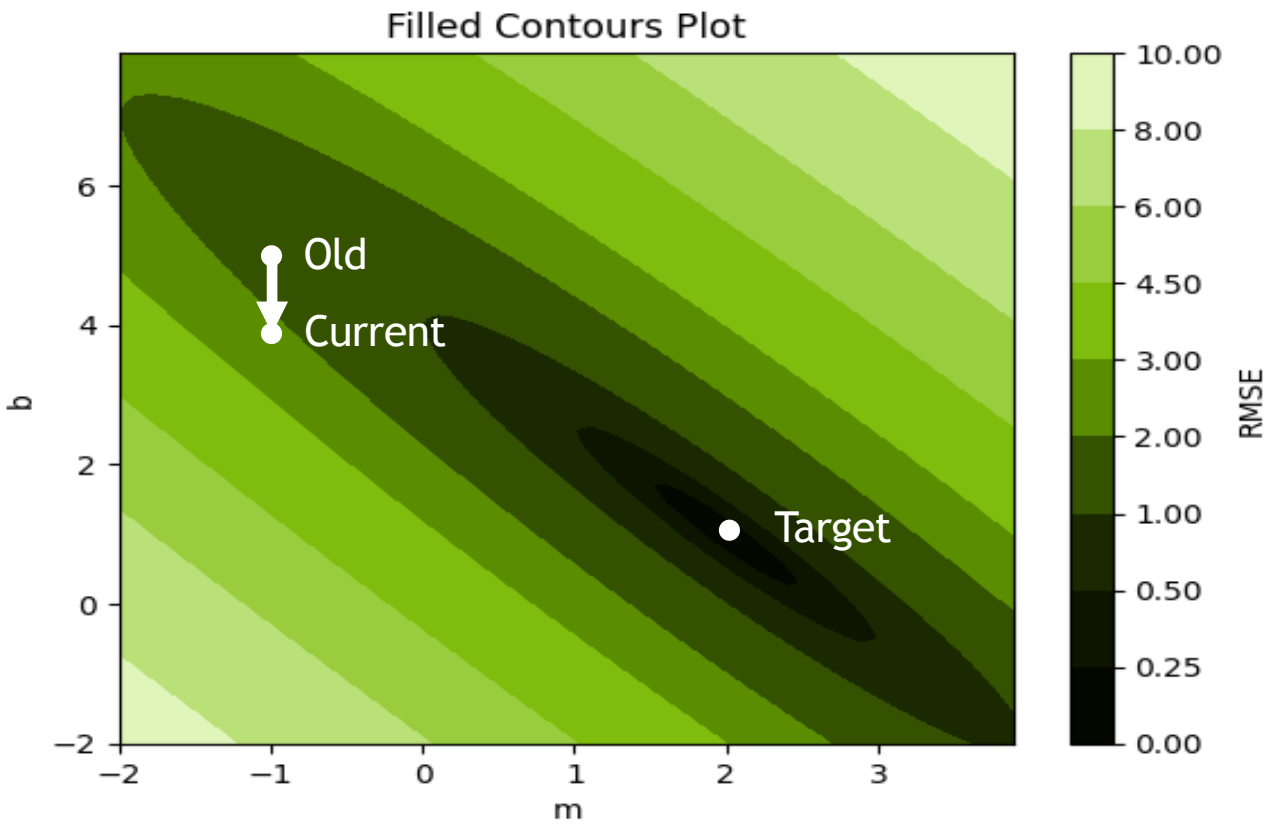
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve



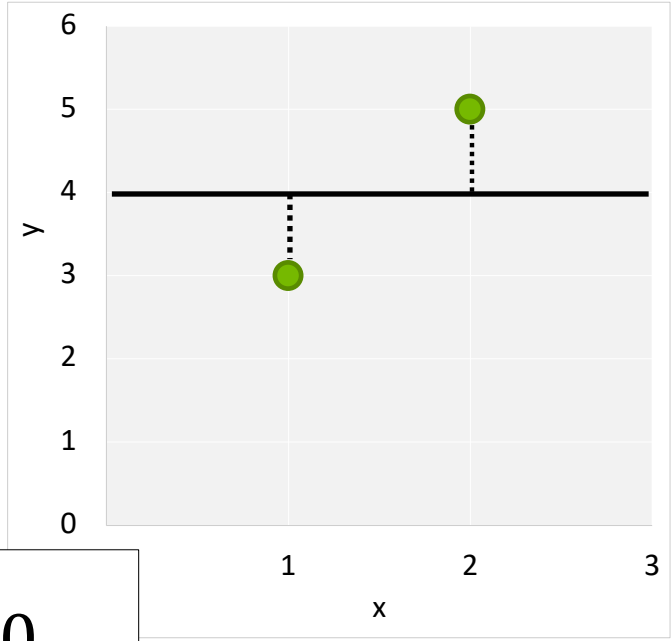
$$m = -1$$

$$\mathrm{b} = 4$$



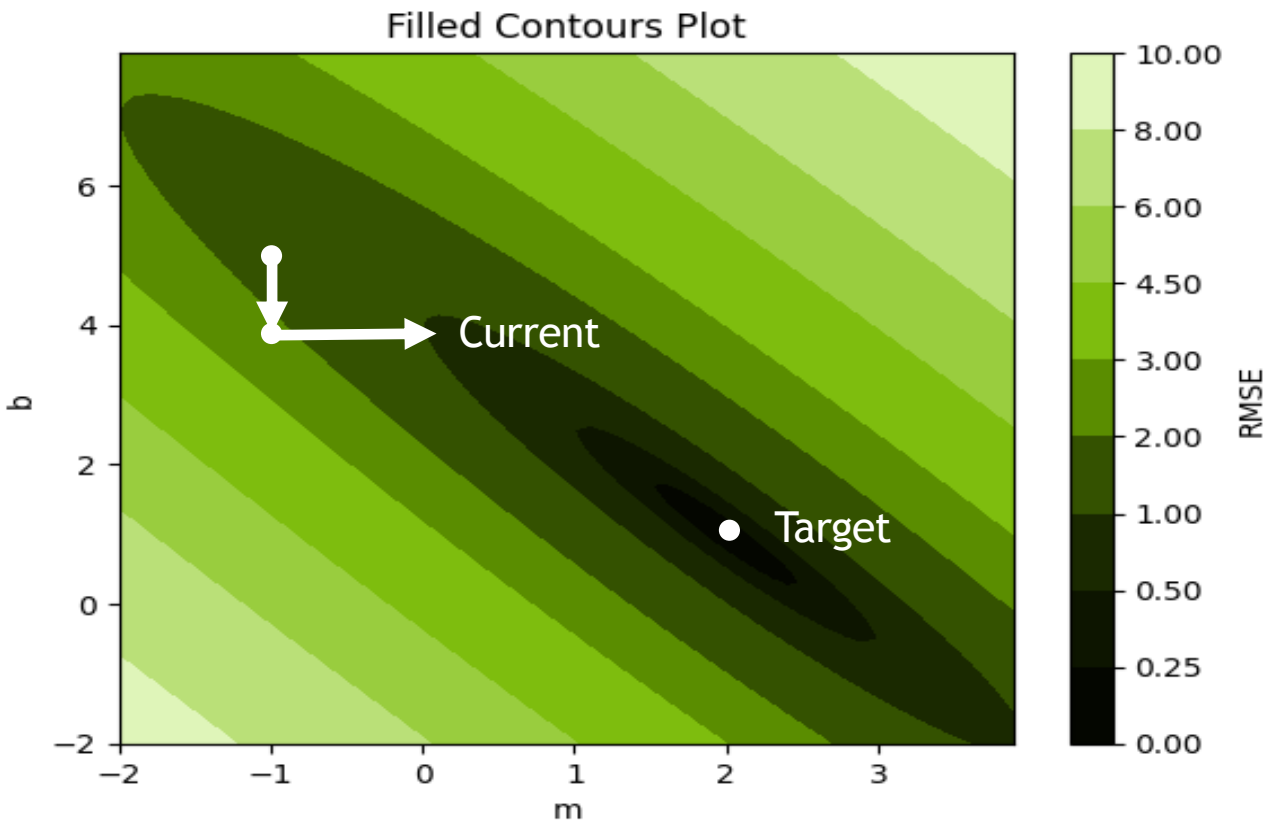
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve



$m = 0$

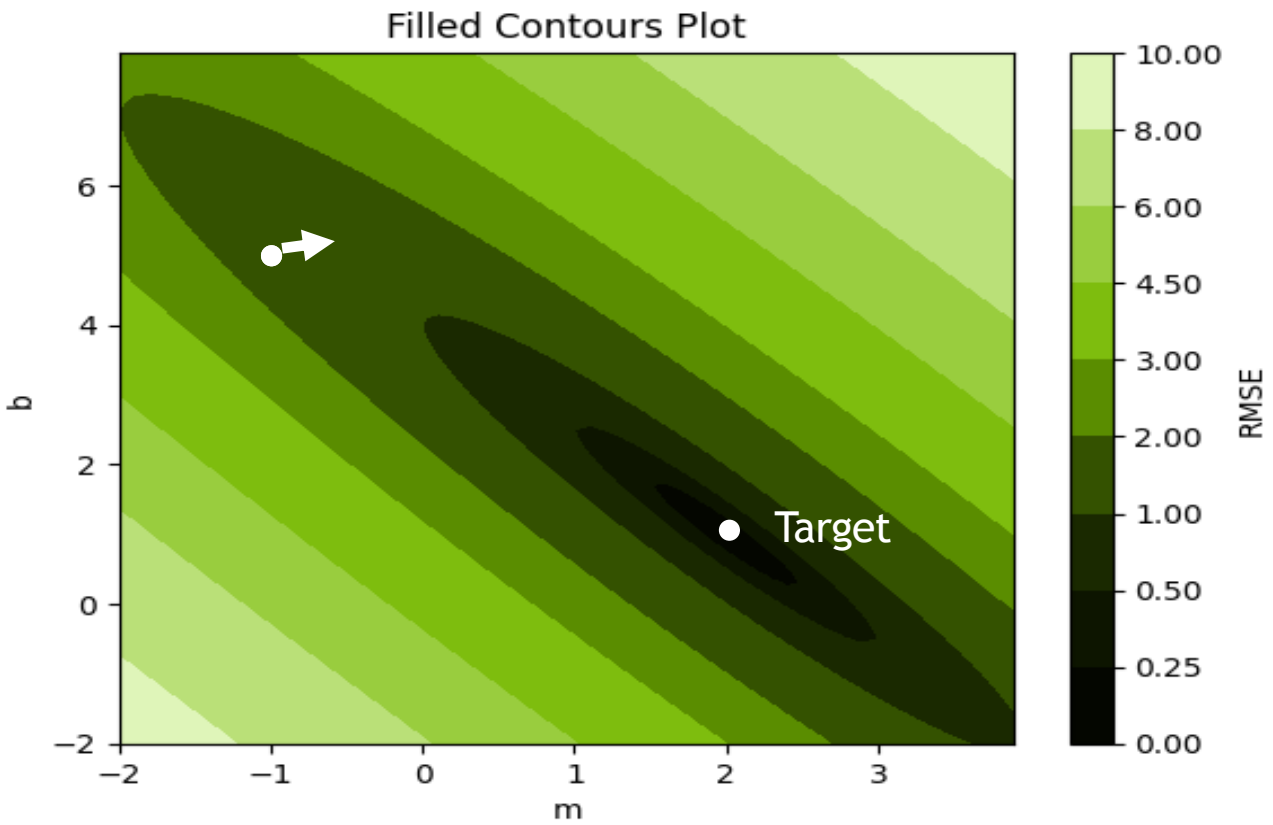
$\mathrm{b} = 4$



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve

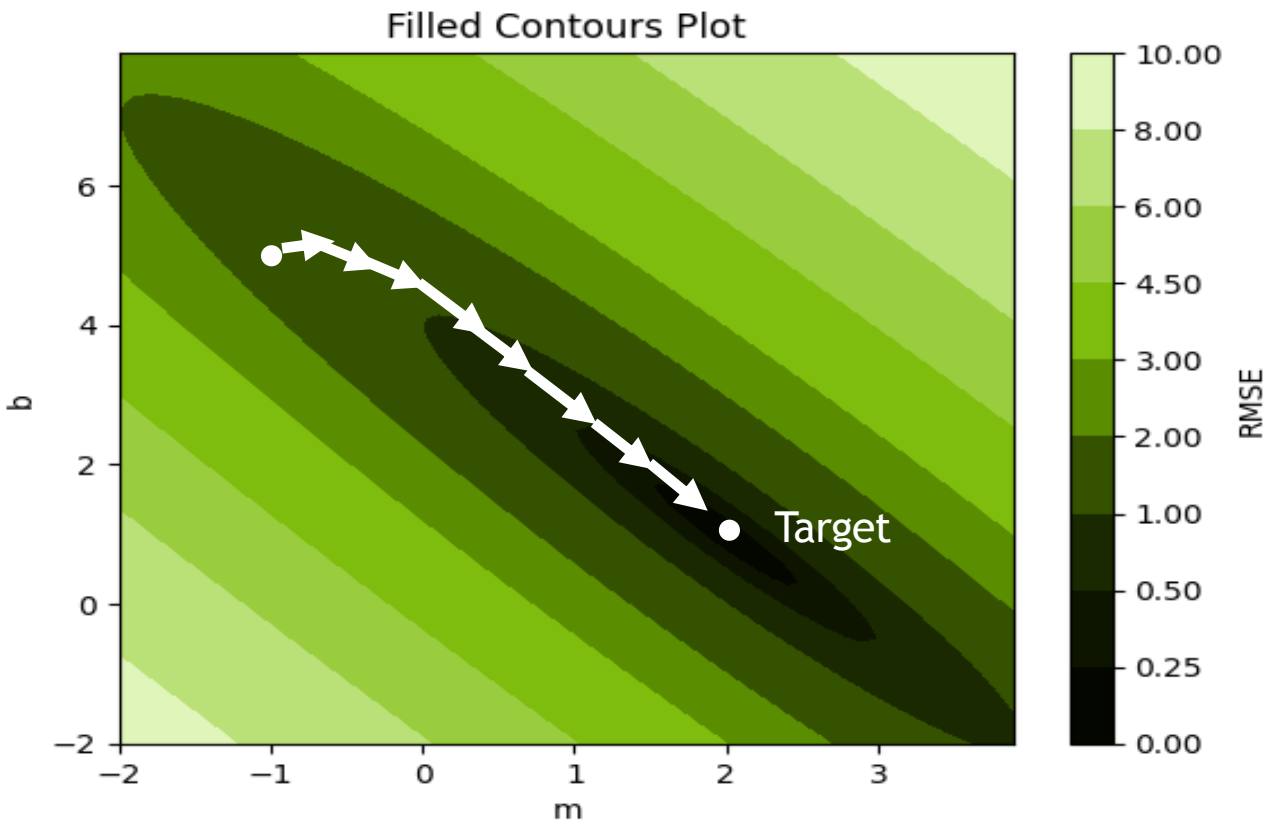
The Gradient	Which direction loss decreases the most
λ: The learning rate	How far to travel
Epoch	A model update with the full dataset
Batch	A sample of the full dataset
Step	An update to the weight parameters



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

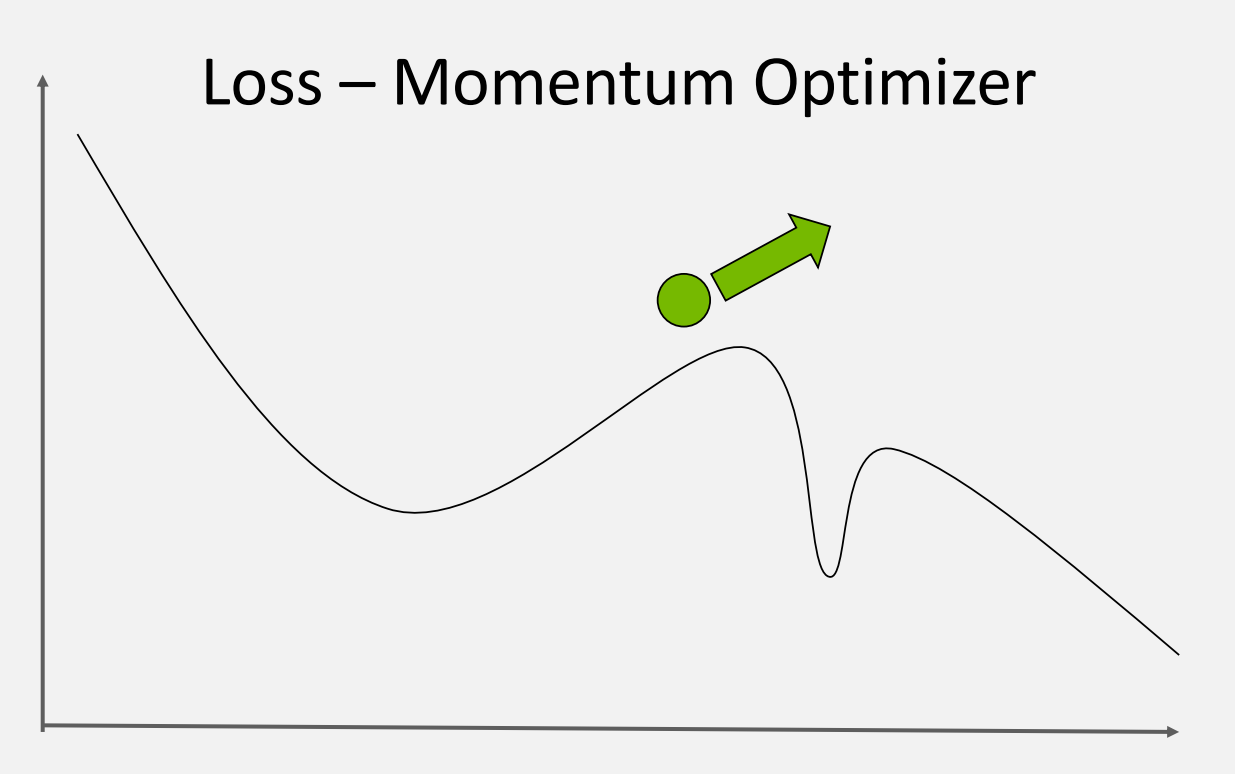
The Loss Curve

The Gradient	Which direction loss decreases the most
λ: The learning rate	How far to travel
Epoch	A model update with the full dataset
Batch	A sample of the full dataset
Step	An update to the weight parameters



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Optimizers



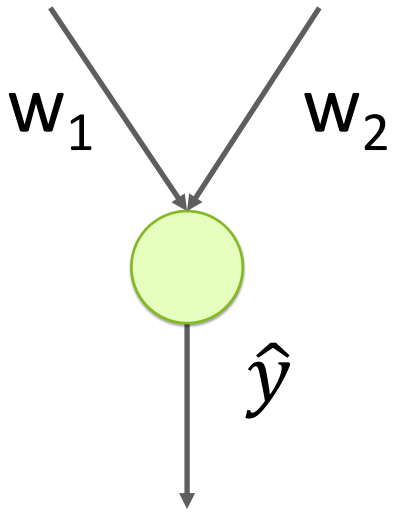
- Adam
- Adagrad
- RMSprop
- SGD

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

From Neuron to Network

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

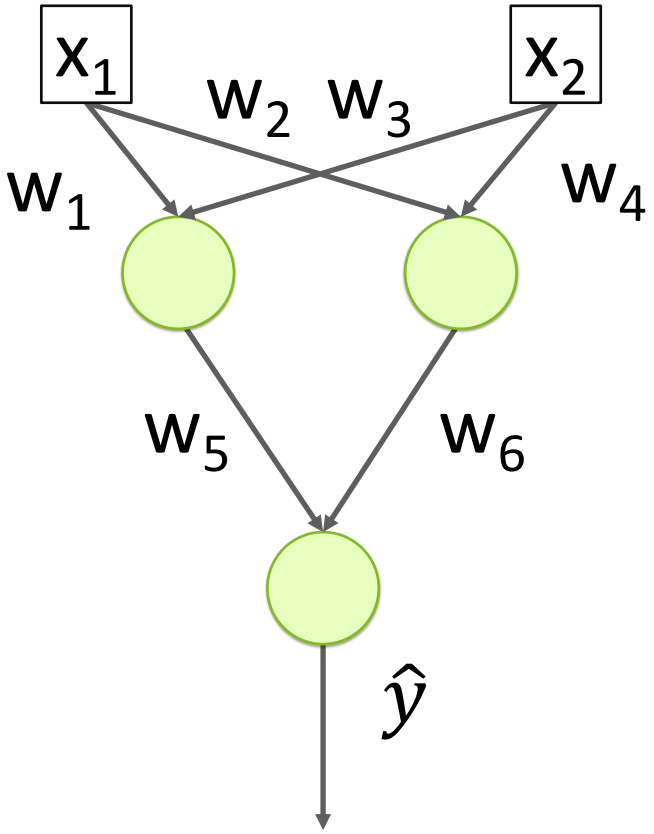
Building a Network



- Scales to more inputs

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

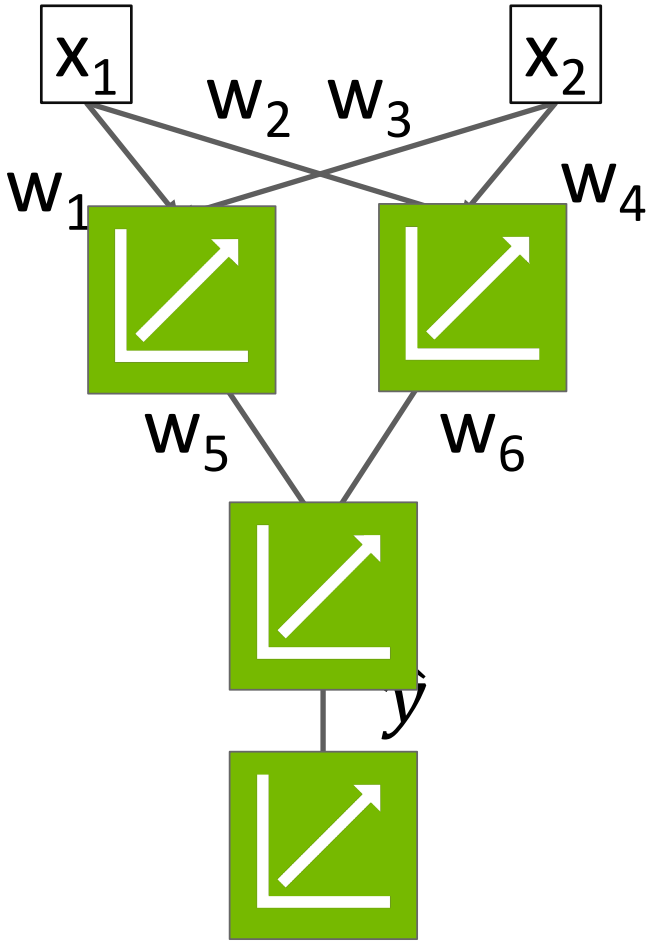
Building a Network



- Scales to more inputs
- Can chain neurons

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Building a Network



- Scales to more inputs
- Can chain neurons
- If all regressions are linear, then output will also be a linear regression

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Activation Functions

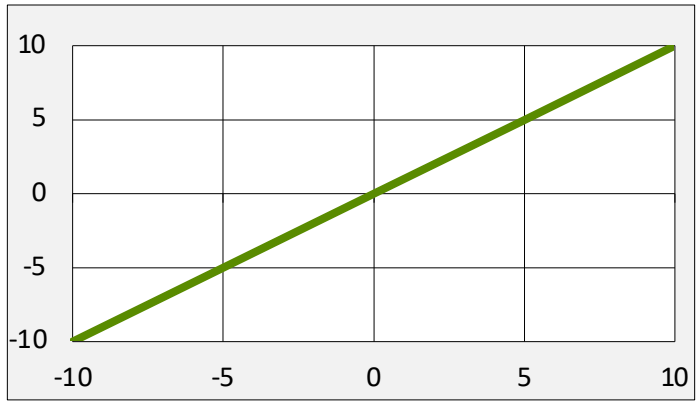
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Activation Functions

Linear

$$\hat{y} = wx + b$$

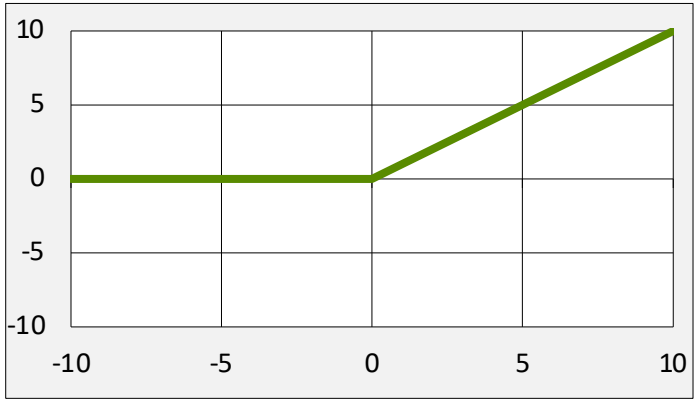
```
# Multiply each input
# with a weight (w) and
# add intercept (b)
y_hat = wx+b
```



ReLU

$$\hat{y} = \begin{cases} wx + b \; if \; wx + b > 0 \\ 0 \;\; otherwise \end{cases}$$

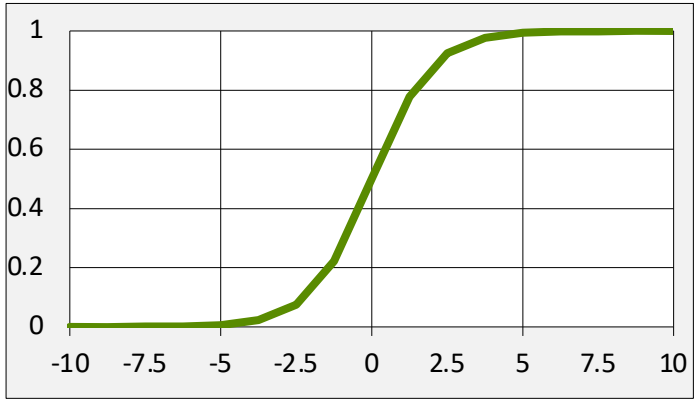
```
# Only return result
# if total is positive
linear = wx+b
y_hat = linear * (linear > 0)
```



Sigmoid

$$\hat{y} = \frac{1}{1 + e^{-(wx+b)}}$$

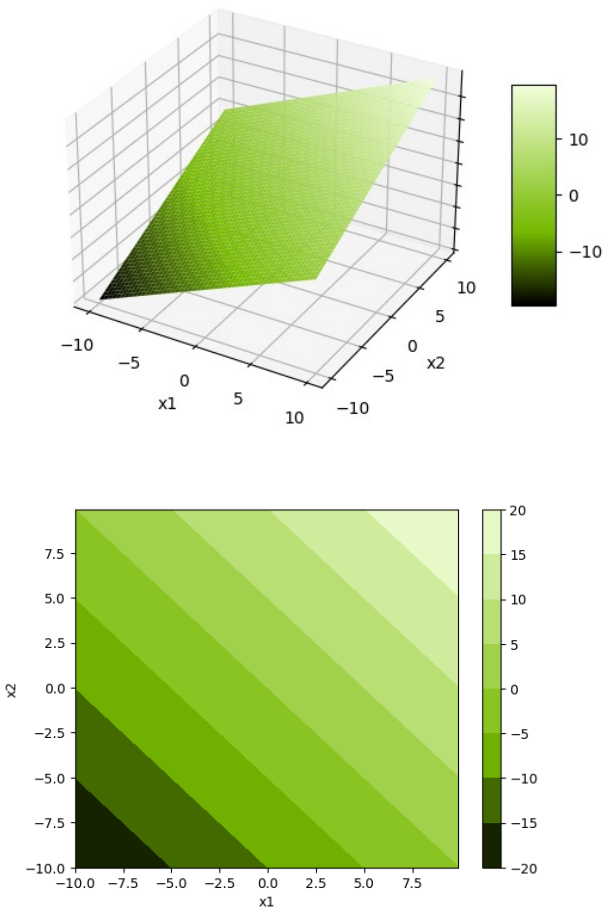
```
# Start with line
linear = wx + b
# Warp to - inf to 0
inf_to_zero = np.exp(-1 * linear)
# Squish to -1 to 1
y_hat = 1 / (1 + inf_to_zero)
```



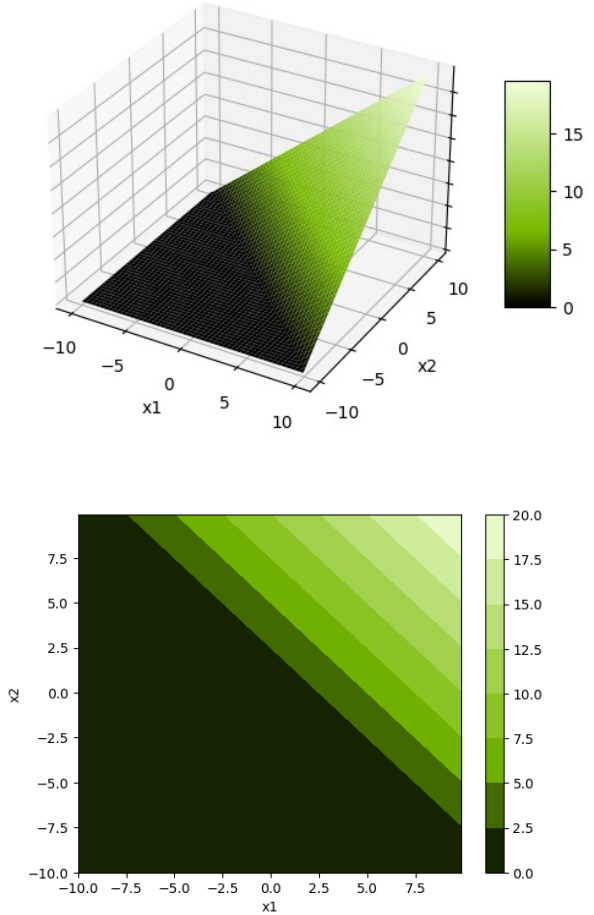
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Activation Functions

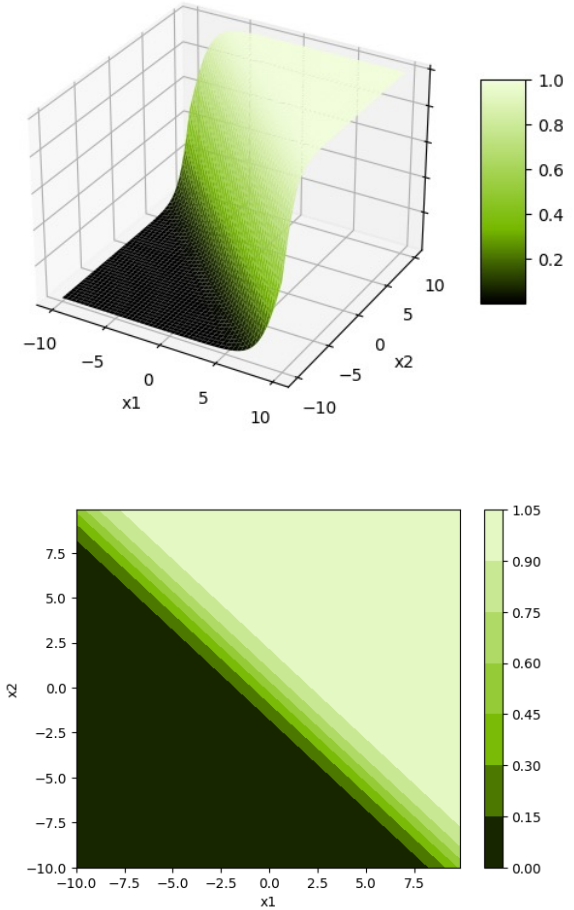
Linear



ReLU

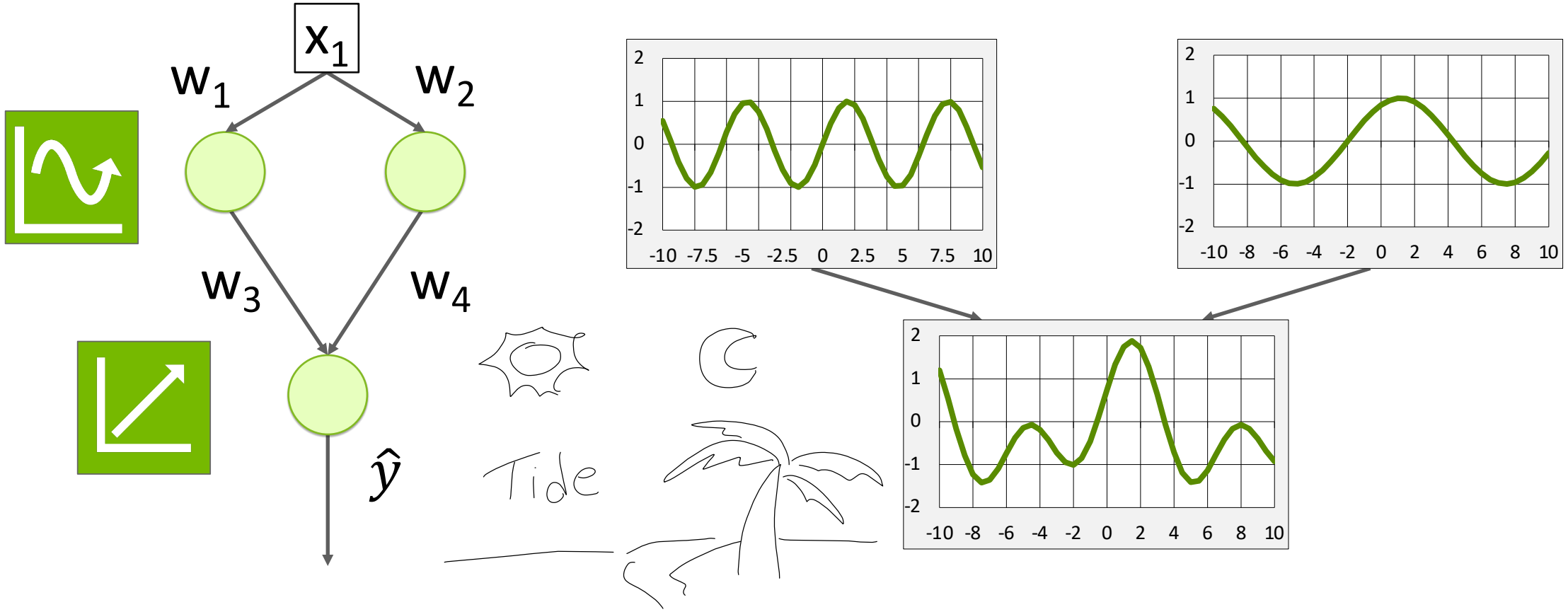


Sigmoid



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Activation Functions



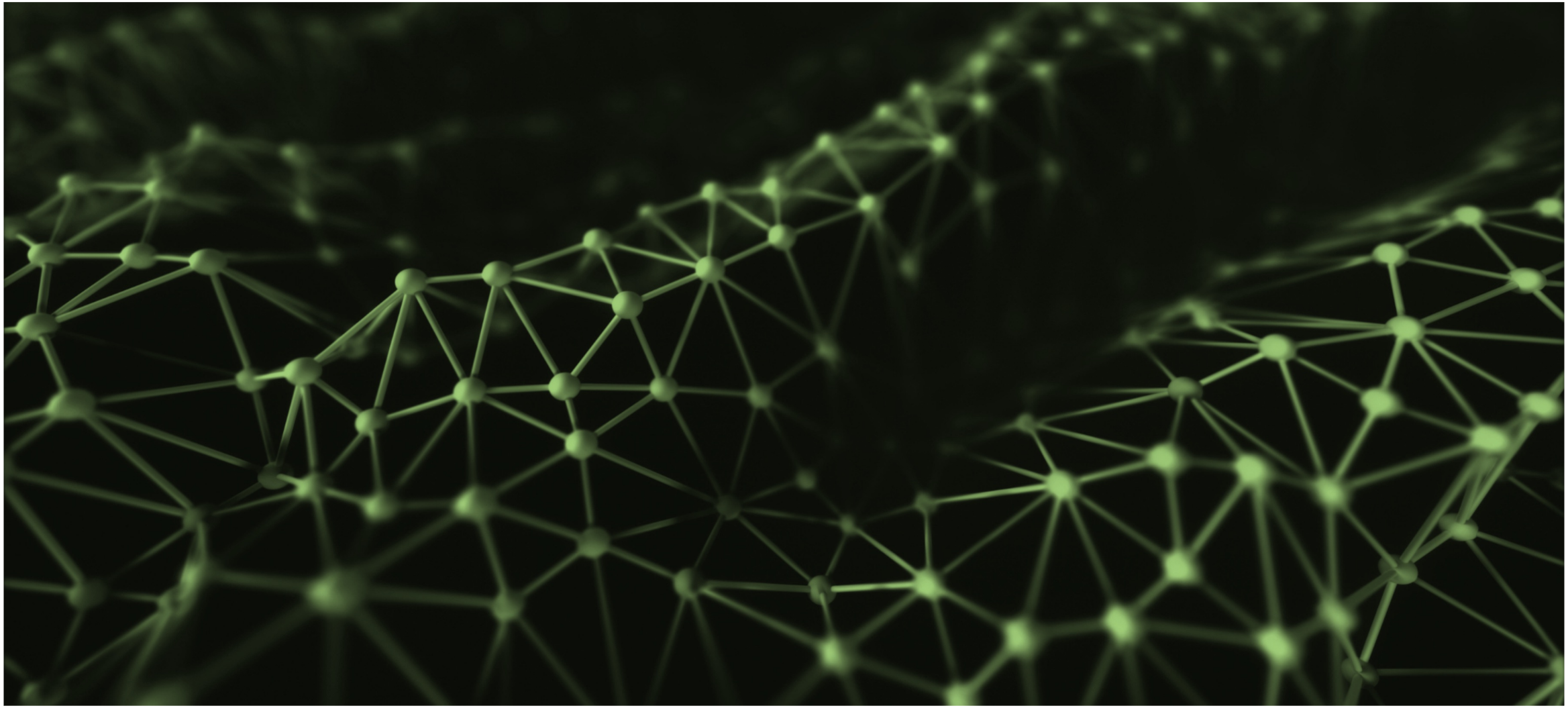
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Overfitting

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Overfitting

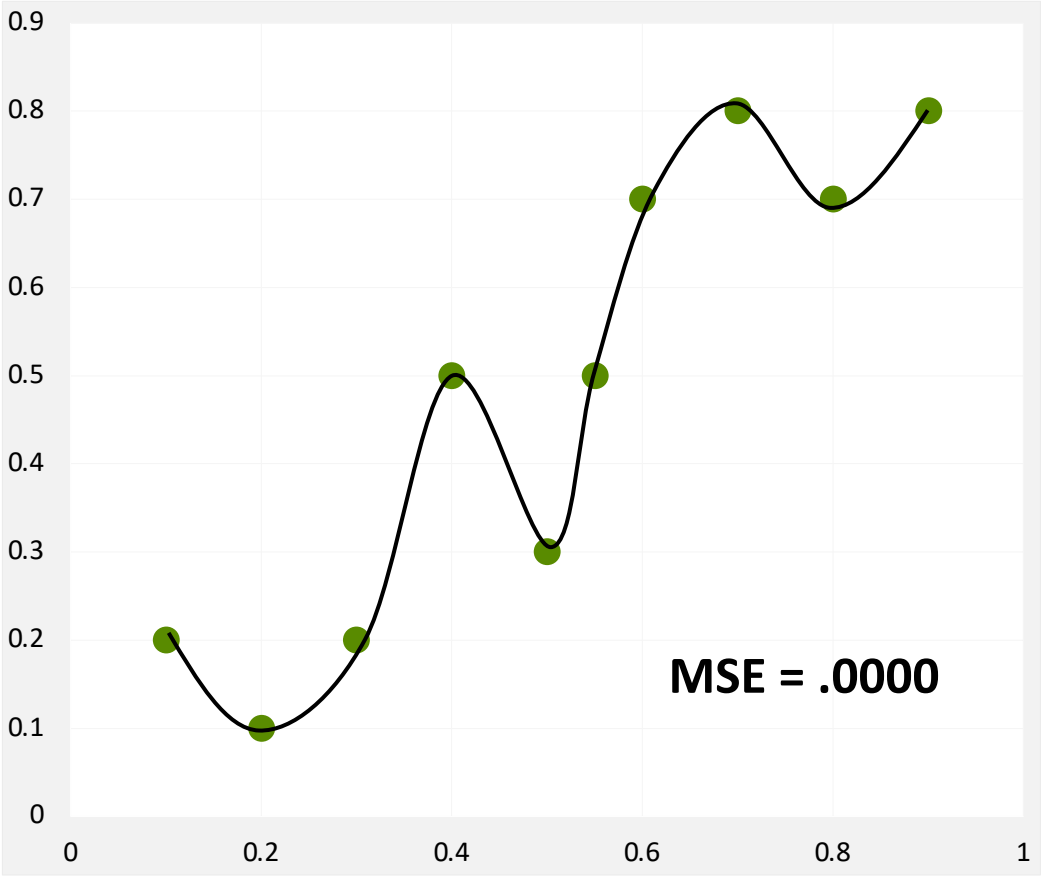
Why not have a super large neural network?

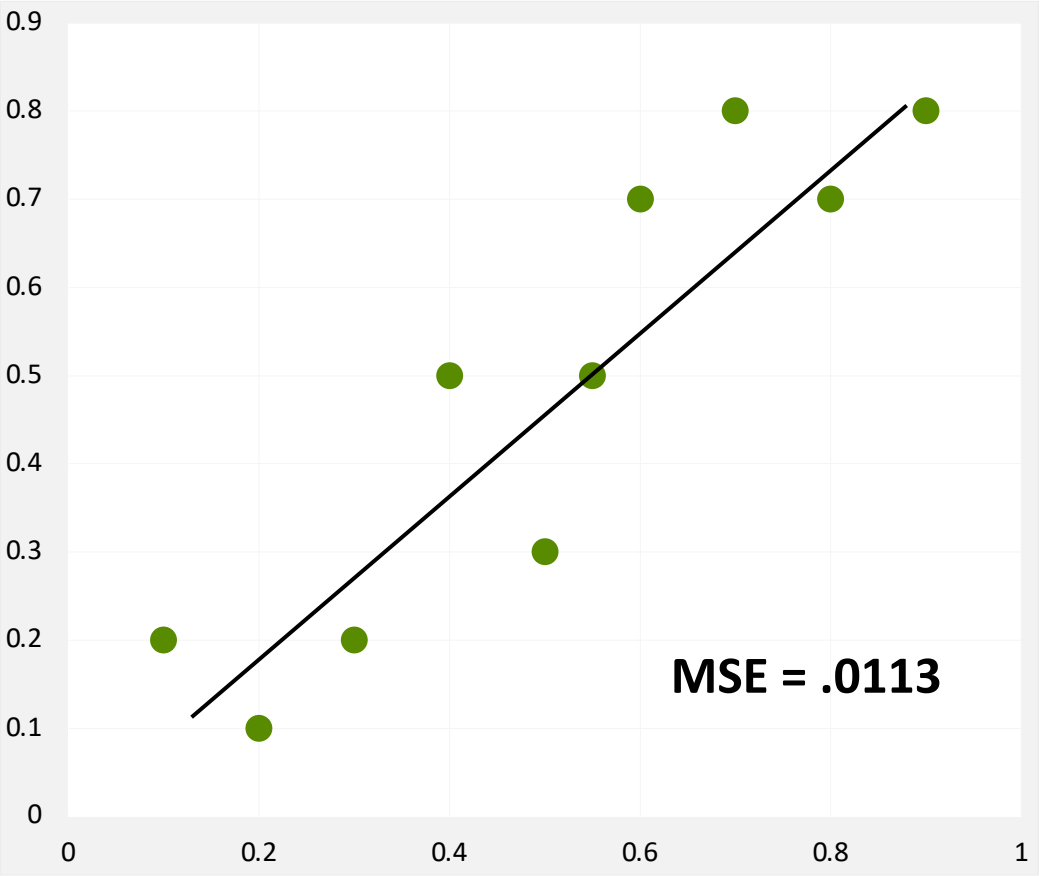


Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Overfitting

Which Trendline is Better?

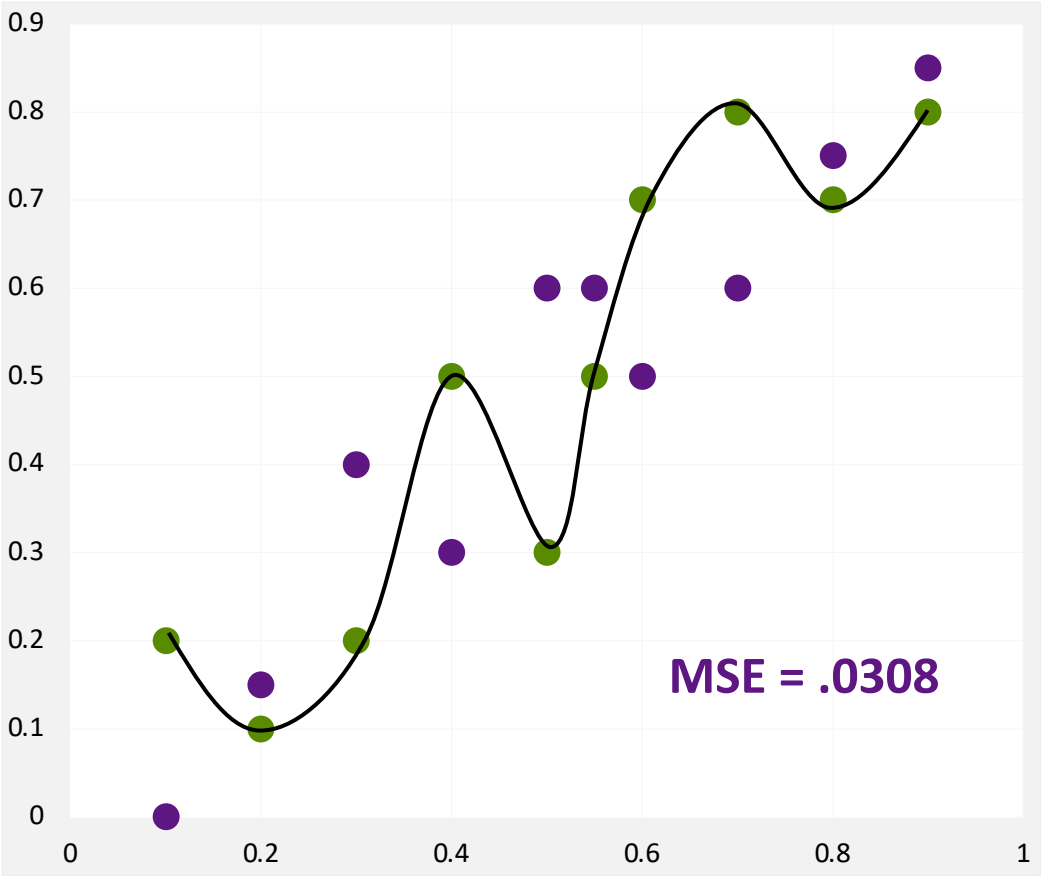


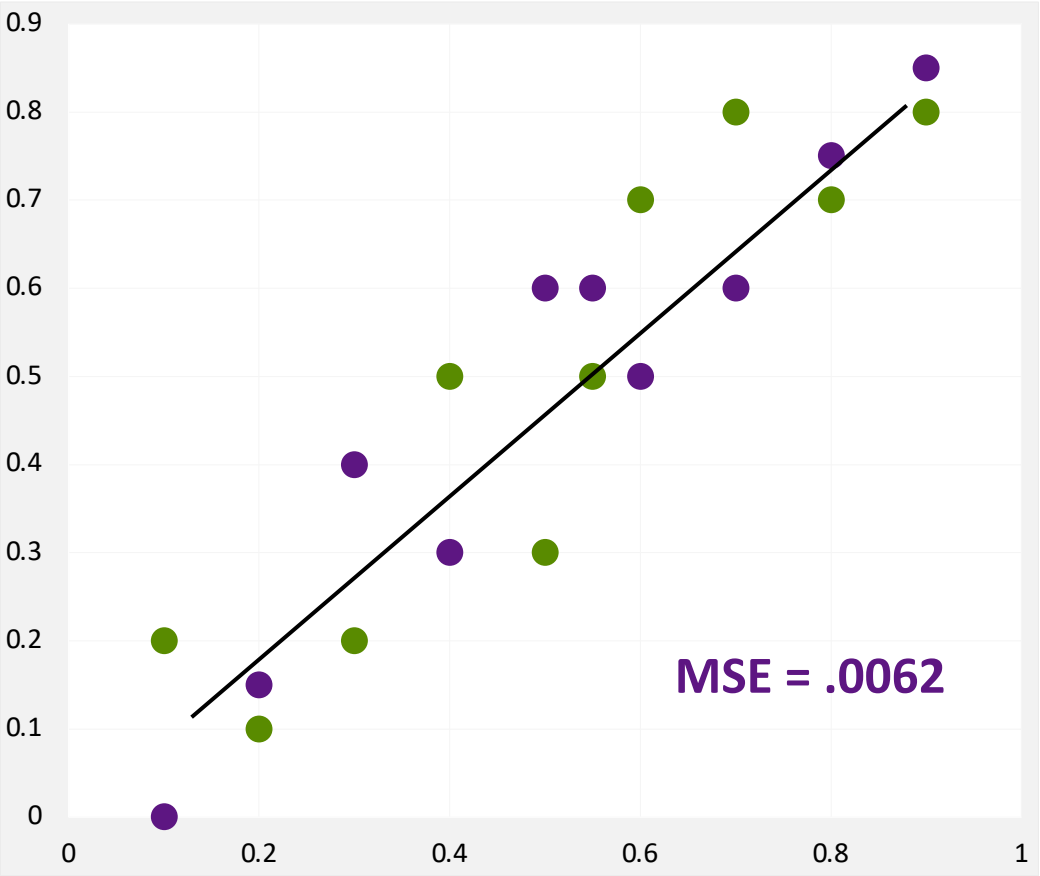


Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Overfitting

Which Trendline is Better?





Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Training vs Validation Data

Avoid memorization

Training data

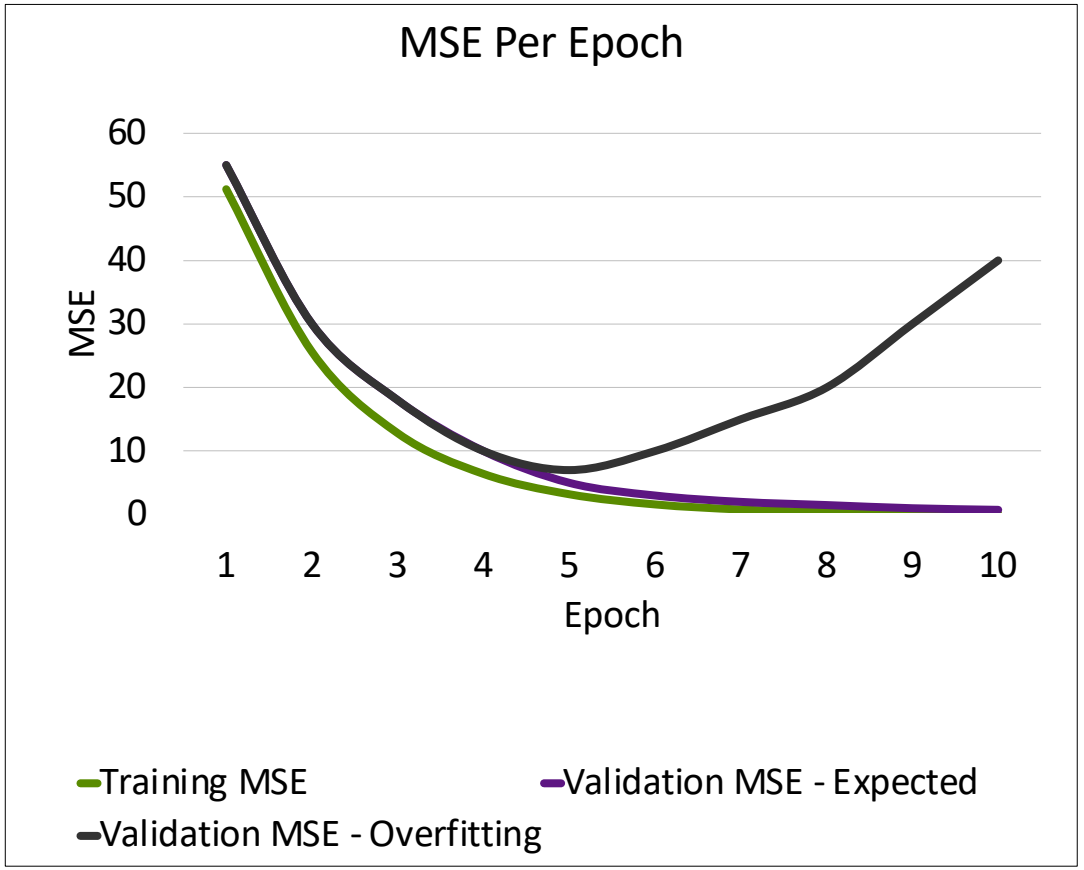
- Core dataset for the model to learn on

Validation data

- New data for model to see if it truly understands (can generalize)

Overfitting

- When model performs well on the training data, but not the validation data (evidence of memorization)
- Ideally the accuracy and loss should be similar between both datasets

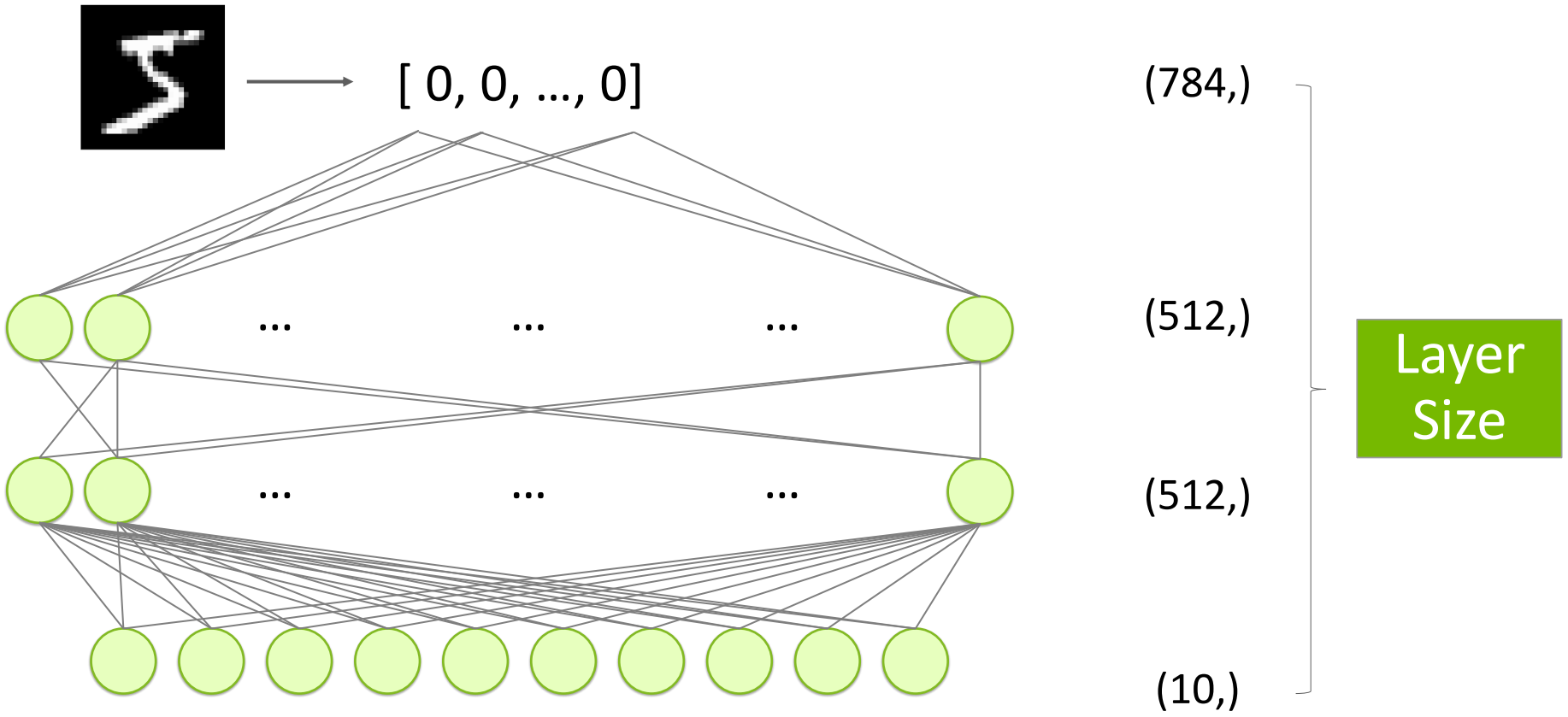


Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

From Regression to Classification

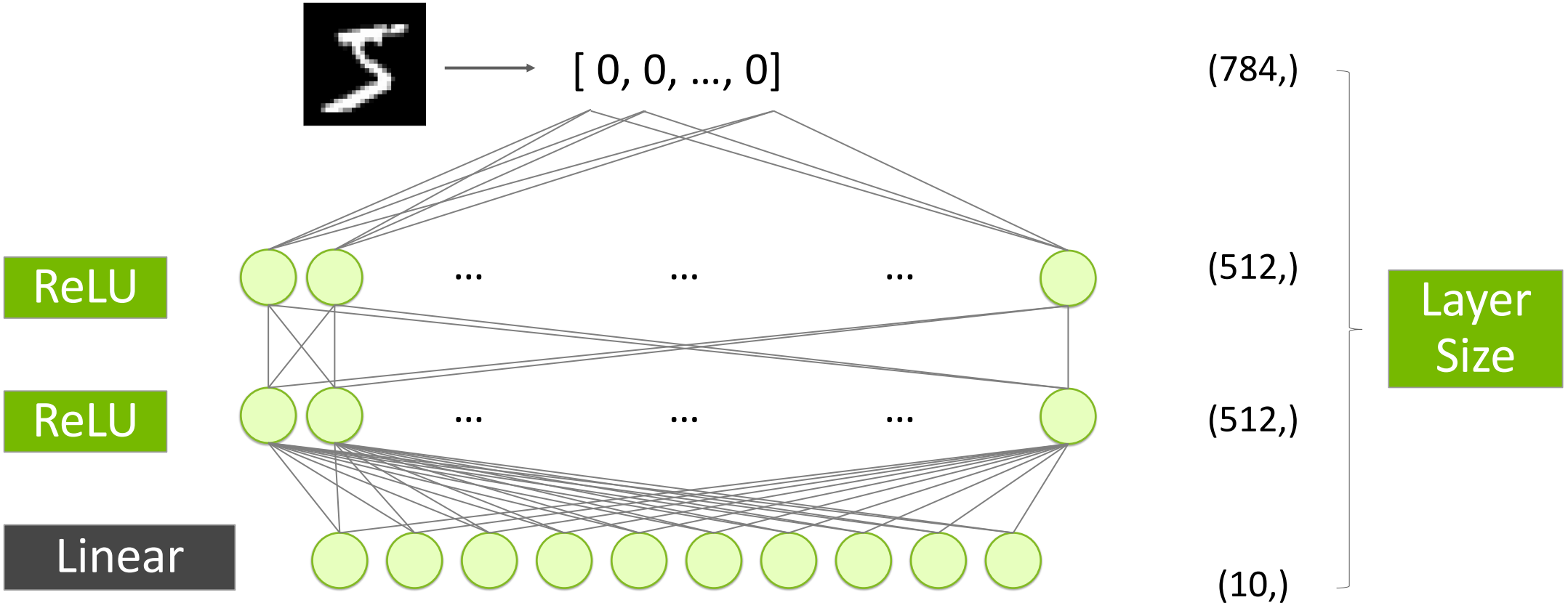
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

An MNIST Model



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

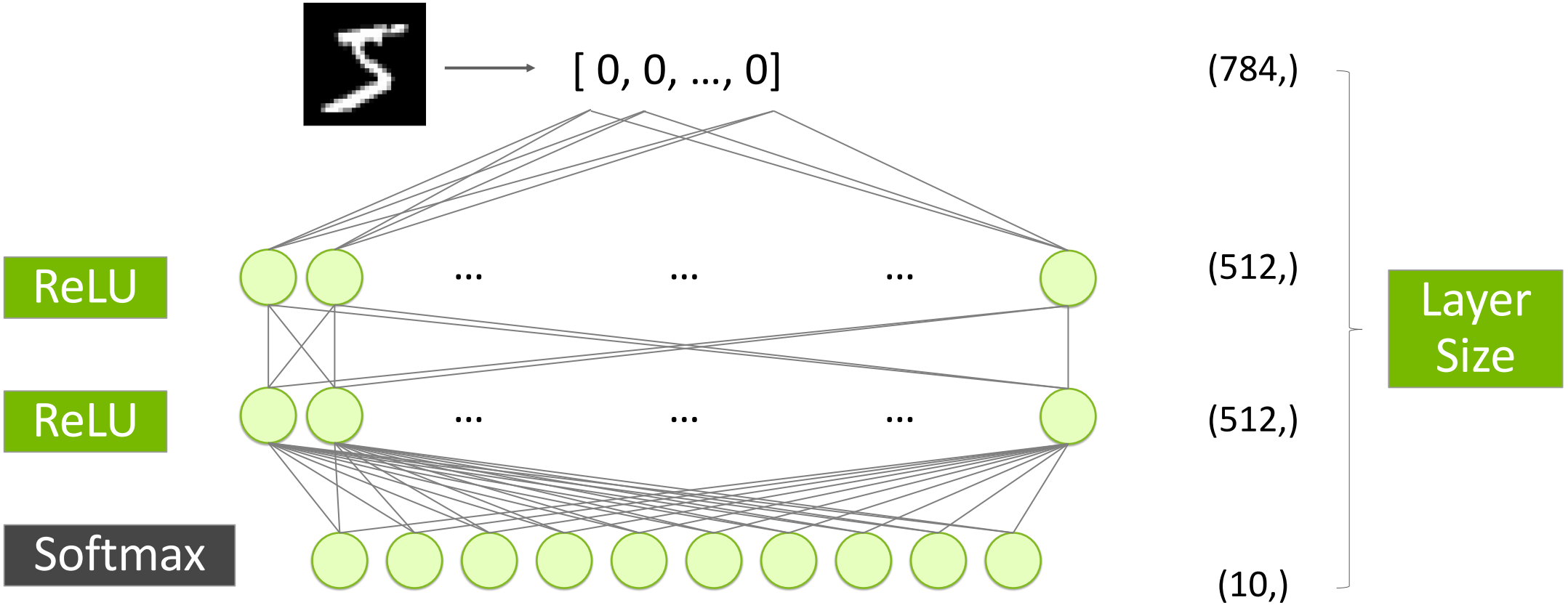
An MNIST Model
During Prediction



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

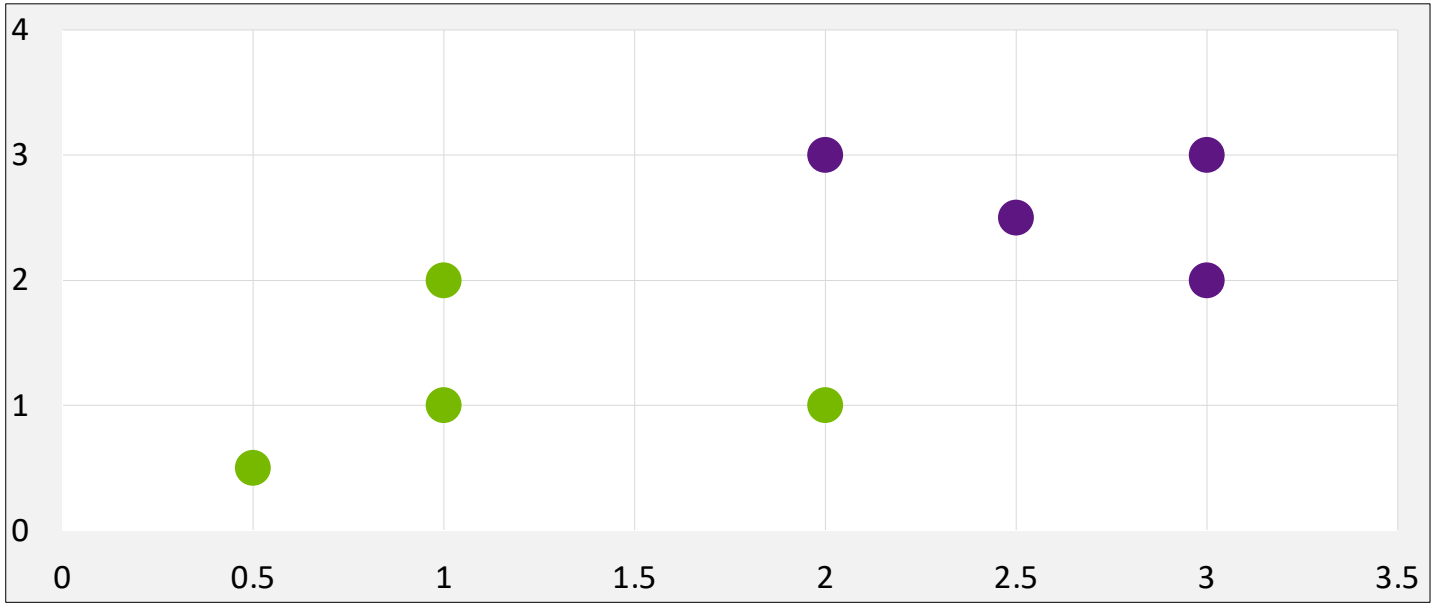
An MNIST Model

During Training



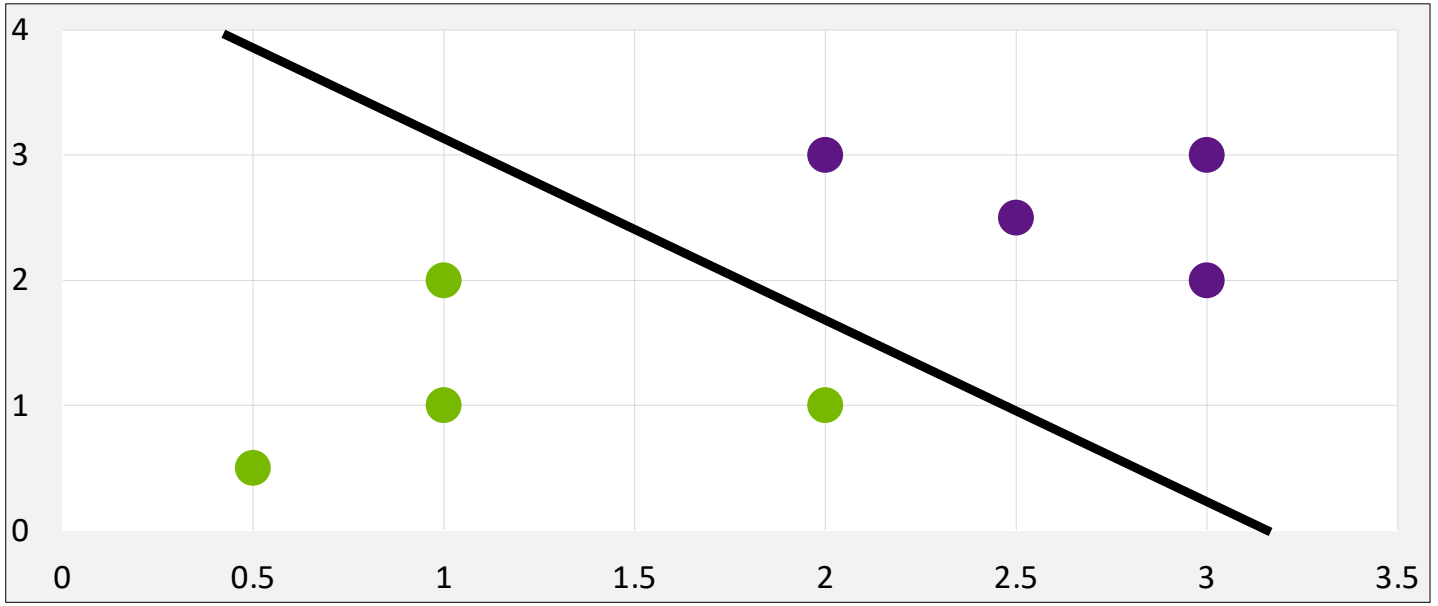
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

RMSE For Probabilities?



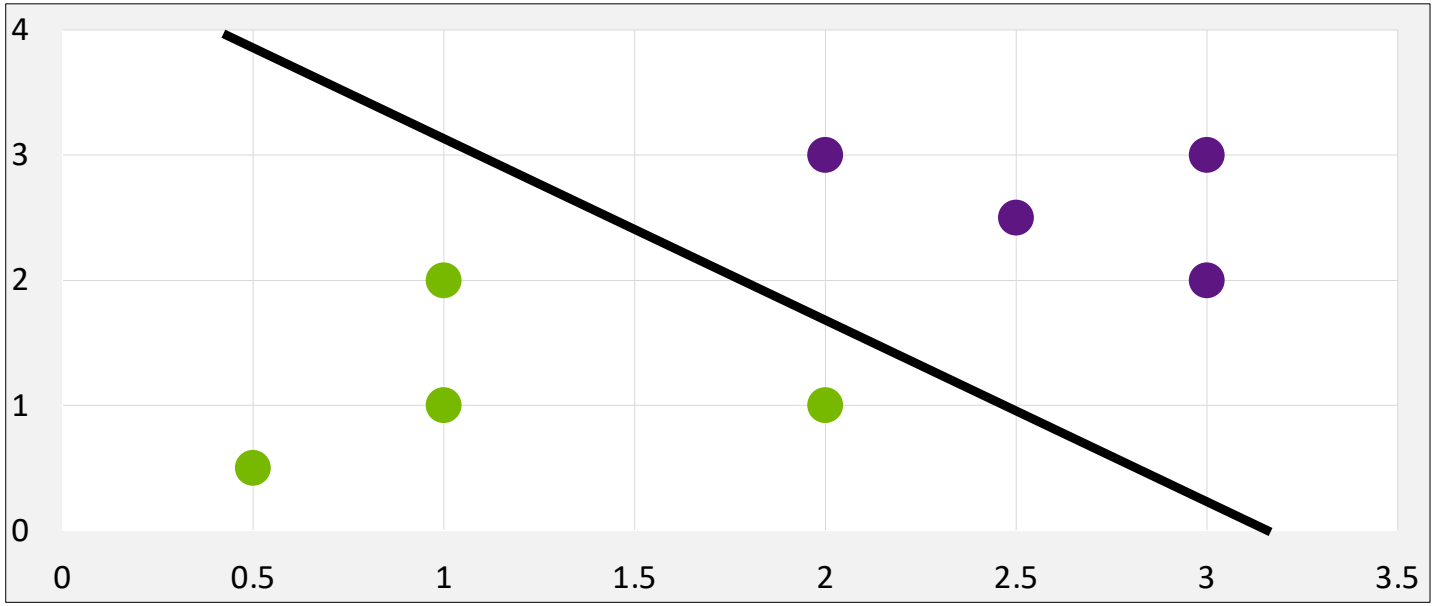
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

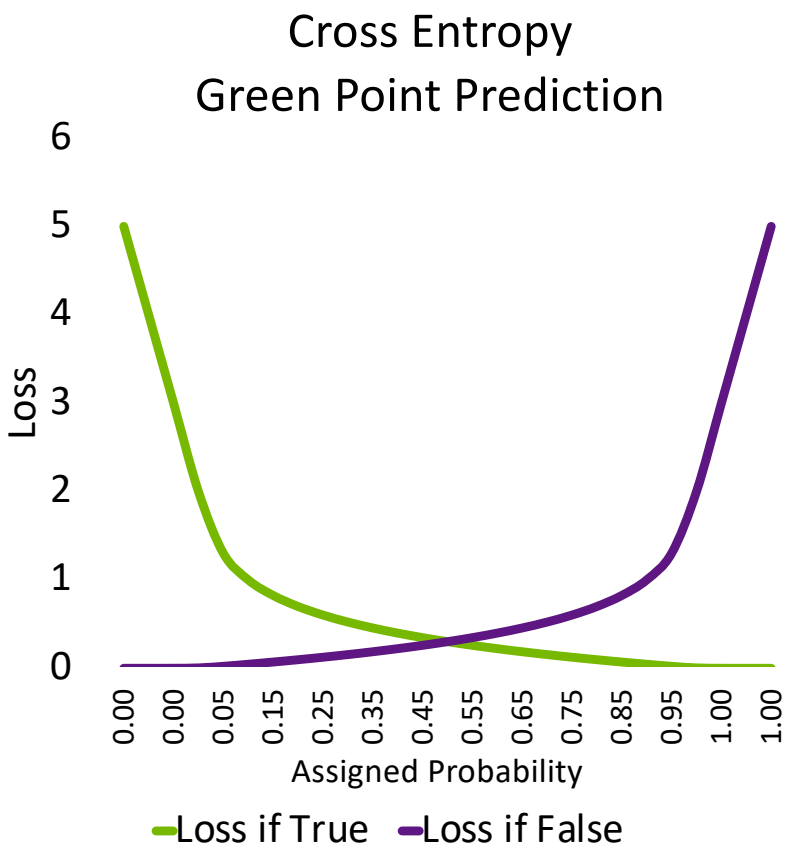
RMSE For Probabilities?



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

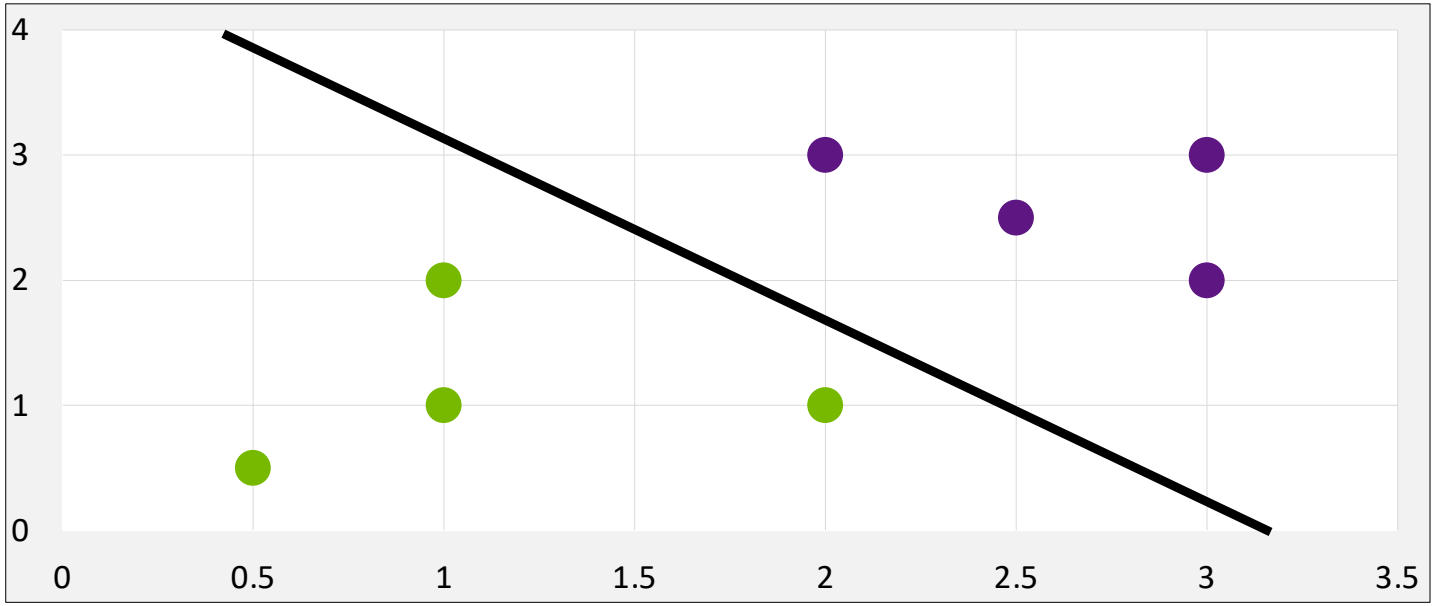
Cross Entropy

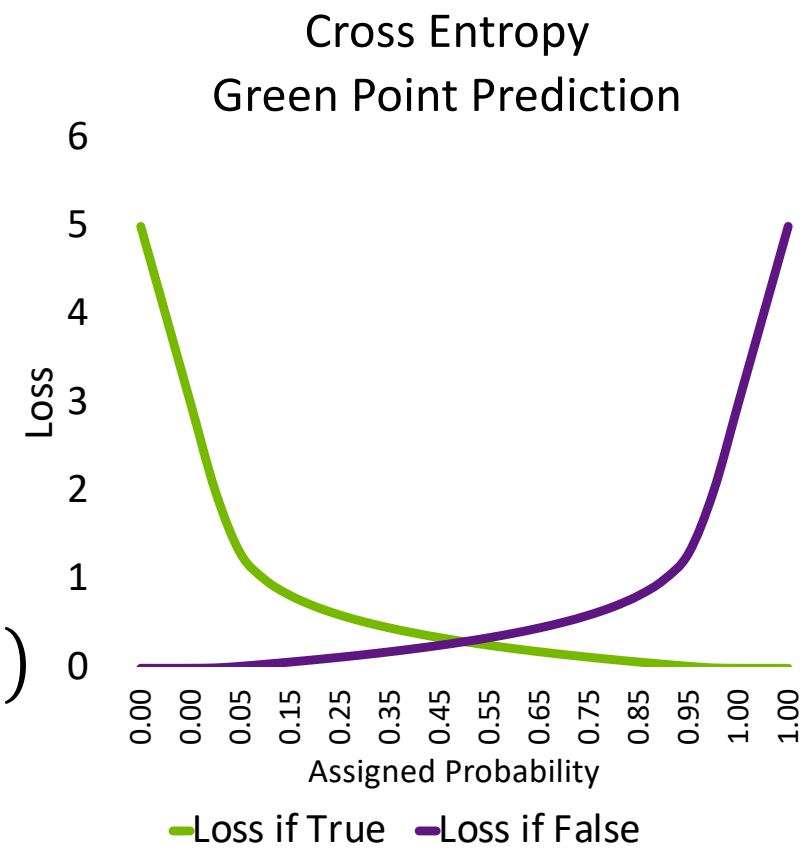




Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Cross Entropy





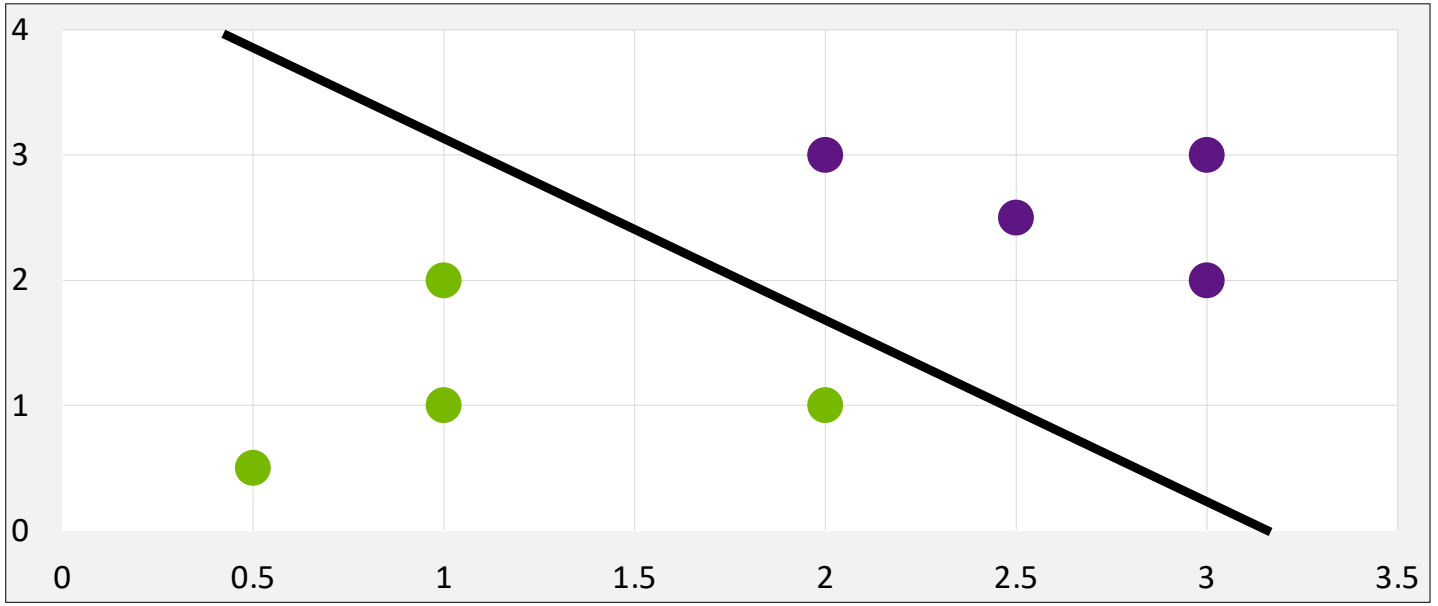
$$Loss = -\big((t(x) \cdot \log(p(x)) + \big(1 - t(x)\big) \cdot \log(1 - p(x))\big)$$

$$t(x) = target\ (0\ if\ False, 1\ if\ True)$$

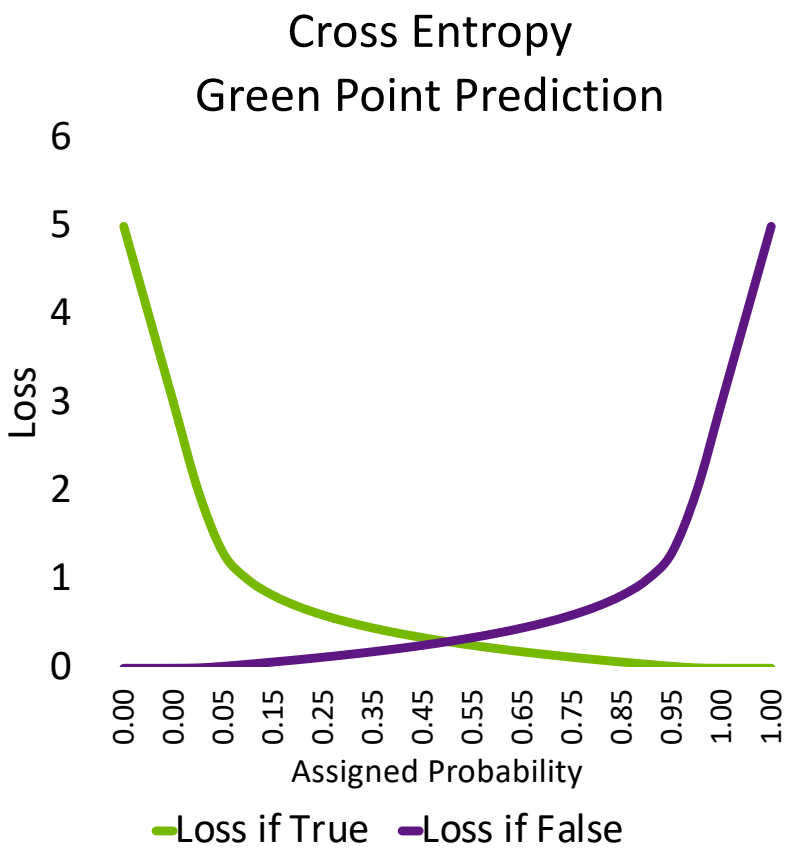
$$p(x) = probability\ prediction\ of\ point\ x$$

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Cross Entropy




```python
def cross_entropy(y_hat, y_actual):
    """Infinite error for misplaced confidence."""
    loss = log(y_hat) if y_actual else log(1-y_hat)
    return -1*loss
```



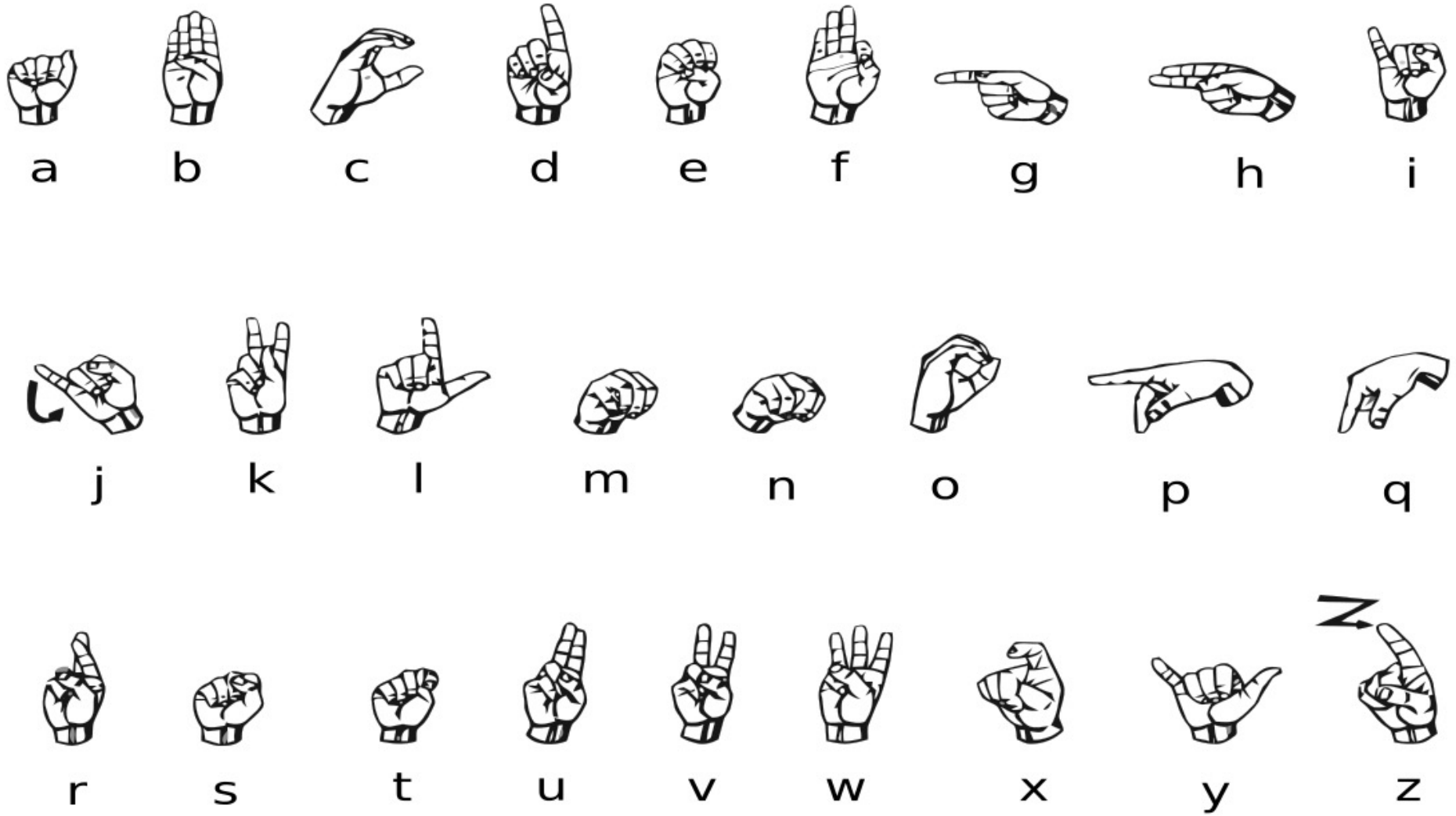
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Bringing it Together

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Next Exercise

The American Sign Language Alphabet



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Let’s go!

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Appendix: Gradient Descent

Helping the Computer Cheat Calculus

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Learning from Error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(y-(mx+b))^2$$

$$MSE = \frac{1}{2}((3 - (m(1)+b))^2 + (5 - (m(2)+b))^2)$$

$$\frac{\partial MSE}{\partial m} = 5m + 3b - 13 \qquad \frac{\partial MSE}{\partial b} = 3m + 2b - 8$$

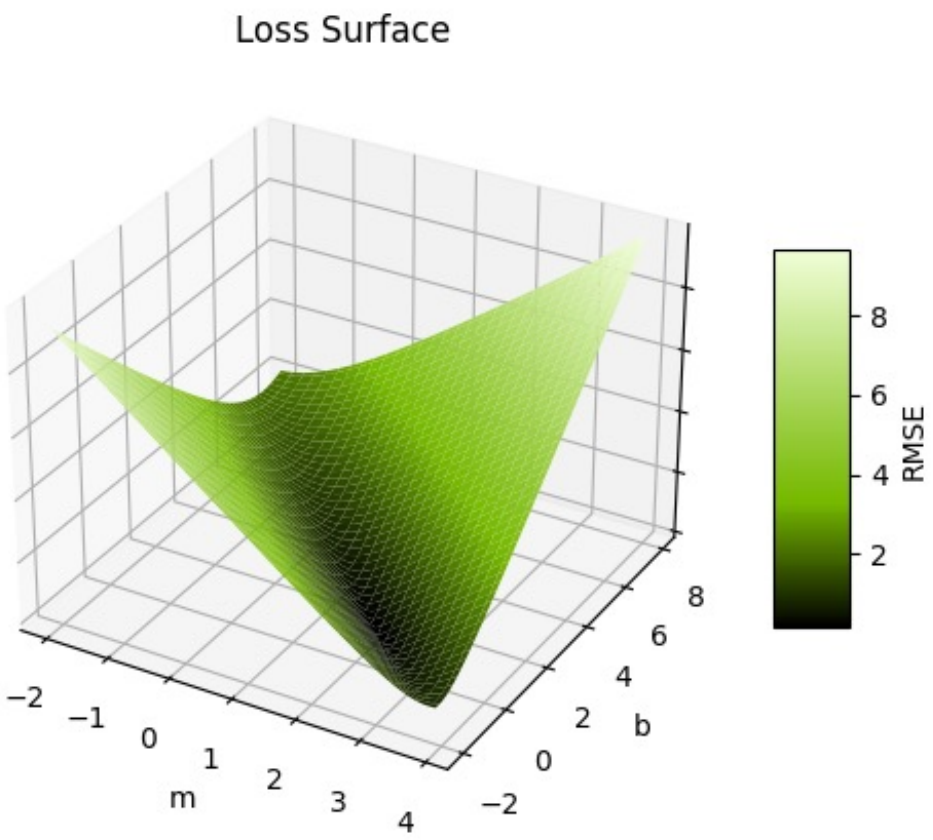
$$\frac{\partial MSE}{\partial m} = -3 \qquad \frac{\partial MSE}{\partial b} = -1$$

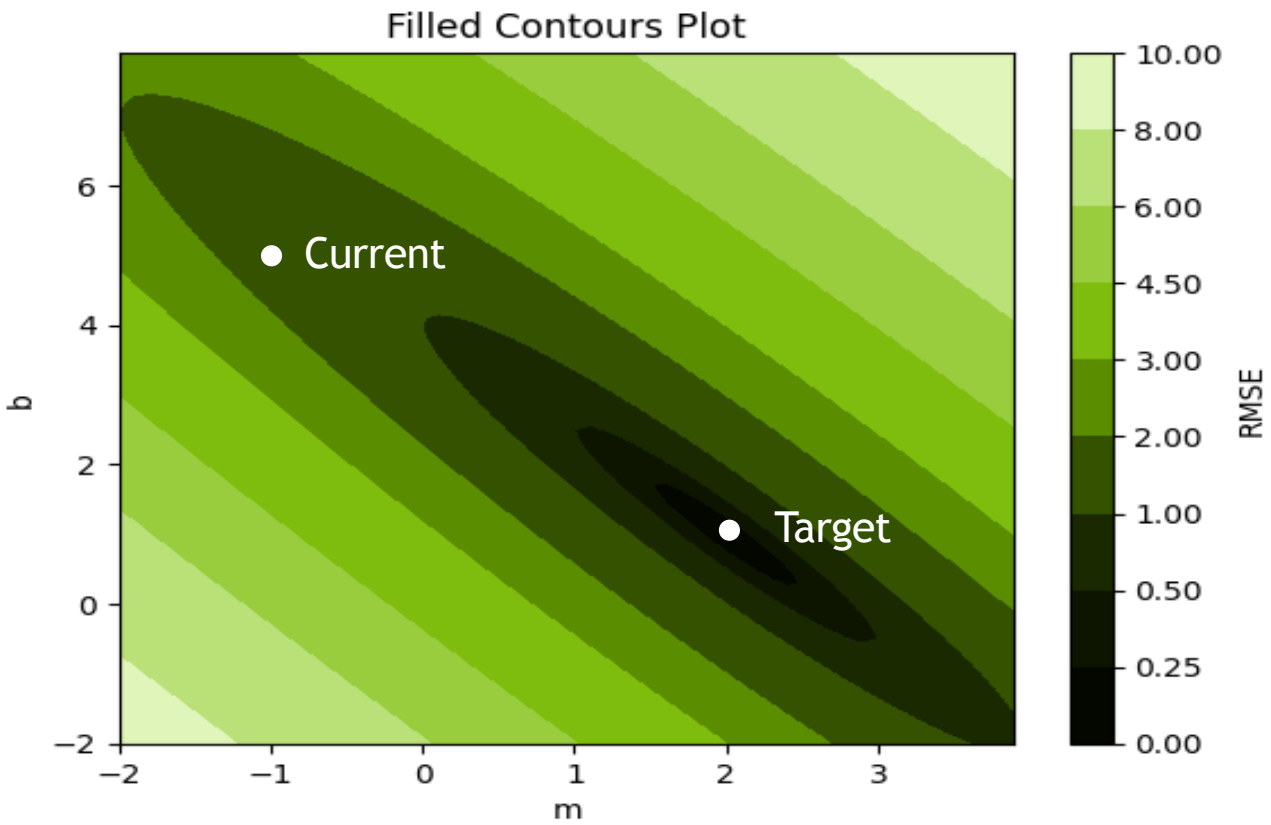
$$m = -1$$

$$\mathrm{b} = 5$$

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve

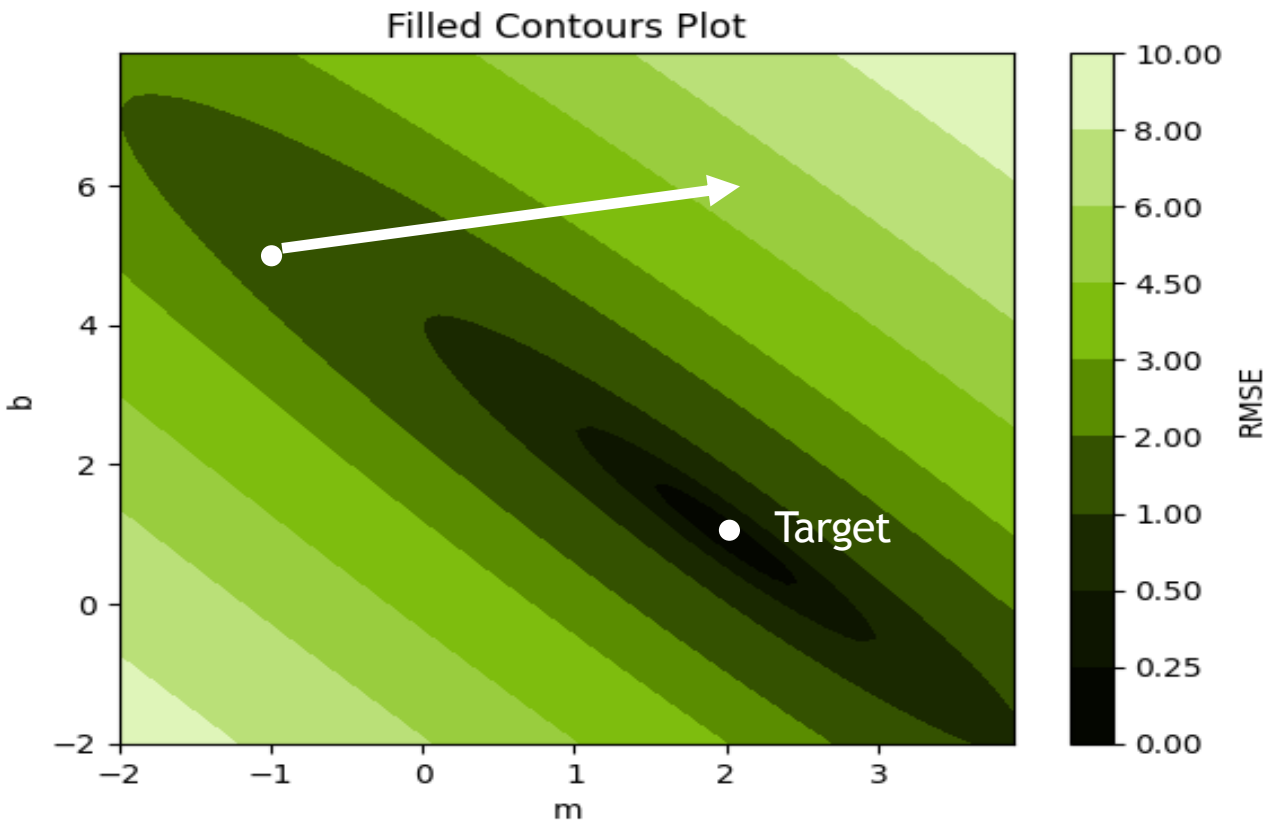




Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve

$$\frac{\partial MSE}{\partial m} = -3 \qquad \frac{\partial MSE}{\partial b} = -1$$



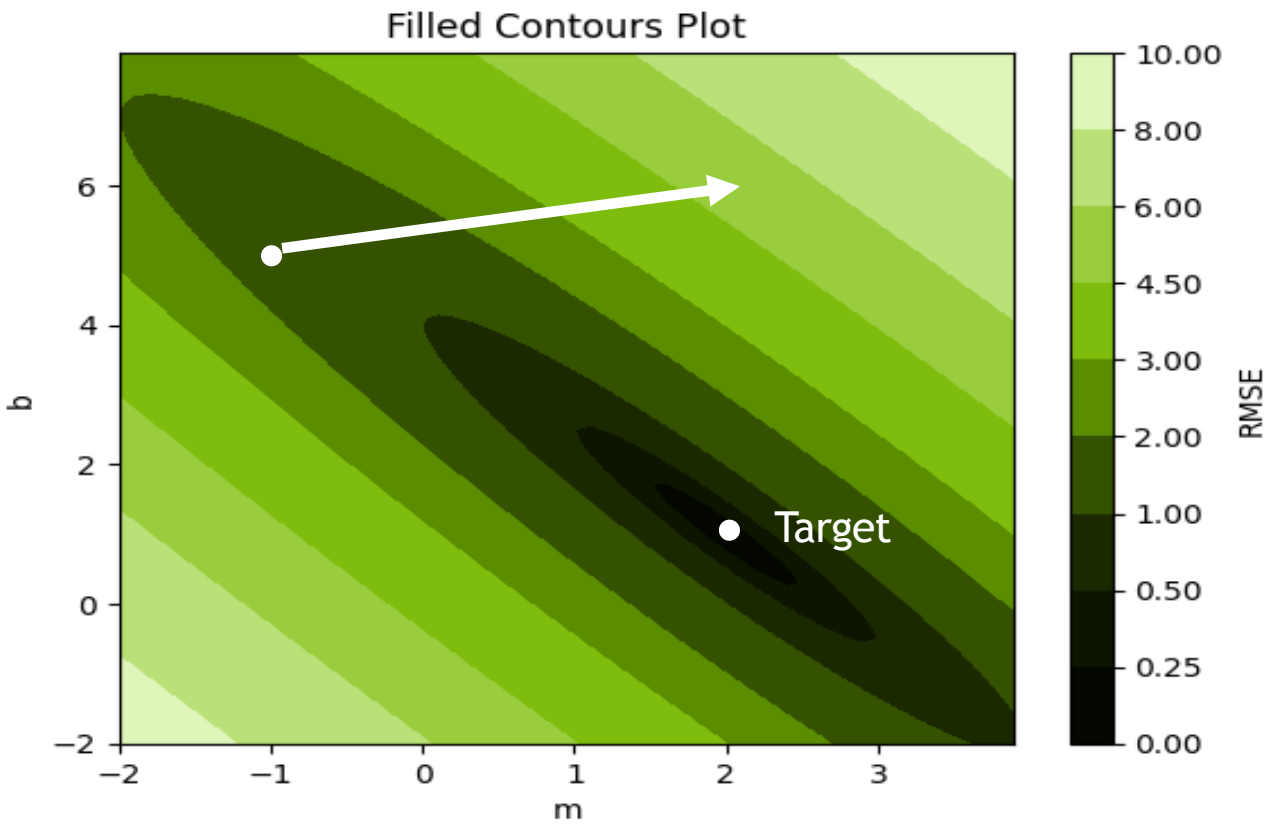
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve

$$\frac{\partial MSE}{\partial m} = -3 \qquad \frac{\partial MSE}{\partial b} = -1$$

$$\mathrm{m} := \mathrm{m} - \lambda \frac{\partial MSE}{\partial m}$$

$$\mathrm{b} := b - \lambda \frac{\partial MSE}{\partial b}$$



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

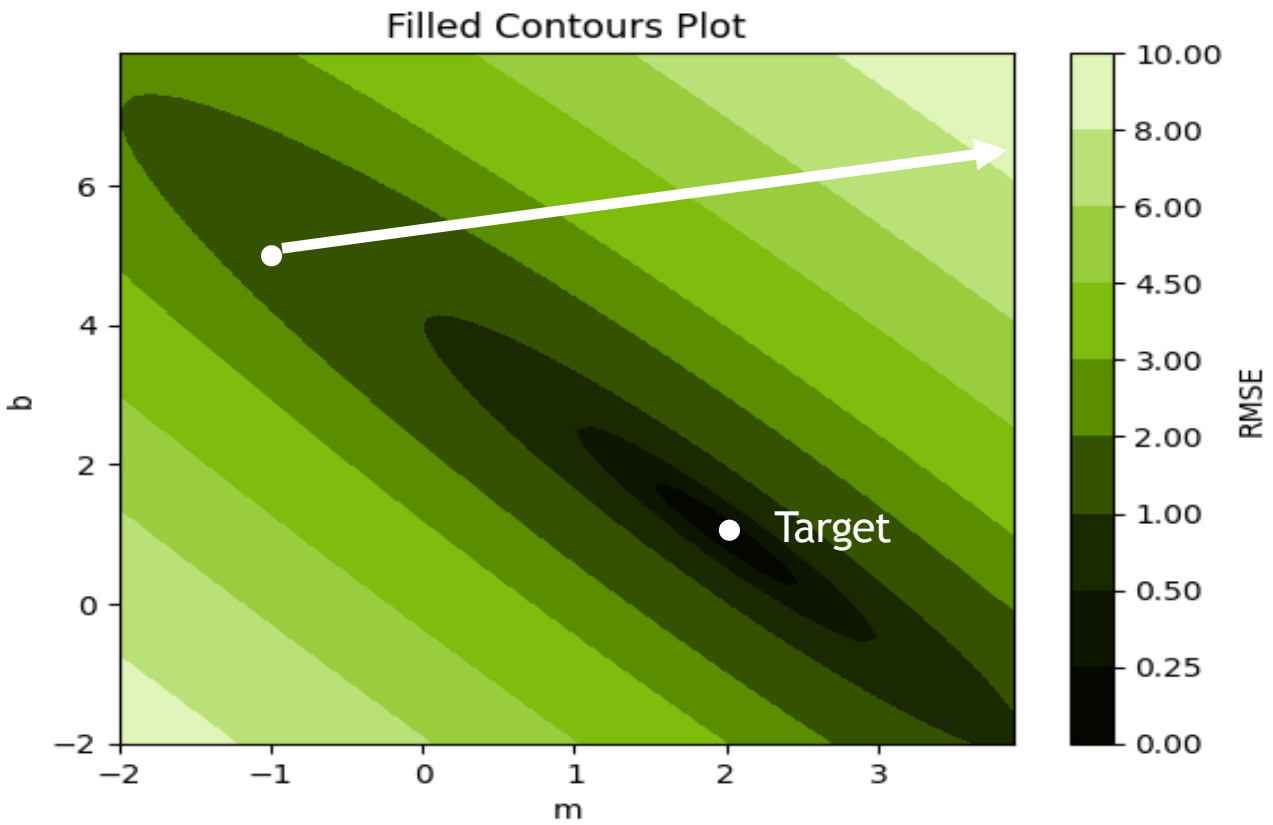
The Loss Curve

$$\frac{\partial MSE}{\partial m} = -3 \qquad \frac{\partial MSE}{\partial b} = -1$$

$$\mathrm{m} := \mathrm{m} - \lambda \frac{\partial MSE}{\partial m}$$

$$\lambda = 2$$

$$\mathrm{b} := b - \lambda \frac{\partial MSE}{\partial b}$$



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

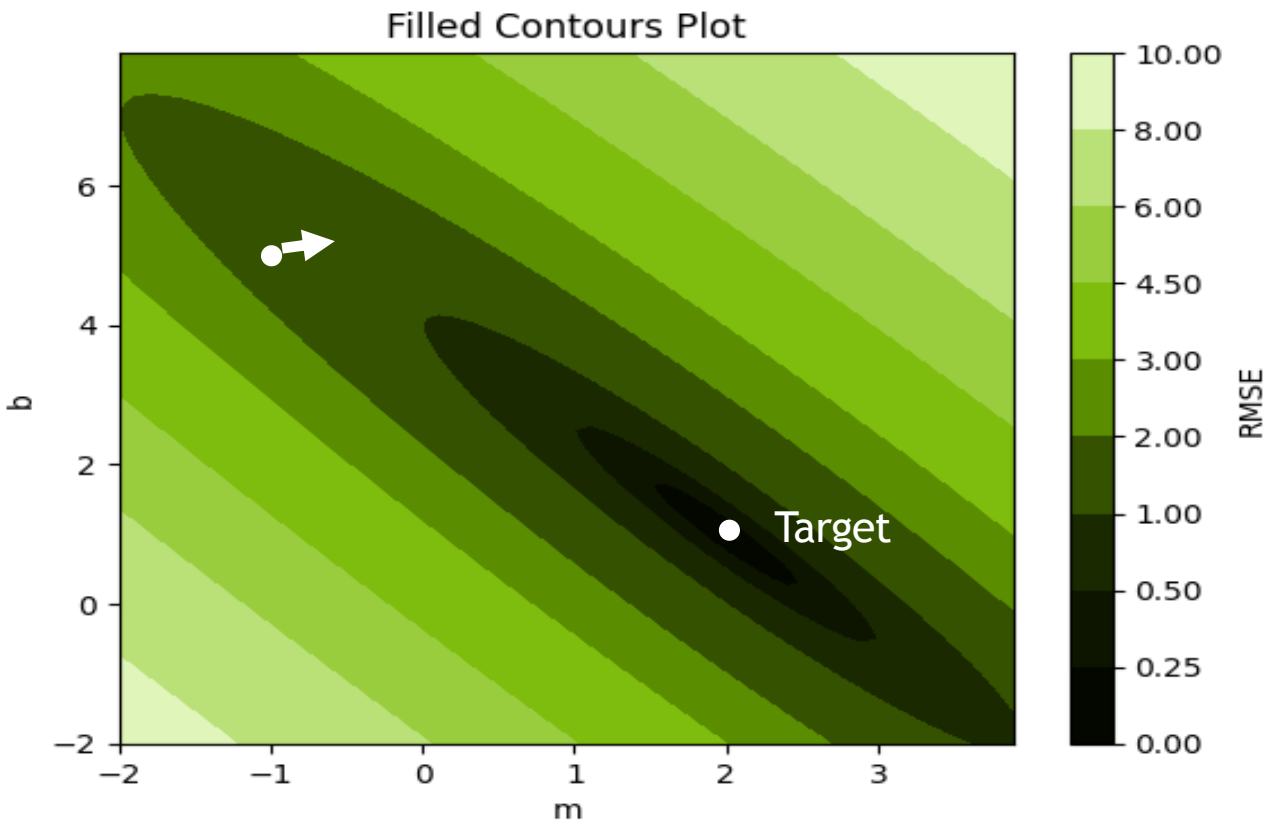
The Loss Curve

$$\frac{\partial MSE}{\partial m} = -3 \qquad \frac{\partial MSE}{\partial b} = -1$$

$$\mathrm{m} := \mathrm{m} - \lambda \frac{\partial MSE}{\partial m}$$

$$\lambda = .1$$

$$\mathrm{b} := b - \lambda \frac{\partial MSE}{\partial b}$$



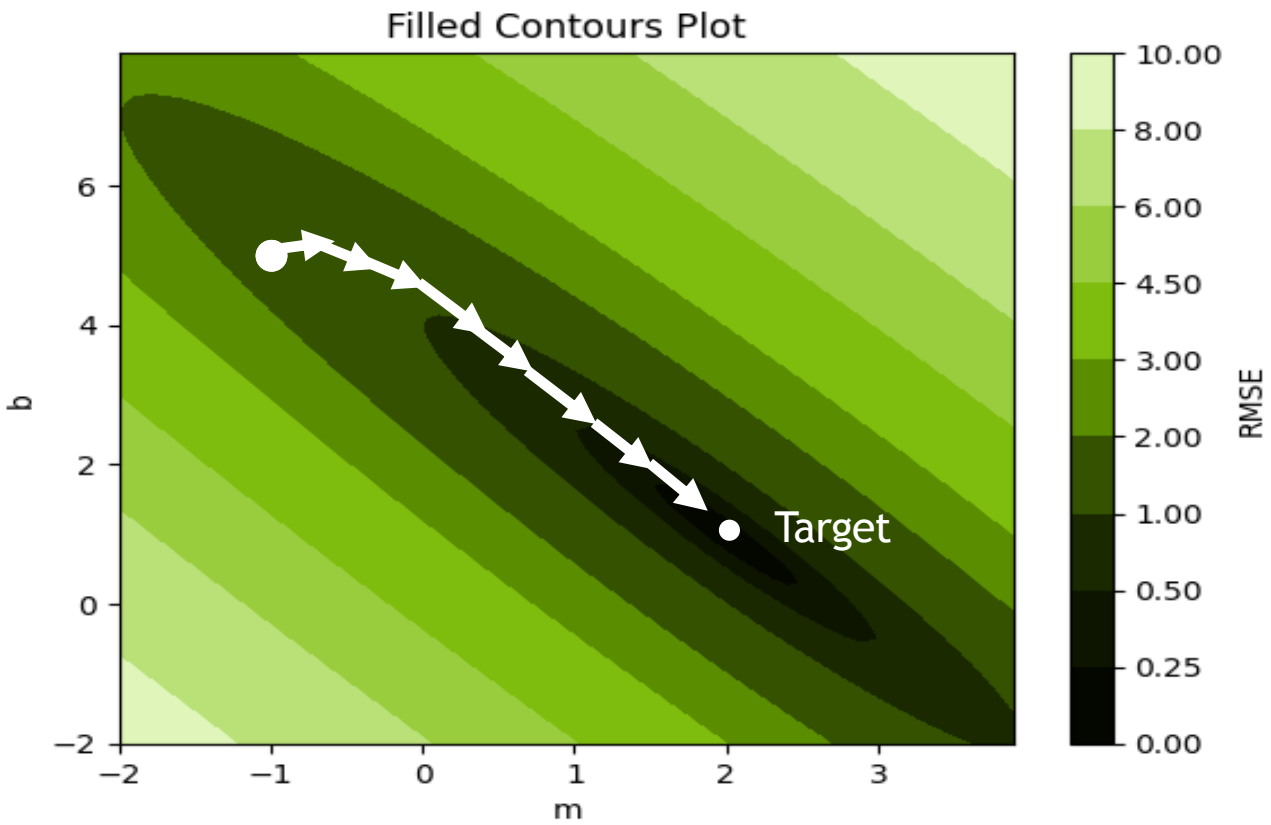
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Loss Curve

$$\lambda = .1$$

$$m := -1 + 3\,\lambda = -0.7$$

$$b := 5 + \lambda = 5.1$$



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

2. Image Classification of an American Sign Language Dataset

In this section, we will perform the data preparation, model creation, and model training steps we observed in the last section using a different dataset: images of hands making letters in American Sign Language.

2.1 Objectives

- Prepare image data for training
- Create and compile a simple model for image classification
- Train an image classification model and observe the results

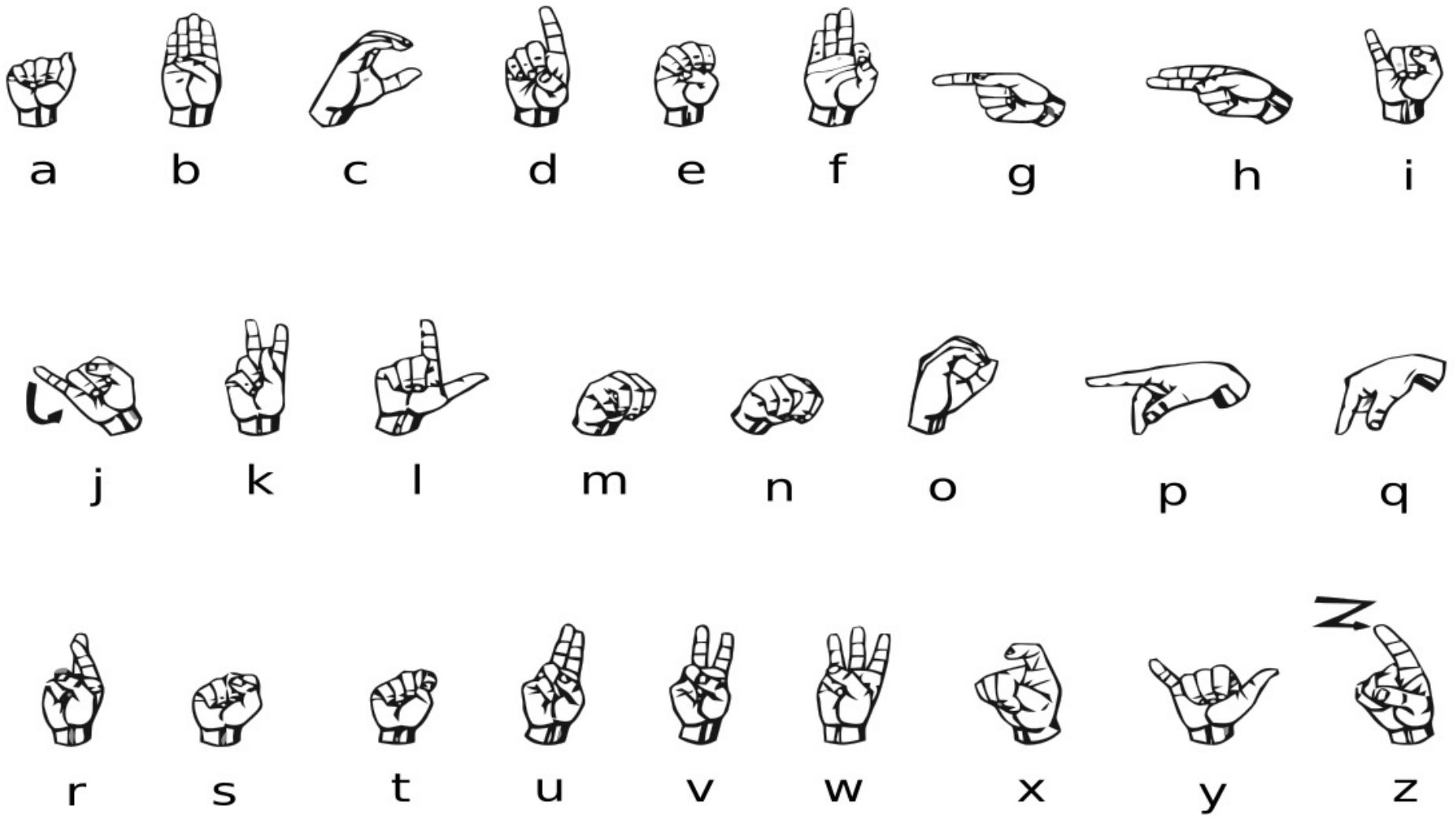
[ ]:
```
import torch.nn as nn
import pandas as pd
import torch
from torch.optim import Adam
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 02_asl.ipynb

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

American Sign Language (ASL) Alphabet



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

1: An Introduction to Deep Learning

2: The Theory Behind Neural Networks

3: Convolutional Neural Networks

4: Data Augmentation and Deployment

5: Pre-trained Models

6: Natural Language Processing

7: Wrap-up and Assessment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Part 3: Convolutional Neural Networks

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Agenda

- Part 1: An Introduction to Deep Learning
- Part 2: How a Neural Network Trains
- Part 3: Convolutional Neural Networks
- Part 4: Data Augmentation and Deployment
- Part 5: Pre-Trained Models
- Part 6: Advanced Architectures

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Recap of the exercise

Trained a dense neural network model

Training accuracy was high

Validation accuracy was low

Evidence of overfitting

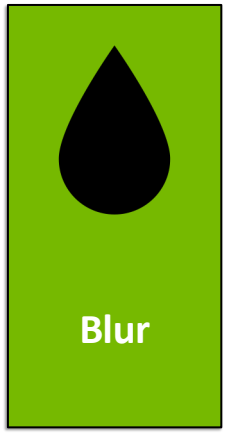
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Original Image



Blur

Sharpen

Brighten

Darken

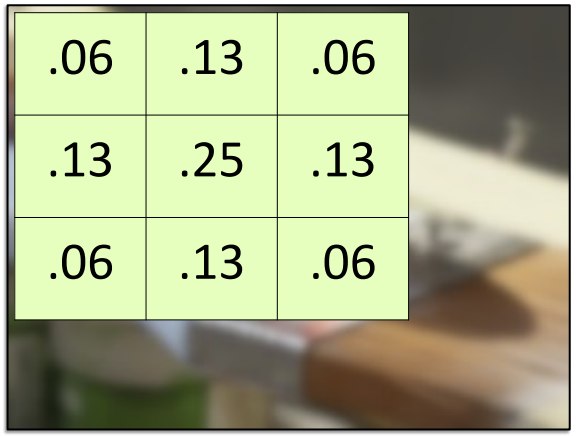
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Original Image

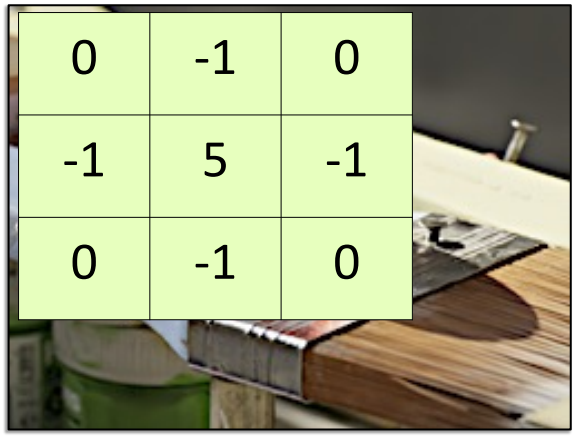
Blur

.06	.13	.06
.13	.25	.13
.06	.13	.06



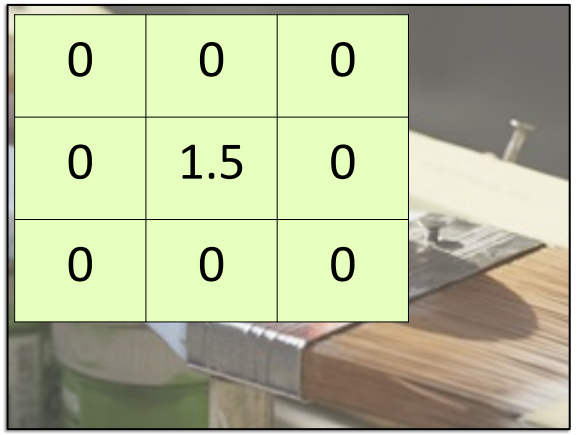
Sharpen

0	-1	0
-1	5	-1
0	-1	0



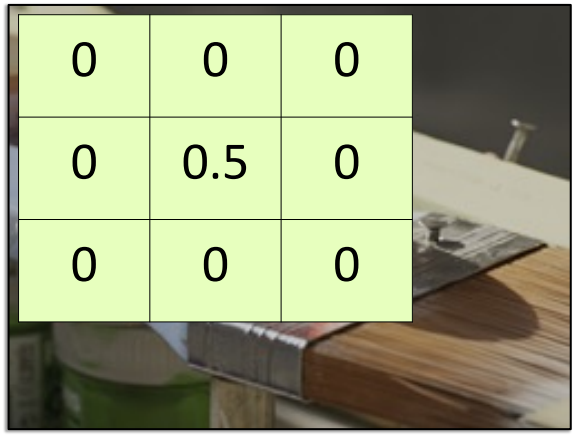
Brighten

0	0	0
0	1.5	0
0	0	0



Darken

0	0	0
0	0.5	0
0	0	0



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Blur Kernel

.06	.13	.06
.13	.25	.13
.06	.13	.06

*

Original Image

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

=

Convolved Image

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Blur Kernel

.06	.13	.06
.13	.25	.13
.06	.13	.06

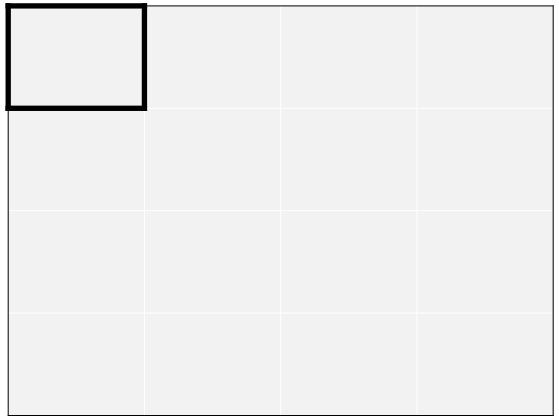
*

Original Image

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

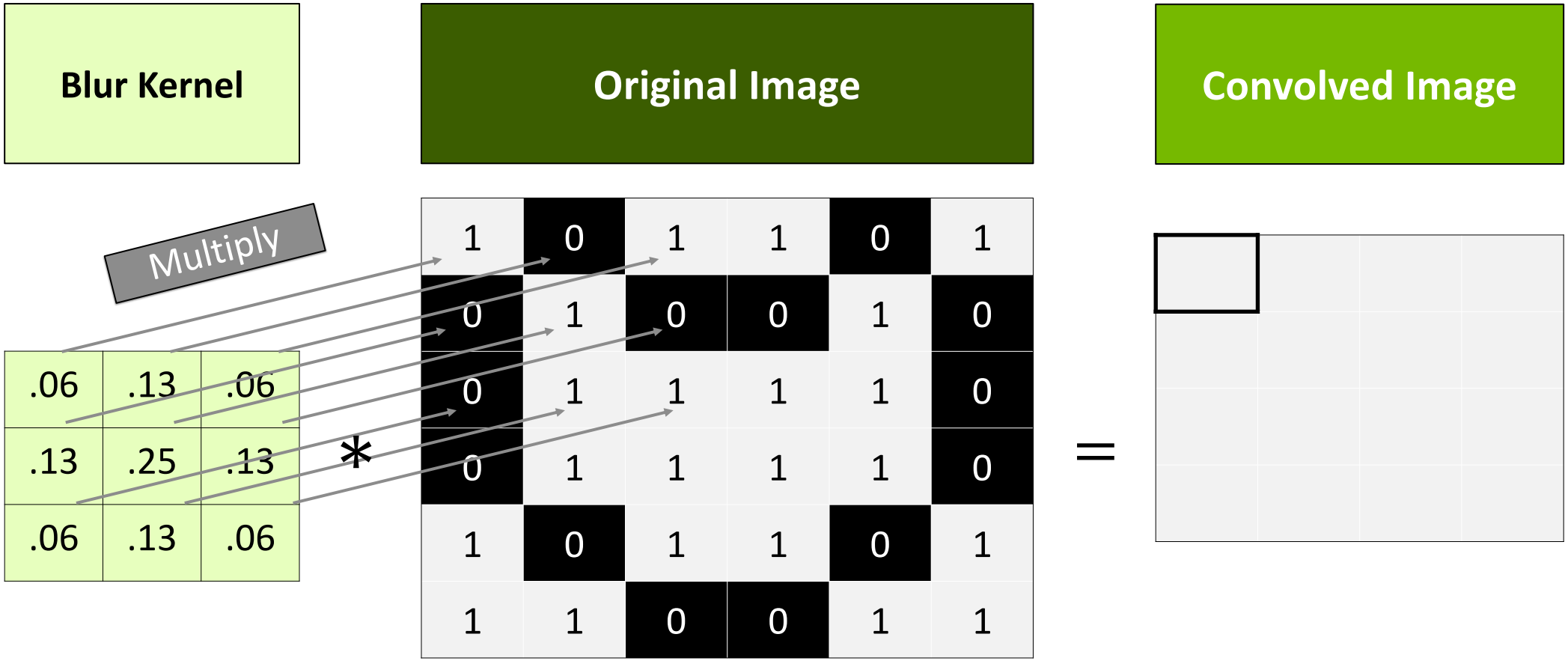
=

Convolved Image



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Blur Kernel

.06	.13	.06
.13	.25	.13
.06	.13	.06

*

Original Image

.06	0	.06	1	0	1
0	.25	0	Total		0
0	.13	.06	1	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

=

Convolved Image

.56			

Kernels and Convolution

Blur Kernel

.06	.13	.06
.13	.25	.13
.06	.13	.06

*

Original Image

1	0	.13	.06	0	1
0	.13	0	0	1	0
0	.06	.13	.06	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

=

Convolved Image

.56	**.57**		

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Convolution

Blur Kernel

.06	.13	.06
.13	.25	.13
.06	.13	.06

*

Original Image

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

=

Convolved Image

.56	.57	.57	.56
.7	.82	.82	.7
.69	.95	.95	.69
.64	.69	.69	.64

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Stride

Stride 1

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0

.56	.57	.57	.56

Stride 2

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0

.56	.57

Stride 3

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0

.56	.56

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Padding

Original Image

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

Zero Padding

0	0	0	0	0	0	0	0
0	1	0	1	1	0	1	0
0	0	1	0	0	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	1	0	1	1	0	1	0
0	1	1	0	0	1	1	0
0	0	0	0	0	0	0	0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Padding

Original Image

1	0	1	1	0	1
0	1	0	0	1	0
0	1	1	1	1	0
0	1	1	1	1	0
1	0	1	1	0	1
1	1	0	0	1	1

Mirror Padding

1	1	0	1	1	0	1	1
1	1	0	1	1	0	1	1
0	0	1	0	0	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
1	1	0	1	1	0	1	1
1	1	1	0	0	1	1	1
1	1	1	0	0	1	1	1

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Neural Networks

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Neural Networks

Kernel

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

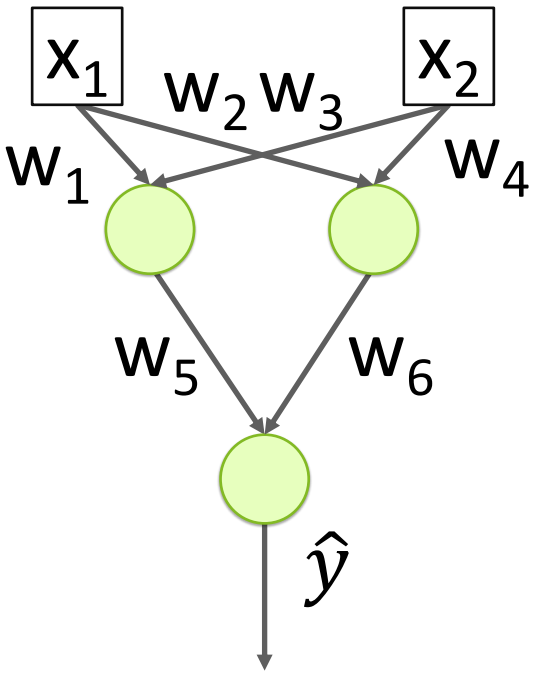
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Kernels and Neural Networks

Kernel

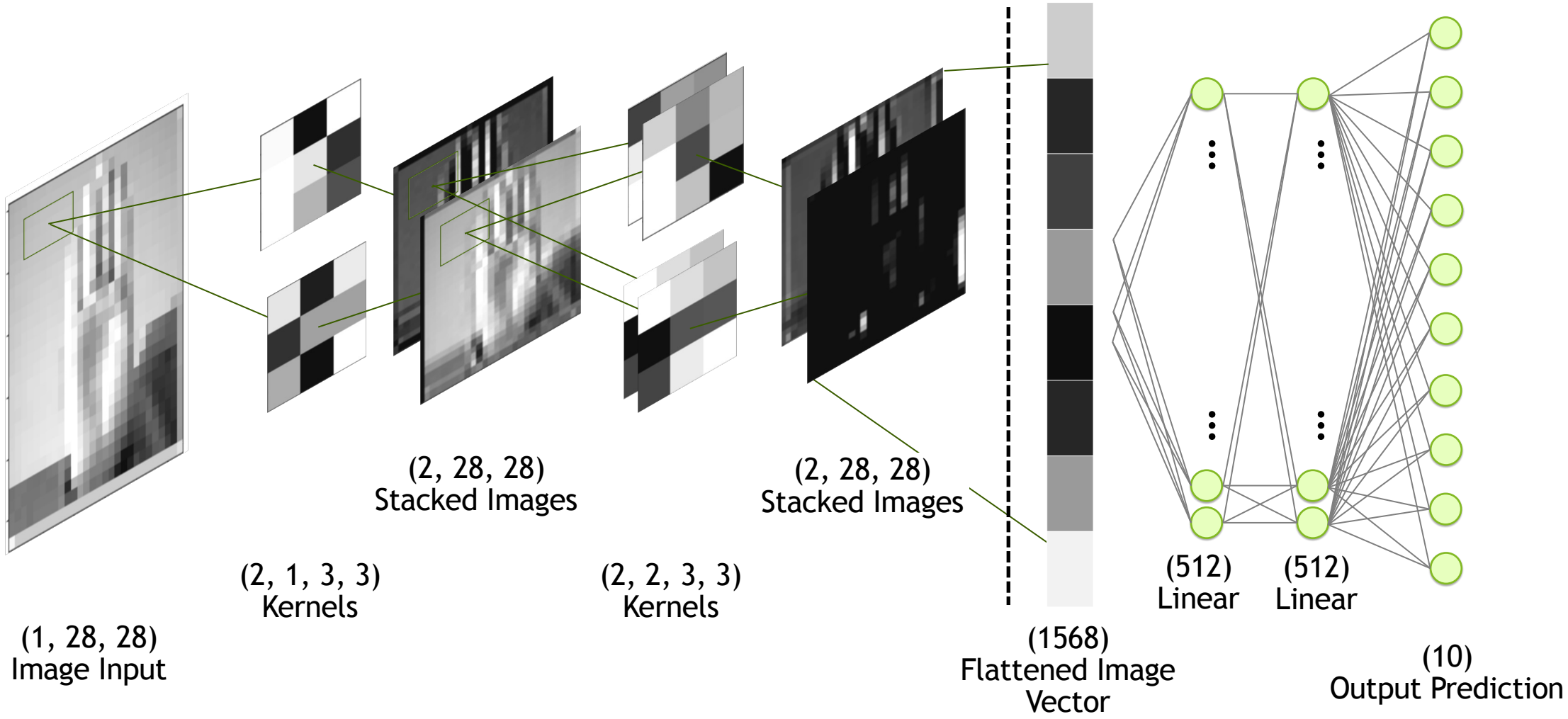
w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

Neuron



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

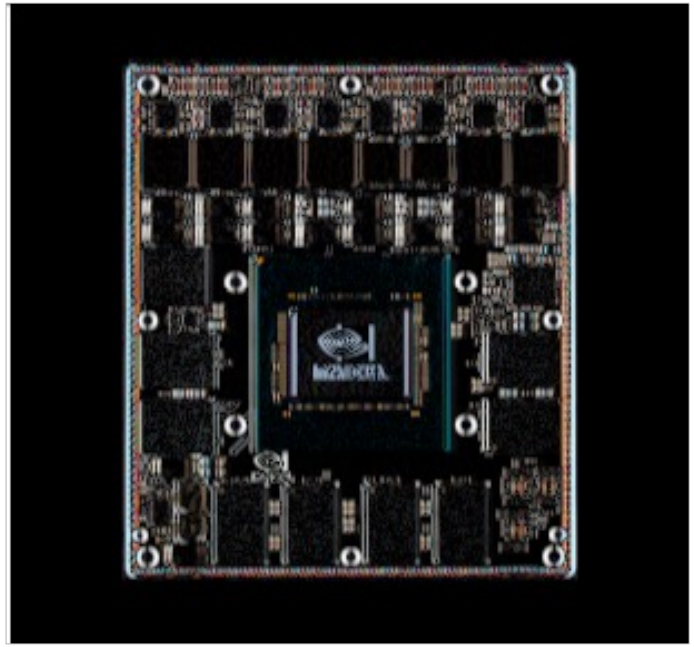
Kernels and Neural Networks



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

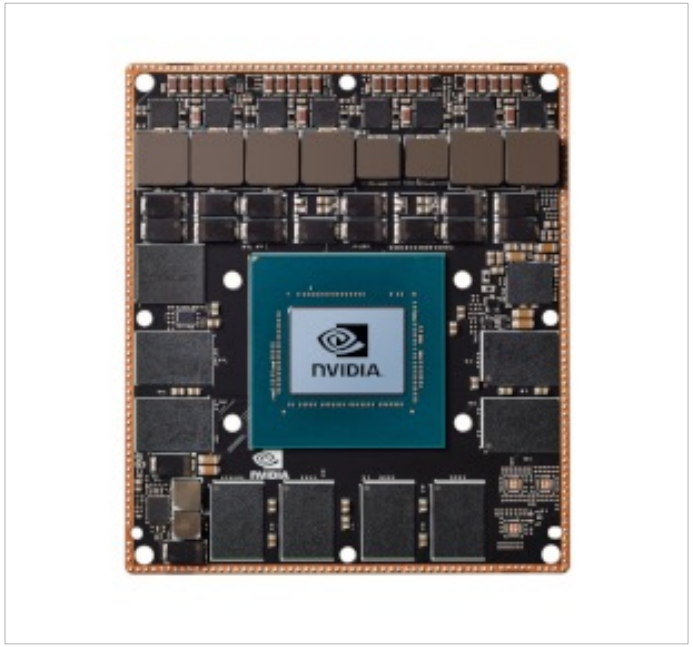
Finding Edges

Vertical Edges



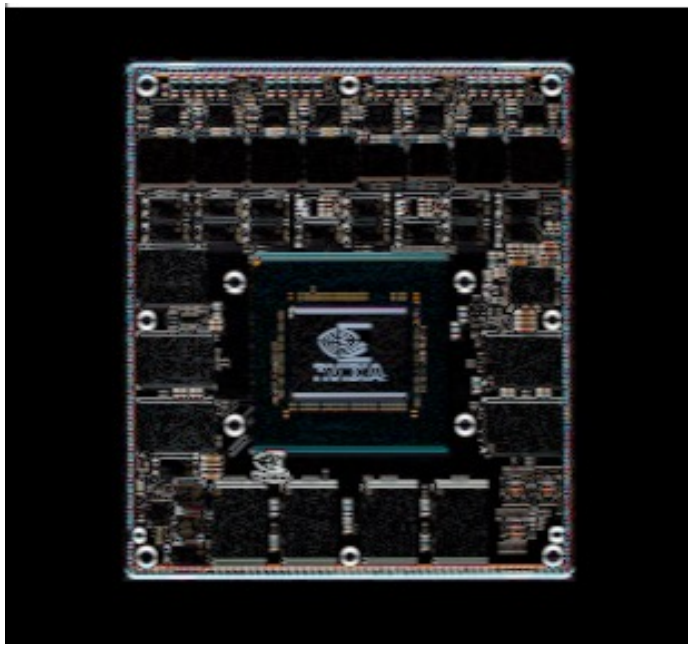
1	0	-1
2	0	-2
1	0	-1

Original Image



0	0	0
0	1	0
0	0	0

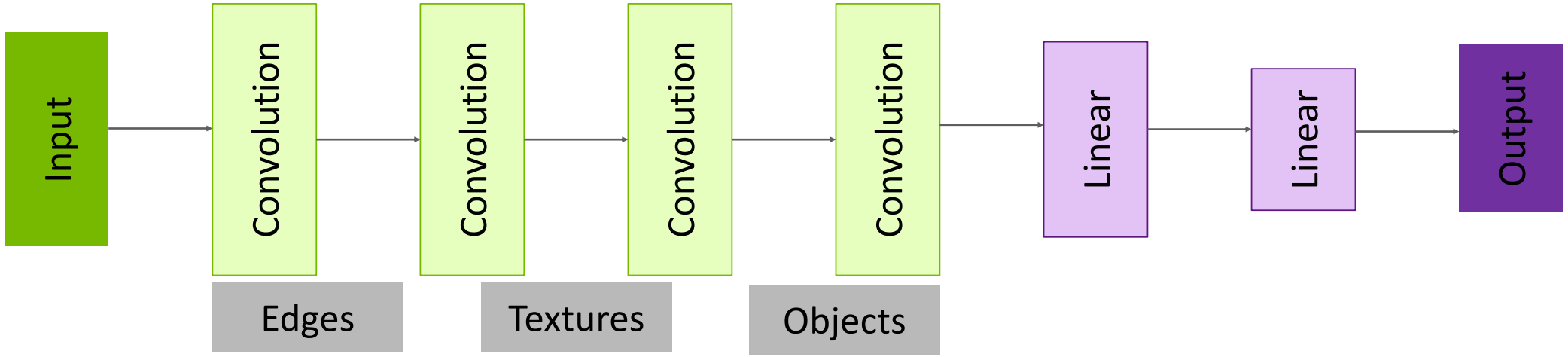
Horizontal Edges

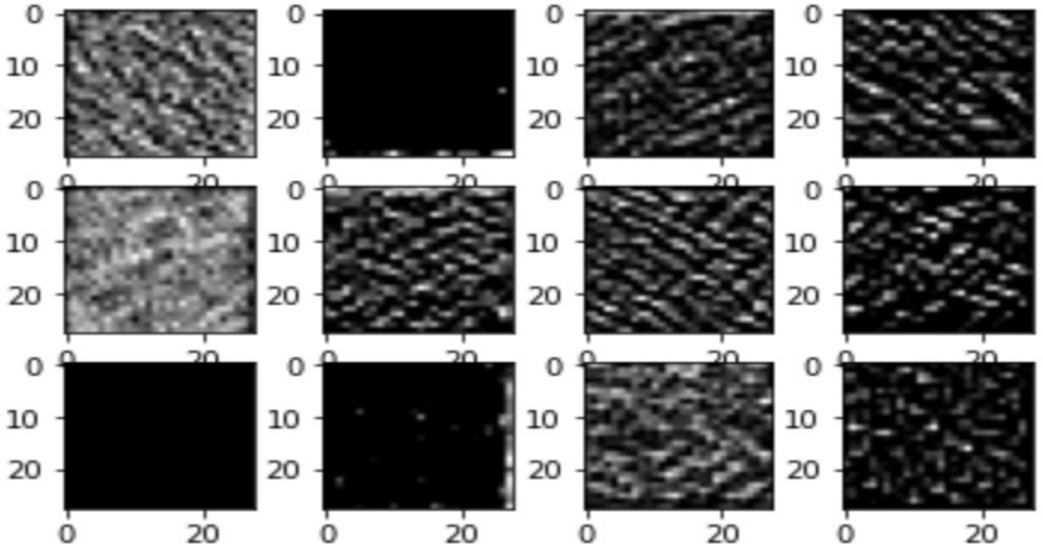


1	2	1
0	0	0
-1	-2	-1

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Neural Network Perception





Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Neural Network Perception

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Other Layers in the Model

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

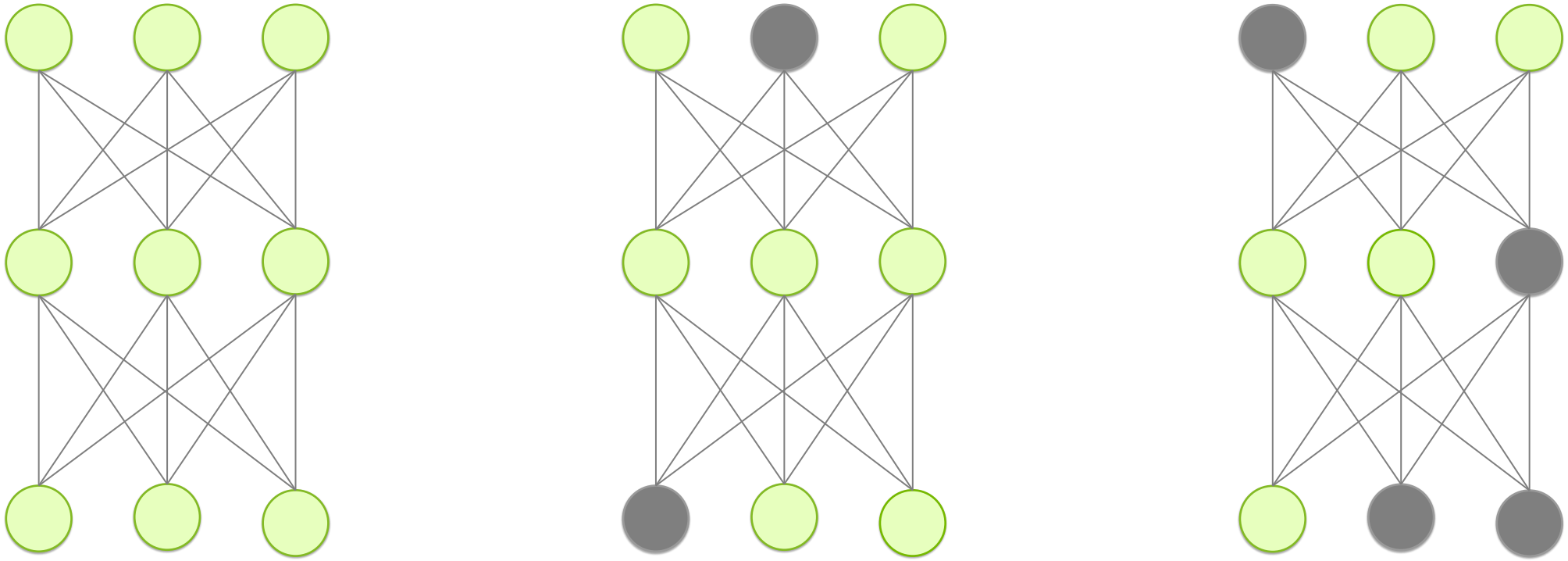
Max Pooling

110	256	153	67
12	89	88	43
10	15	50	55
23	9	49	23

256	153
23	55

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Dropout



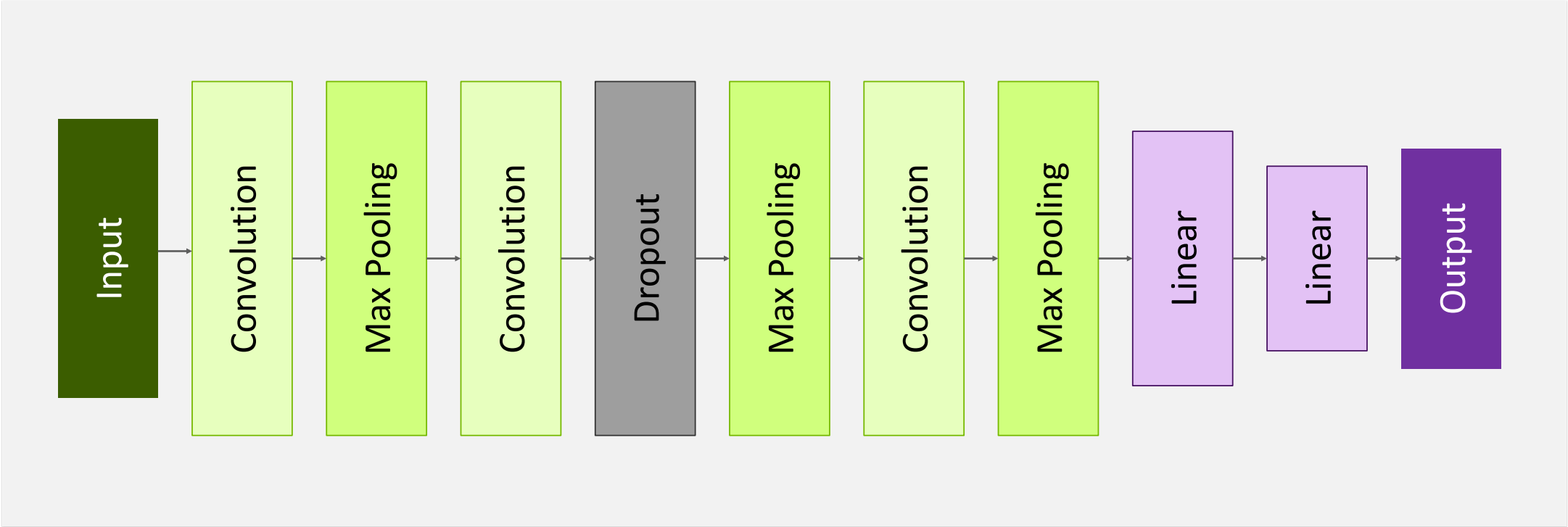
rate = 0

rate = .2

rate = .4

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Whole Architecture



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Let’s go!

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

3. Convolutional Neural Networks

In the previous section, we built and trained a simple model to classify ASL images. The model was able to learn how to correctly classify the training dataset with very high accuracy, but, it did not perform nearly as well on validation dataset. This behavior of not generalizing well to non-training data is called overfitting, and in this section, we will introduce a popular kind of model called a convolutional neural network that is especially good for reading images and classifying them.

3.1 Objectives

- Prep data specifically for a CNN
- Create a more sophisticated CNN model, understanding a greater variety of model layers
- Train a CNN model and observe its performance

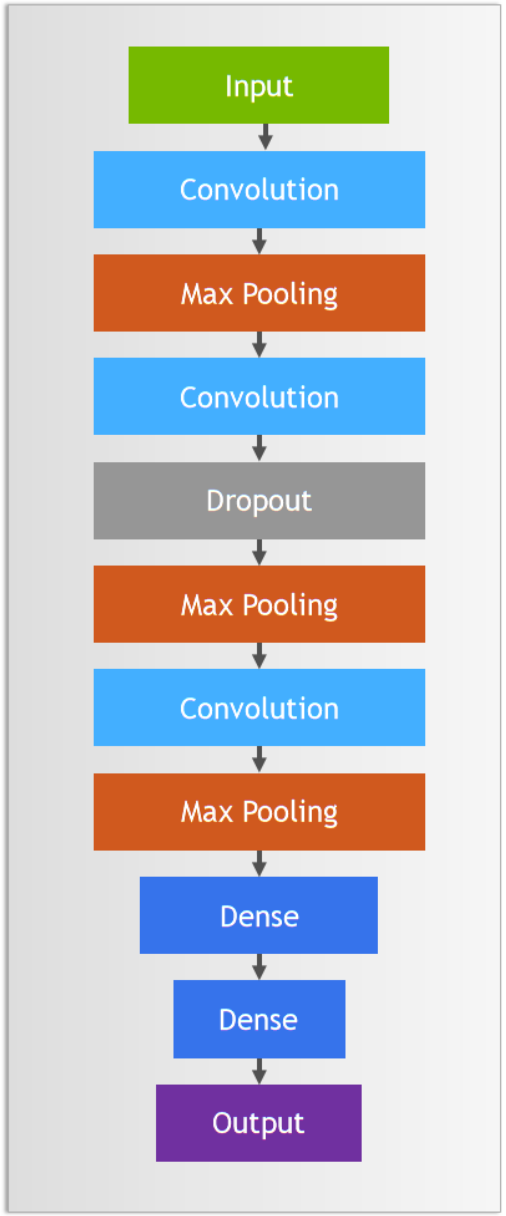
```
[ ]: import torch.nn as nn
     import pandas as pd
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 03_asl_cnn.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

CNN

```
n_classes = 24
kernel_size = 3
flattened_img_size = 75 * 3 * 3


model = nn.Sequential(
      # First convolution
      nn.Conv2d(IMG_CHS, 25, kernel_size, stride=1, padding=1), # 25 x 28 x 28
      nn.BatchNorm2d(25),
      nn.ReLU(),
      nn.MaxPool2d(2, stride=2), # 25 x 14 x 14
      # Second convolution
      nn.Conv2d(25, 50, kernel_size, stride=1, padding=1), # 50 x 14 x 14
      nn.BatchNorm2d(50),
      nn.ReLU(),
      nn.Dropout(.2),
      nn.MaxPool2d(2, stride=2), # 50 x 7 x 7
      # Third convolution
      nn.Conv2d(50, 75, kernel_size, stride=1, padding=1), # 75 x 7 x 7
      nn.BatchNorm2d(75),
      nn.ReLU(),
      nn.MaxPool2d(2, stride=2), # 75 x 3 x 3
      # Flatten to Dense
      nn.Flatten(),
      nn.Linear(flattened_img_size, 512),
      nn.Dropout(.3),
      nn.ReLU(),
      nn.Linear(512, n_classes)
)
```



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Part 4: Data Augmentation and Deployment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Agenda

- Part 1: An Introduction to Deep Learning
- Part 2: How a Neural Network Trains
- Part 3: Convolutional Neural Networks
- Part 4: Data Augmentation and Deployment
- Part 5: Pre-Trained Models
- Part 6: Advanced Architectures

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Recap of the Exercise

Analysis

- CNN increased validation accuracy
- Still seeing training accuracy higher than validation

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Recap of the Exercise

Analysis	Solution
• CNN increased validation accuracy • Still seeing training accuracy higher than validation	• Clean data provides better examples • Dataset variety helps the model generalize



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Data Augmentation

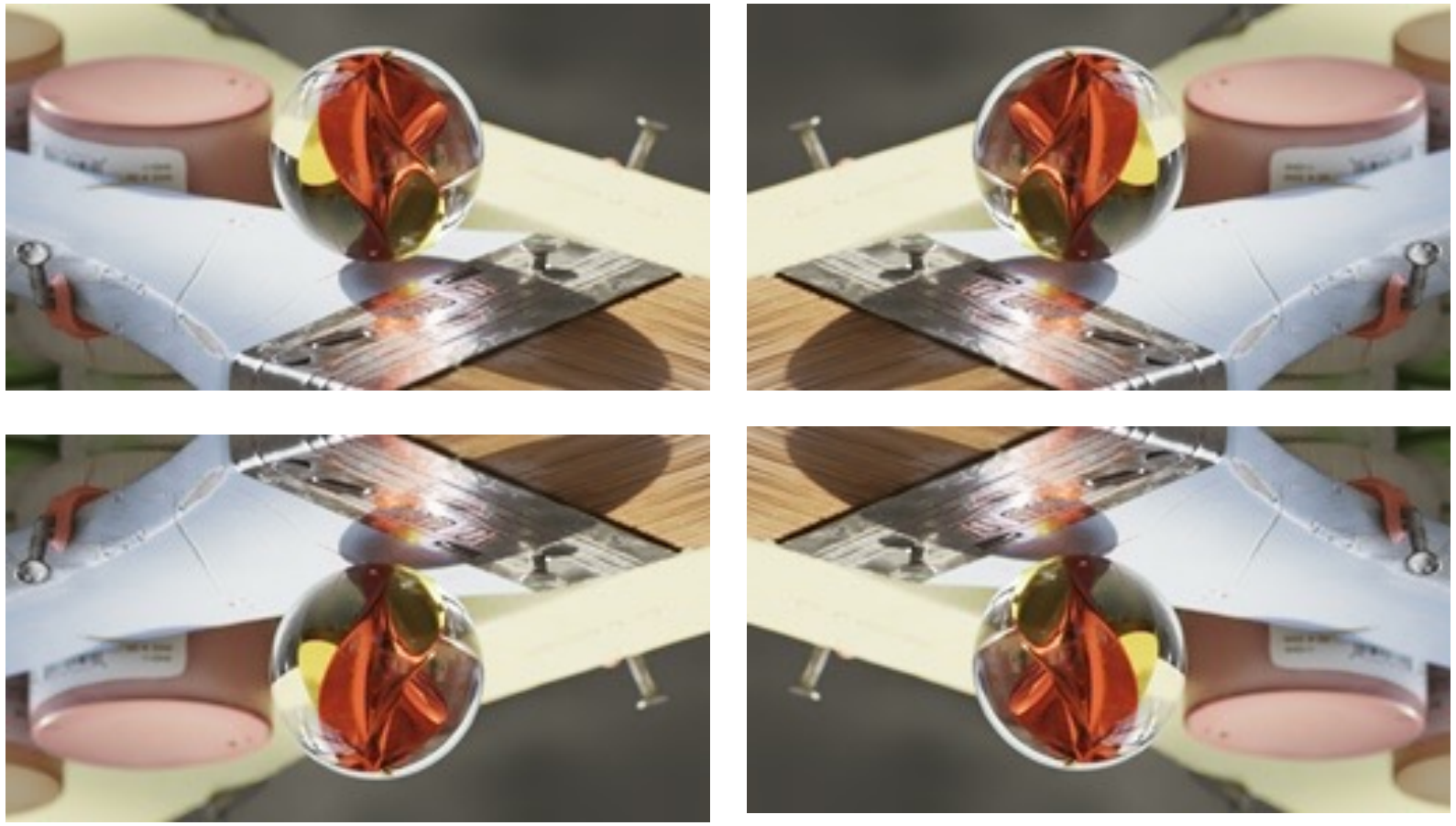
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Data Augmentation

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Image Flipping

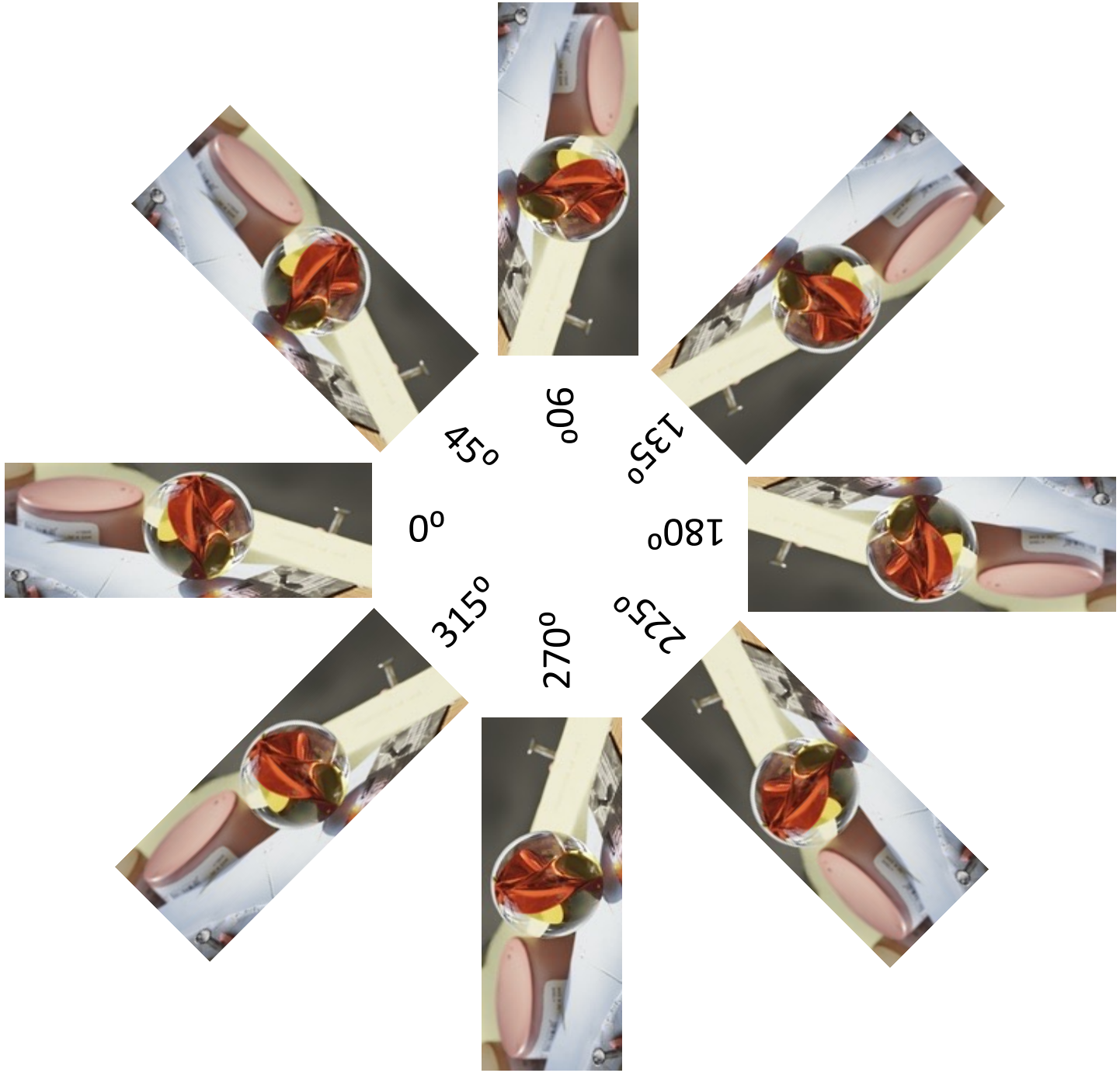
Horizontal Flip



Vertical Flip

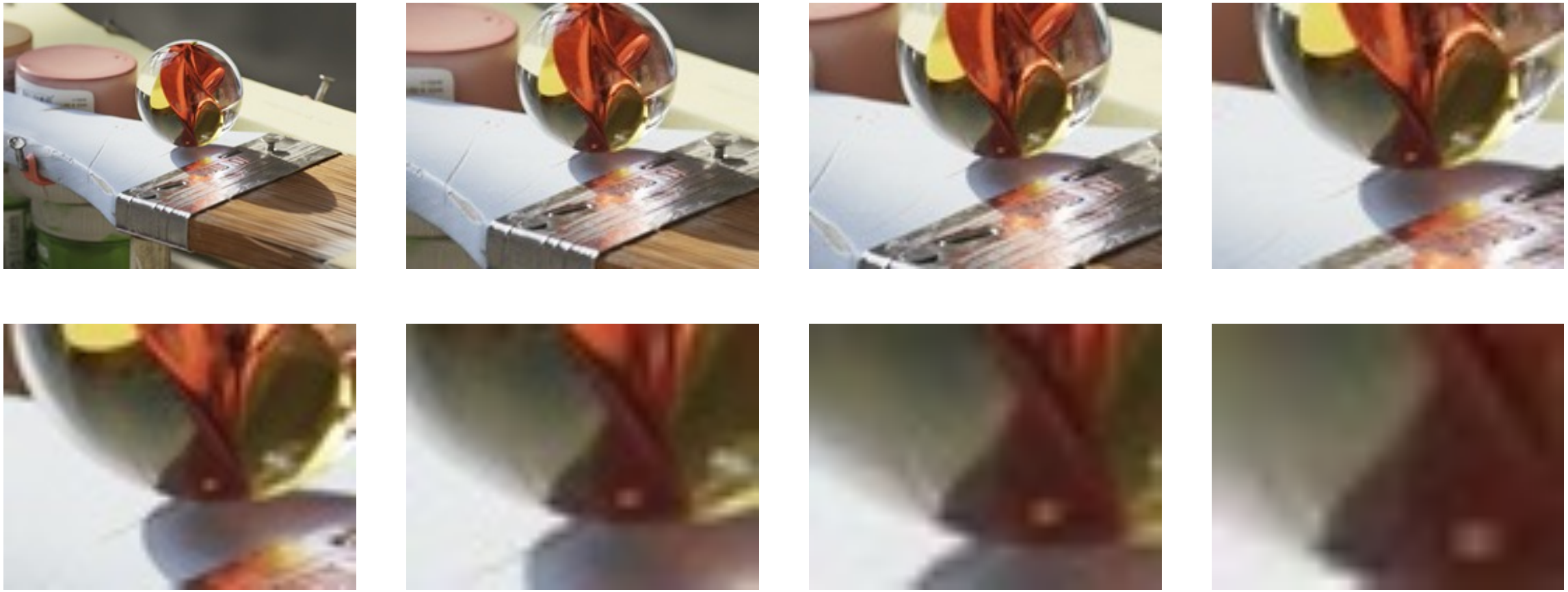
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Rotation



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Zooming

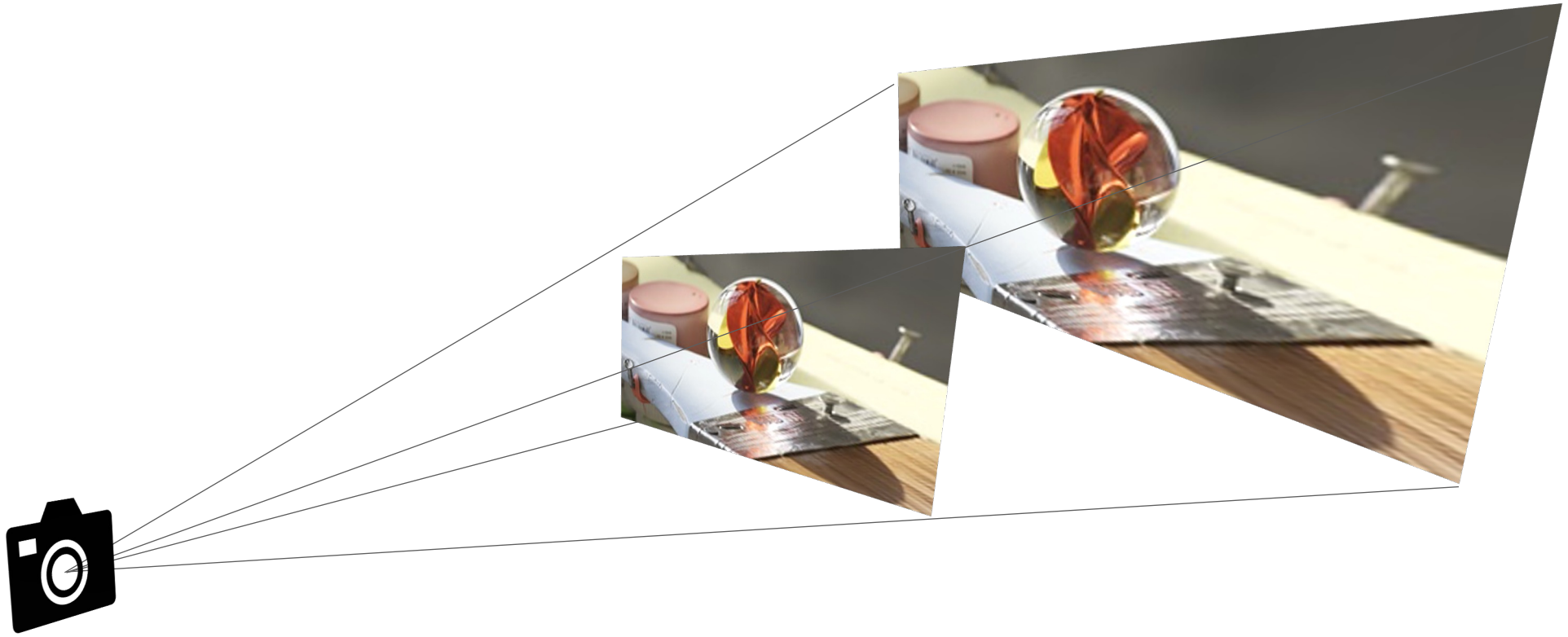


Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Width And Height Shifting

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Homography



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Brightness

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

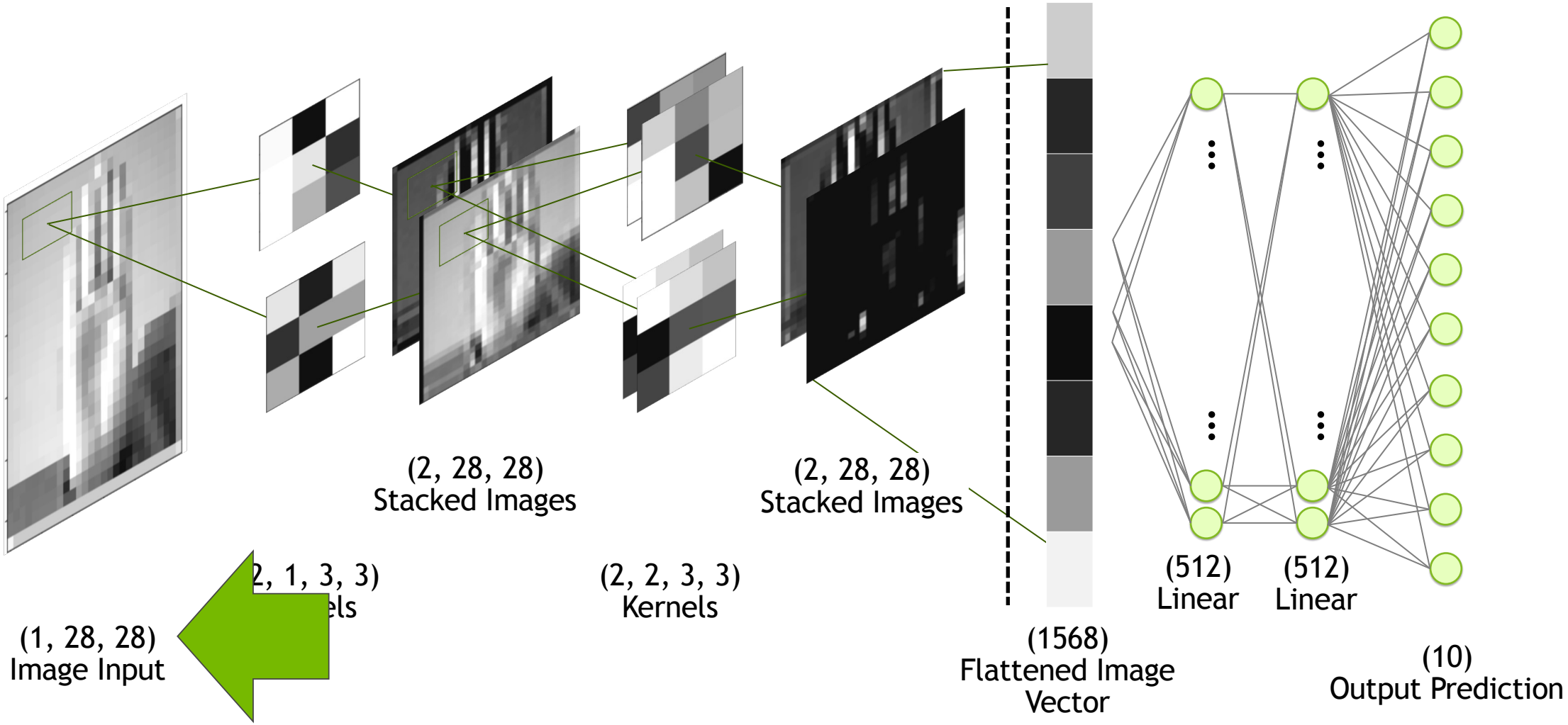
Channel Shifting

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Model Deployment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Model Deployment



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Model Deployment

Training Batch Input

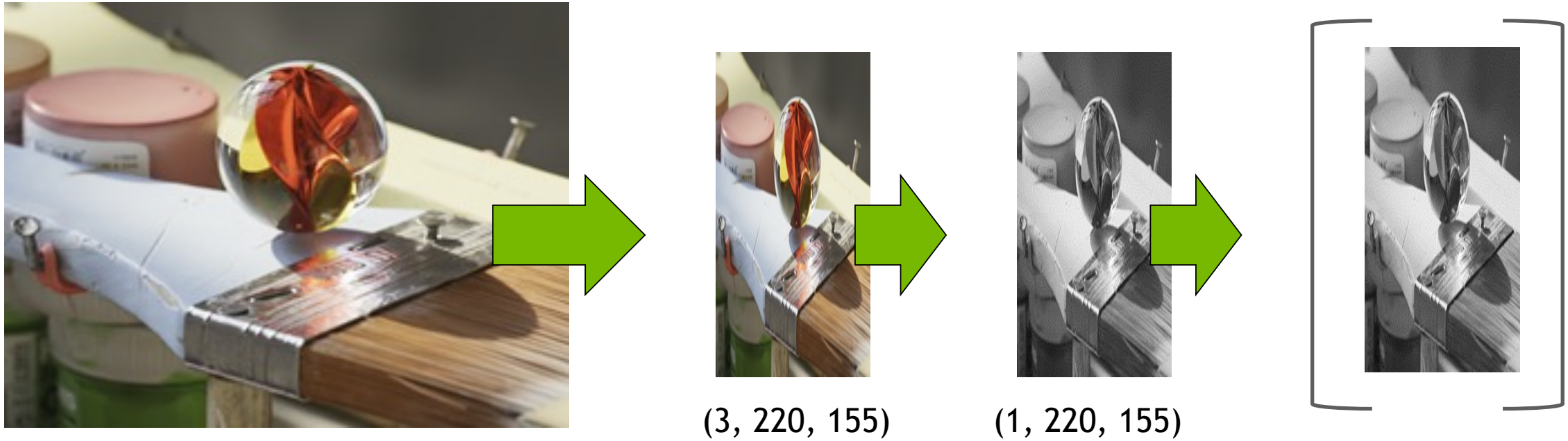
Convolution

Max Pooling

...

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Model Deployment



Resize

Greyscale

“Batch”

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Let’s Try It Out!

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_n | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

4a. Data Augmentation

So far, we've selected a model architecture that vastly improves the model's performance, as it is designed to recognize important features in the images. The validation accuracy is still lagging behind the training accuracy, which is a sign of overfitting: the model is getting confused by things it has not seen before when it tests against the validation dataset.

In order to teach our model to be more robust when looking at new data, we're going to programmatically increase the size and variance in our dataset. This is known as *data augmentation*, a useful technique for many deep learning applications.

The increase in size gives the model more images to learn from while training. The increase in variance helps the model ignore unimportant features and select only the features that are truly important in classification, allowing it to generalize better.

4a.1 Objectives

- Augment the ASL dataset

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 04a_asl_augmentation.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_n | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

4b. Deploying Your Model

Now that we have a well trained model, it's time to use it. In this exercise, we'll expose new images to our model and detect the correct letters of the sign language alphabet. Let's get started!

4b.1 Objectives

- Load an already-trained model from disk
- Reformat images for a model trained on images of a different format
- Perform inference with new images, never seen by the trained model and evaluate its performance

[ ]:
```
import pandas as pd
import torch
import torch.nn as nn
from torch.optim import Adam
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 04b_asl_predictions.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

1: An Introduction to Deep Learning

2: The Theory Behind Neural Networks

3: Convolutional Neural Networks

4: Data Augmentation and Deployment

5: Pre-trained Models

6: Natural Language Processing

7: Wrap-up and Assessment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Part 5: Pre-trained Models

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

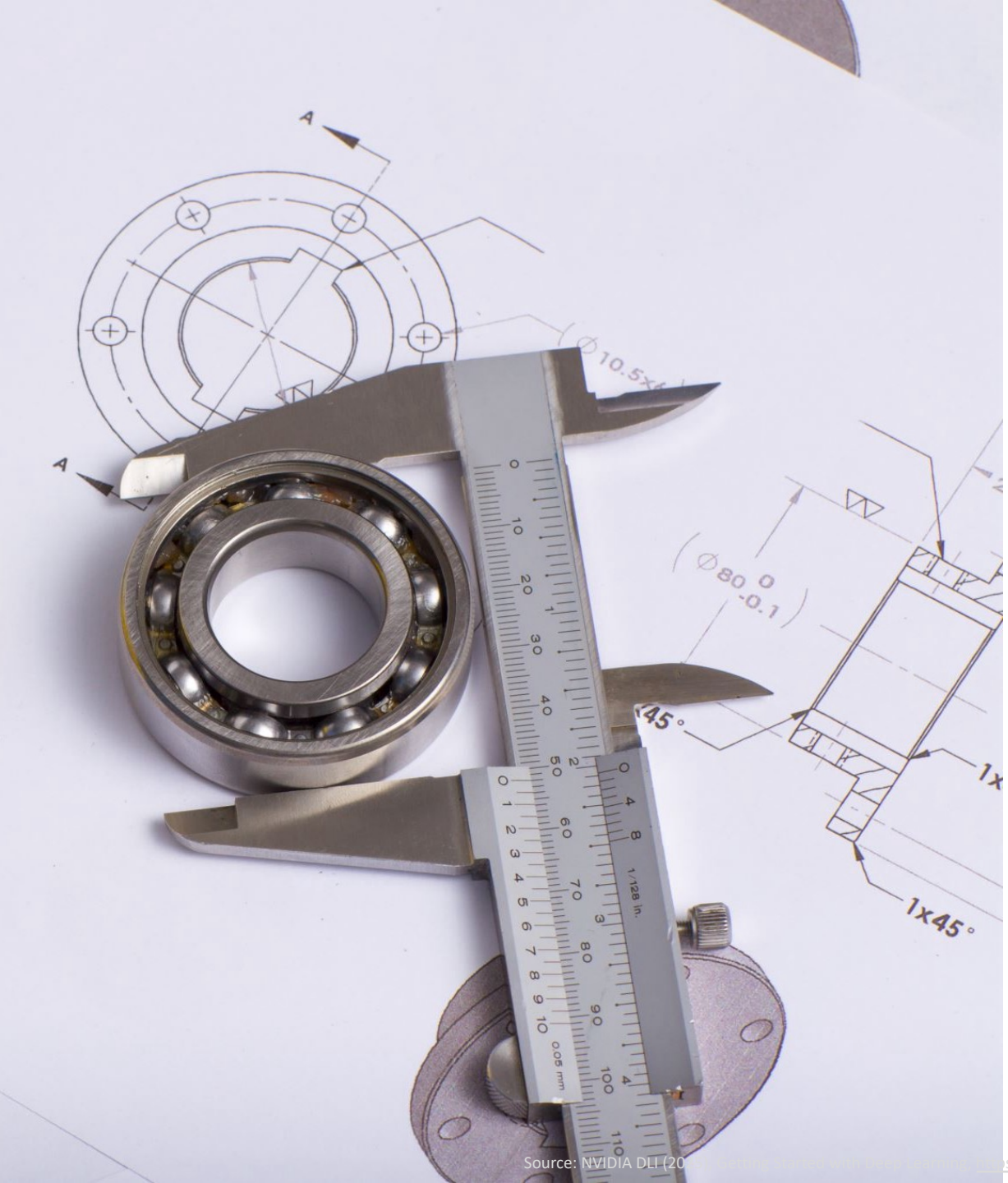
Agenda

- Part 1: An Introduction to Deep Learning
- Part 2: How a Neural Network Trains
- Part 3: Convolutional Neural Networks
- Part 4: Data Augmentation and Deployment
- Part 5: Pre-Trained Models
- Part 6: Advanced Architectures

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Review So Far

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1



Review So Far

- Learning Rate
- Number of Layers
- Neurons per Layer
- Activation Functions
- Dropout
- Data

Source: NVIDIA DLI (2019). Getting Started with Deep Learning. https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Pre-Trained Models

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Pre-Trained Models

TensorFlow Hub

Keras

NVIDIA NGC

PYTORCH HUB

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Pre-Trained Models

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan* **& Andrew Zisserman+**
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen,az}@robots.ox.ac.uk

IMAGENET

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Next Challenge

An Automated Doggy Door

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transfer Learning

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

The Challenge After

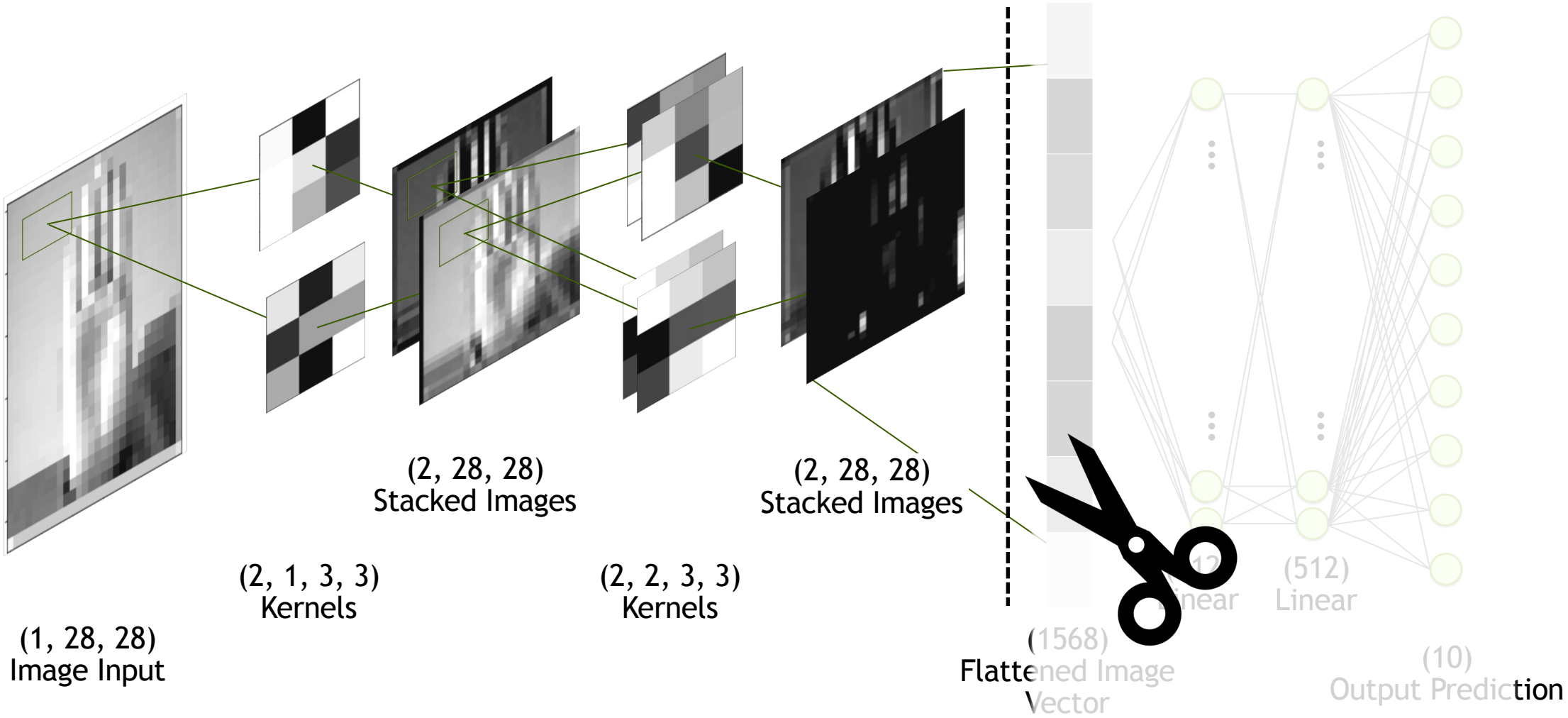
An Automated Presidential Doggy Door

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transfer Learning

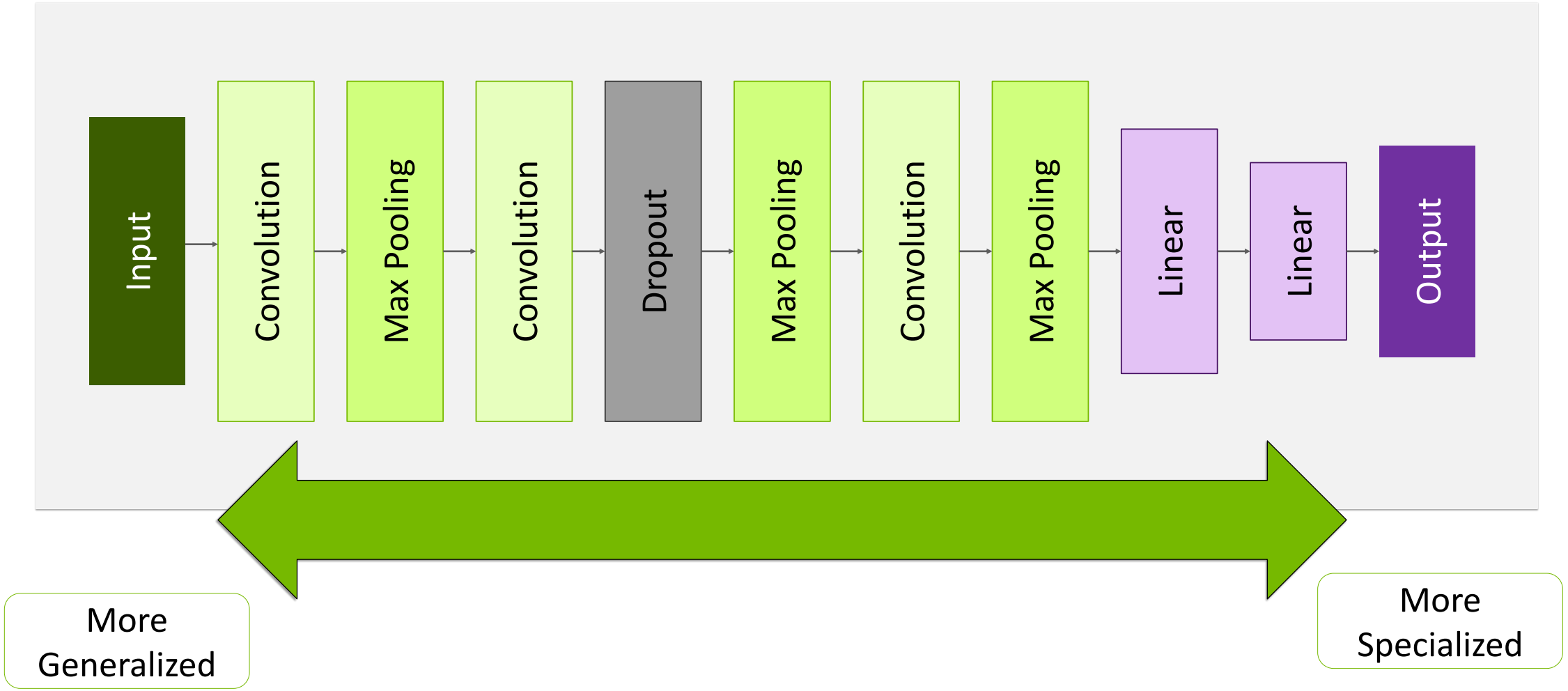
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transfer Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transfer Learning



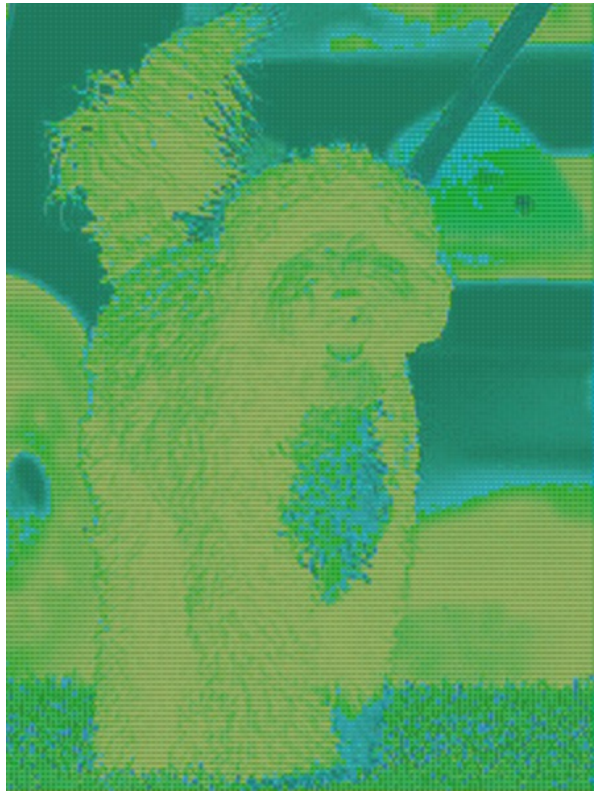
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transfer Learning

Freezing the Model?

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transfer Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Let’s Get Started!

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

5a. Pre-Trained Models ¶

Though it is often necessary to have a large, well annotated dataset to solve a deep learning challenge, there are many freely available pre-trained models that we can use right out of the box. As you decide to take on your own deep learning project, it is a great idea to start by looking for existing models online that can help you achieve your goal. A great place to explore available models is NGC. There are also many models hosted on GitHub that you can find through searching on Google.

5a.1 Objectives

- Use TorchVision to load a very well-trained pretrained model
- Preprocess our own images to work with the pretrained model
- Use the pretrained model to perform accurate inference on your own images

[ ]:
```
import torch
import torchvision.transforms.v2 as transforms
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 05a_doggy_door.ipynb 0 Trash

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_ji | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

5b. Transfer Learning

So far, we have trained accurate models on large datasets, and also downloaded a pre-trained model that we used with no training necessary. But what if we cannot find a pre-trained model that does exactly what you need, and what if we do not have a sufficiently large dataset to train a model from scratch? In this case, there is a very helpful technique we can use called transfer learning.

With transfer learning, we take a pre-trained model and retrain it on a task that has some overlap with the original training task. A good analogy for this is an artist who is skilled in one medium, such as painting, who wants to learn to practice in another medium, such as charcoal drawing. We can imagine that the skills they learned while painting would be very valuable in learning how to draw with charcoal.

As an example in deep learning, say we have a pre-trained model that is very good at recognizing different types of cars, and we want to train a model to recognize types of motorcycles. A lot of the learnings of the car model would likely be very useful, for instance the ability to recognize headlights and wheels.

Transfer learning is especially powerful when we do not have a large and varied dataset. In this case, a model trained from

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 05b_presidential_doggy_door.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Part 6: Advanced Architectures

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fundamentals of Deep Learning

Agenda

- Part 1: An Introduction to Deep Learning
- Part 2: How a Neural Network Trains
- Part 3: Convolutional Neural Networks
- Part 4: Data Augmentation and Deployment
- Part 5: Pre-Trained Models
- Part 6: Advanced Architectures

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Moving Forward

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

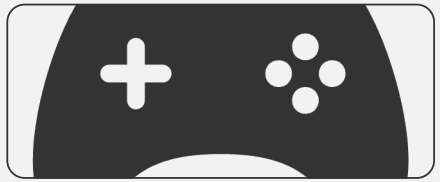
Fields of AI

Computer Vision

- Optometry

Natural Language Processing

- Linguistics



Reinforcement Learning

- Game Theory
- Psychology

Anomaly Detection

- Security
- Medicine

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

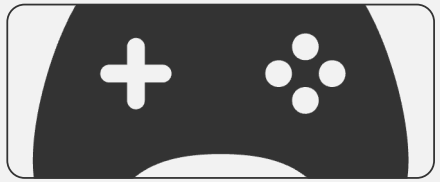
Fields of AI

Computer Vision

- Optometry

Natural Language Processing

- Linguistics



Reinforcement Learning

- Game Theory
- Psychology

Anomaly Detection

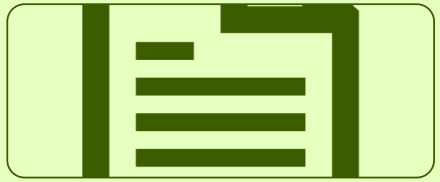
- Security
- Medicine

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Fields of AI

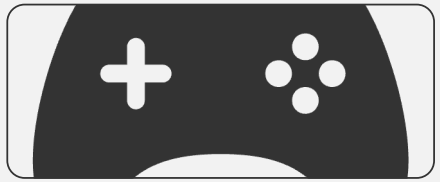
Computer Vision

- Optometry



Natural Language Processing

- Linguistics



Reinforcement Learning

- Game Theory
- Psychology

Anomaly Detection

- Security
- Medicine

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

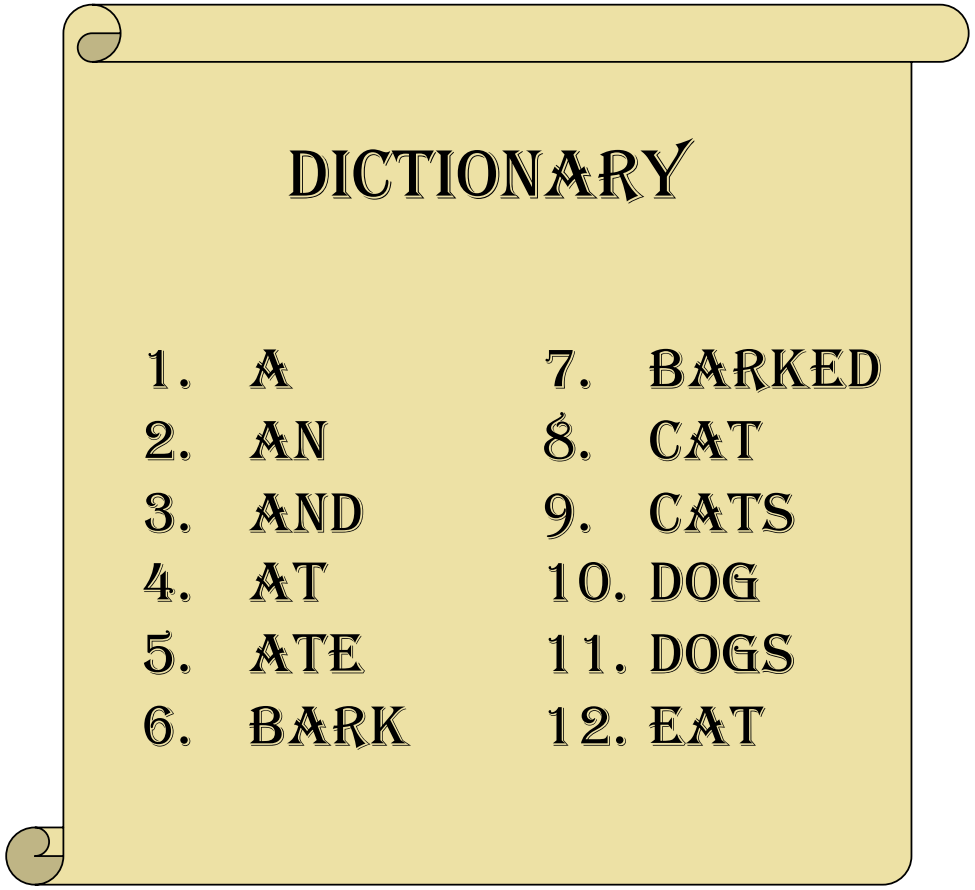
Natural Language Processing

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

From Words to Numbers

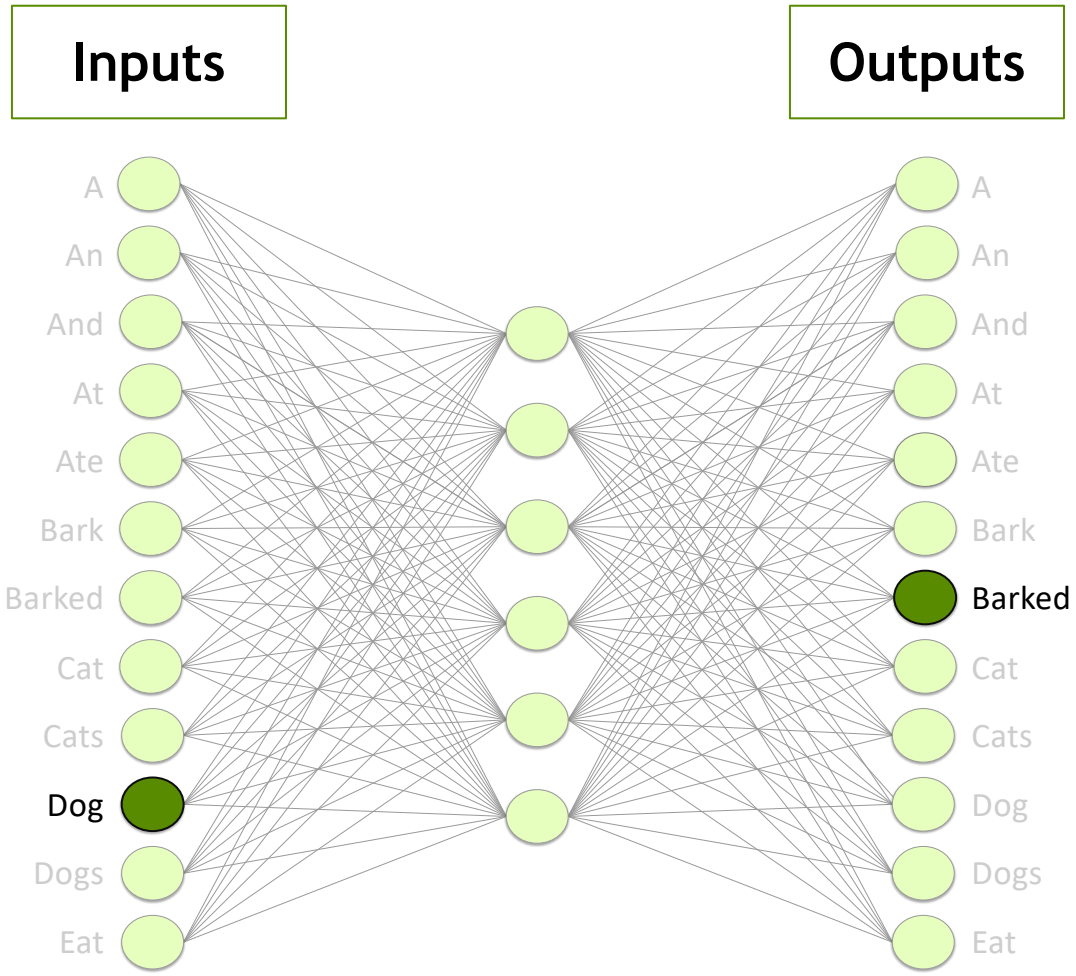
"A dog barked at a cat."

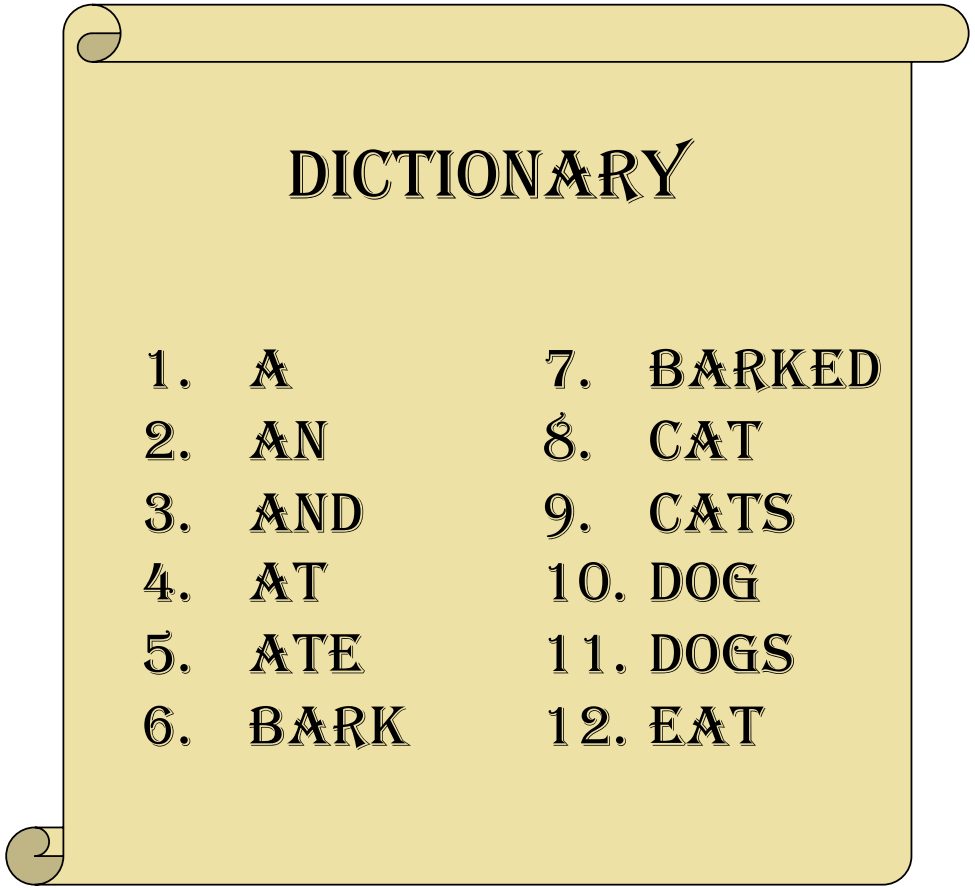
[1, 10, 7, 4, 1, 8]



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

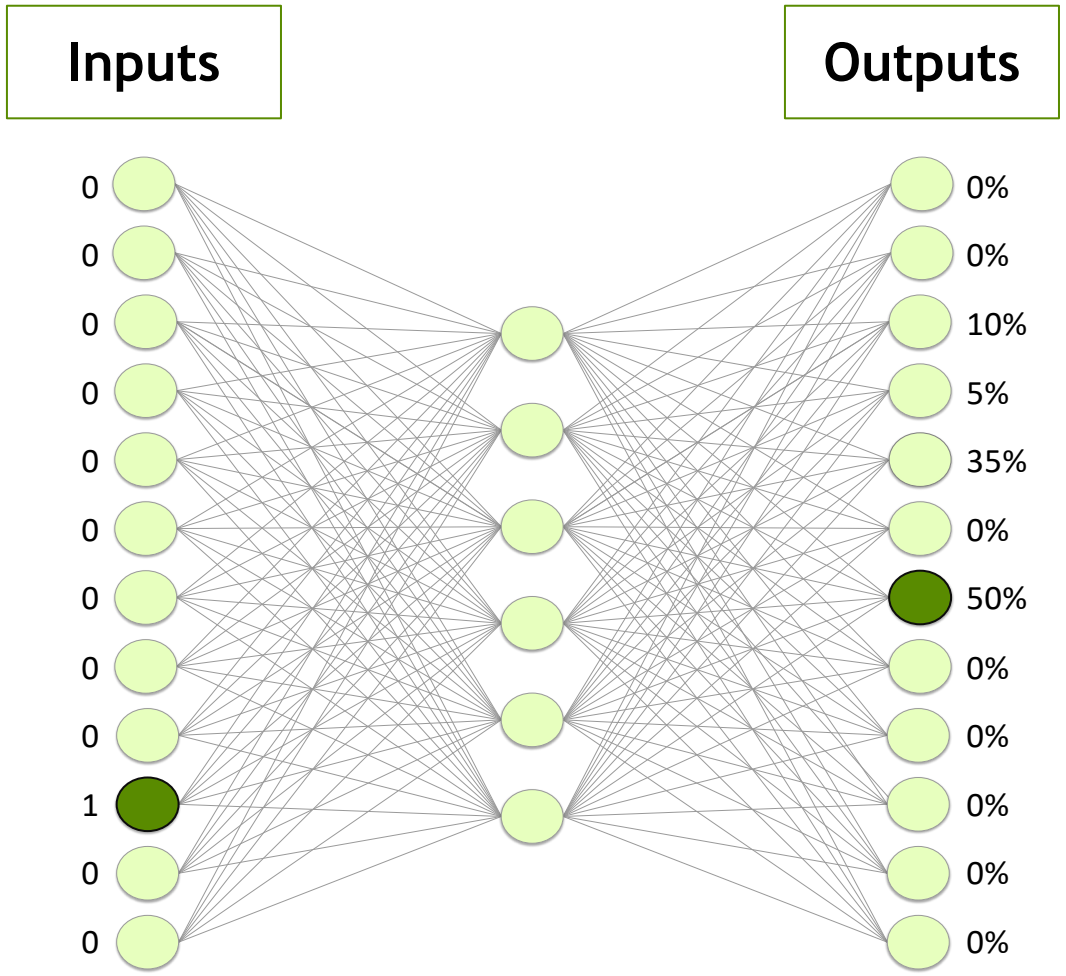
From Words to Numbers

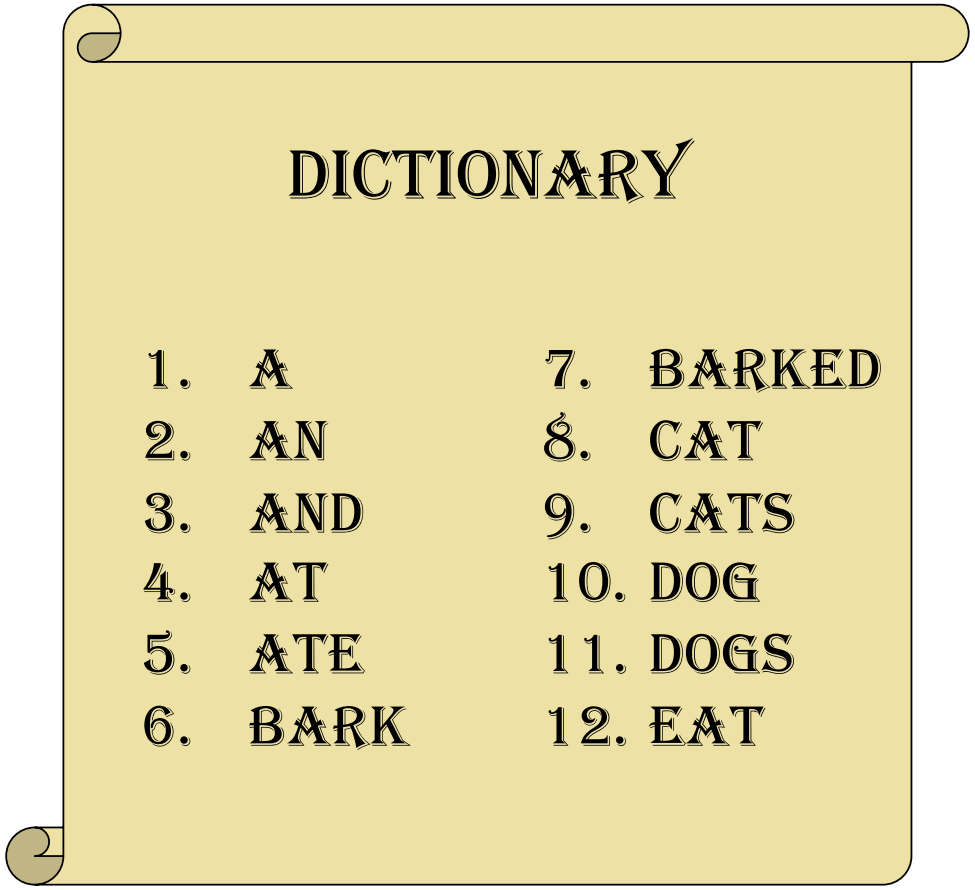




Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

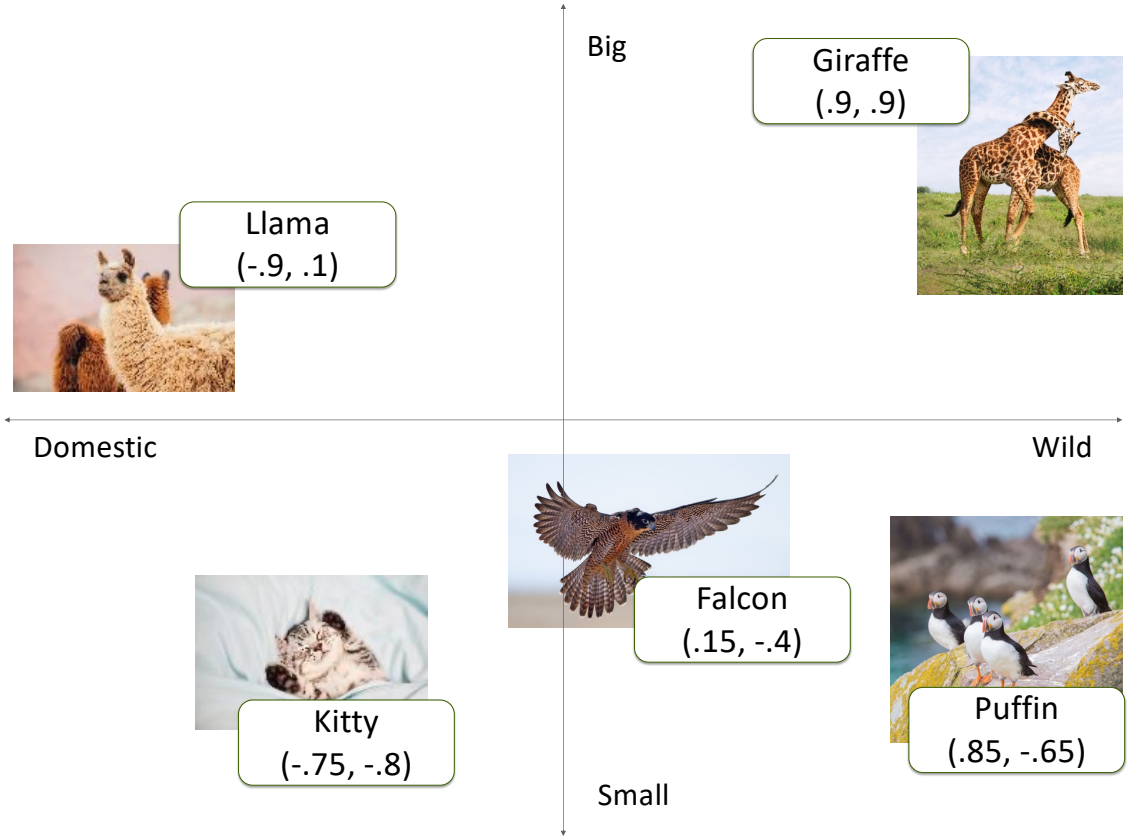
From Words to Numbers

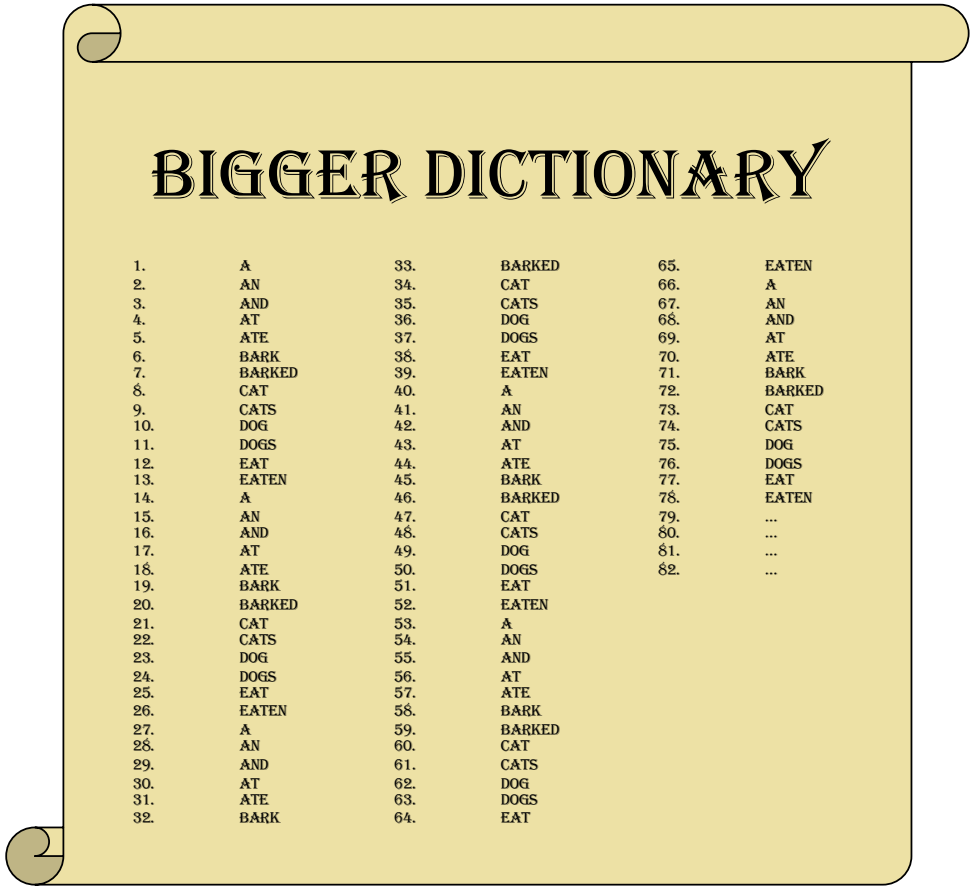




Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

From Words to Numbers



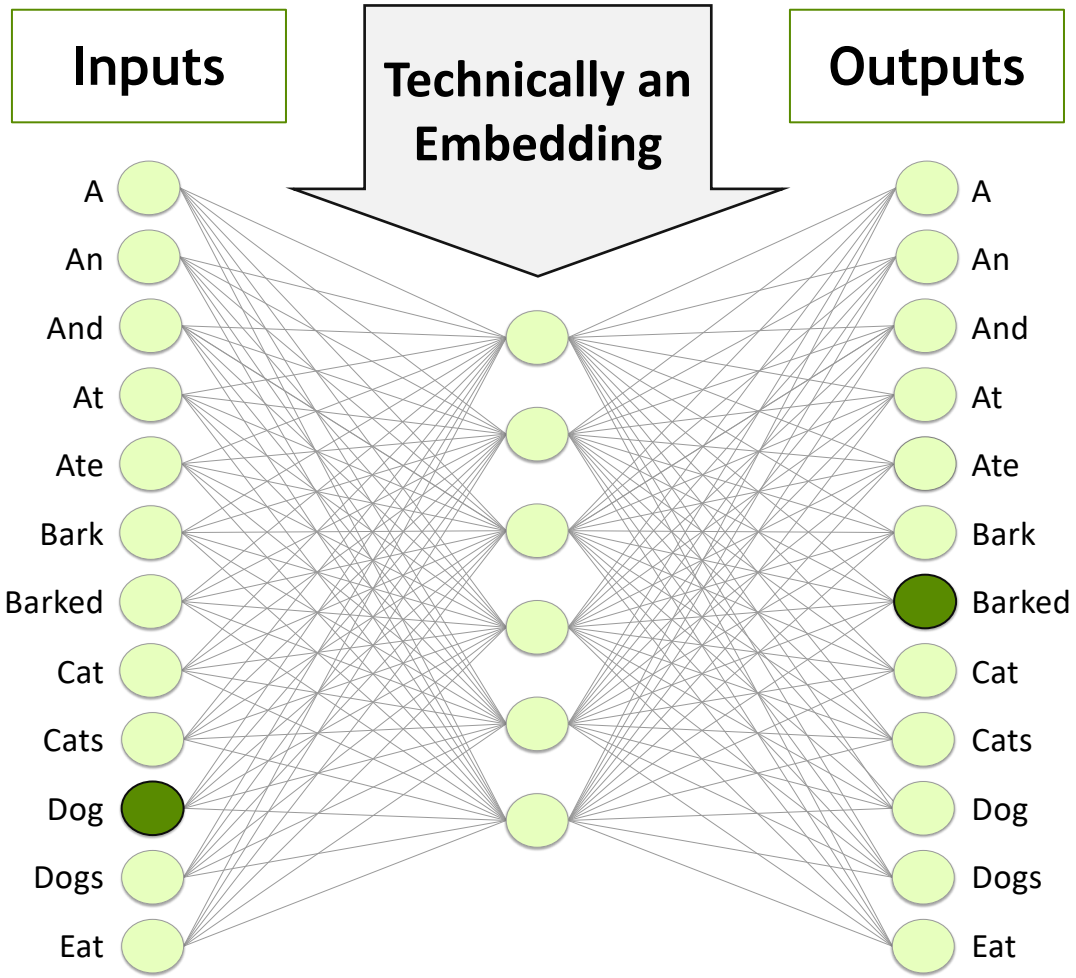


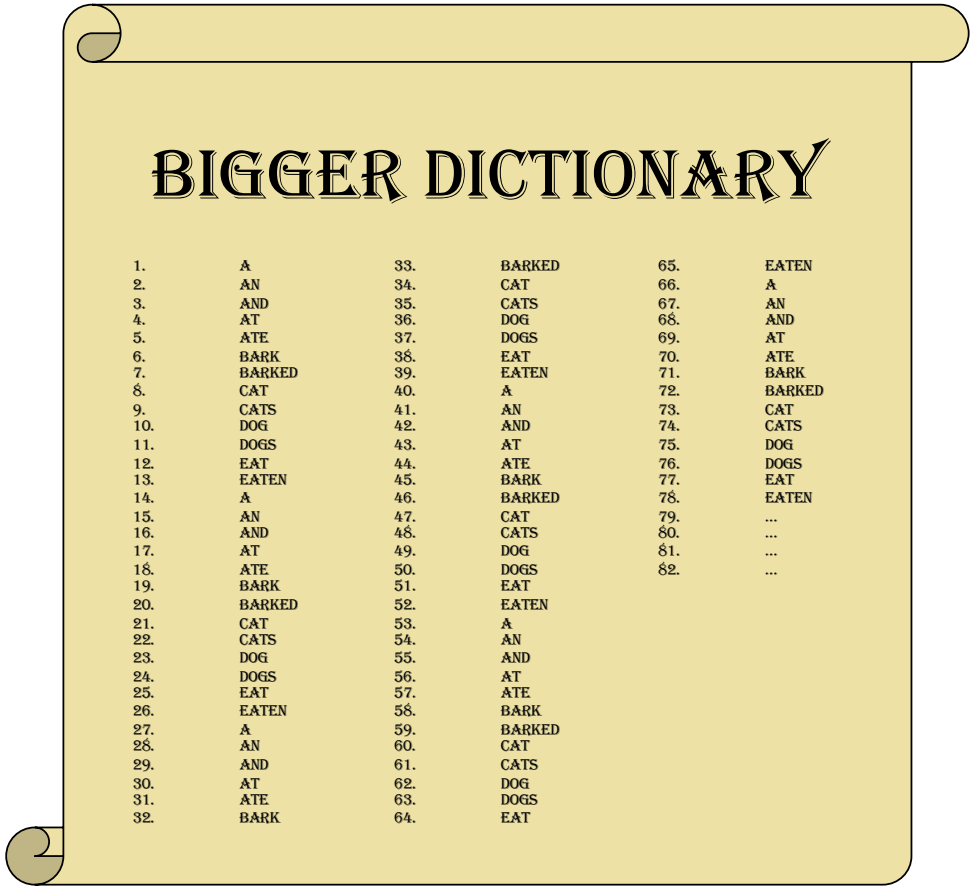
BIGGER DICTIONARY

1.	A	33.	BARKED	65.	EATEN
2.	AN	34.	CAT	66.	A
3.	AND	35.	CATS	67.	AN
4.	AT	36.	DOG	68.	AND
5.	ATE	37.	DOGS	69.	AT
6.	BARK	38.	EAT	70.	ATE
7.	BARKED	39.	EATEN	71.	BARK
8.	CAT	40.	A	72.	BARKED
9.	CATS	41.	AN	73.	CAT
10.	DOG	42.	AND	74.	CATS
11.	DOGS	43.	AT	75.	DOG
12.	EAT	44.	ATE	76.	DOGS
13.	EATEN	45.	BARK	77.	EAT
14.	A	46.	BARKED	78.	EATEN
15.	AN	47.	CAT	79.	...
16.	AND	48.	CATS	80.	...
17.	AT	49.	DOG	81.	...
18.	ATE	50.	DOGS	82.	...
19.	BARK	51.	EAT		
20.	BARKED	52.	EATEN		
21.	CAT	53.	A		
22.	CATS	54.	AN		
23.	DOG	55.	AND		
24.	DOGS	56.	AT		
25.	EAT	57.	ATE		
26.	EATEN	58.	BARK		
27.	A	59.	BARKED		
28.	AN	60.	CAT		
29.	AND	61.	CATS		
30.	AT	62.	DOG		
31.	ATE	63.	DOGS		
32.	BARK	64.	EAT		

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

From Words to Numbers





Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Sentence Prediction

I am the very model of a modern Major-Gineral,
I've information vegetable, animal, and mineral,

...

I'm very good at integral and differential calculus;
I know the scientific names of beings animalculous:
In short, in matters vegetable, animal, and mineral,
I am the very model of a m

~ Major-General Stanley

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Sentence Prediction

I am the very model of a modern Major-Gineral,
I've information vegetable, animal, and mineral,

...

I'm very good at integral and differential calculus;
I know the scientific names of beings animalculous:
In short, in matters vegetable, animal, and mineral,
I am the very model of a modern Major-Gineral.

~ Major-General Stanley

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention

I
am
the
very
model

5 x 3

Q

Query

5 x 3

K

Key

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention

I						
am						
the						
very						
model						

5 x 3

Q

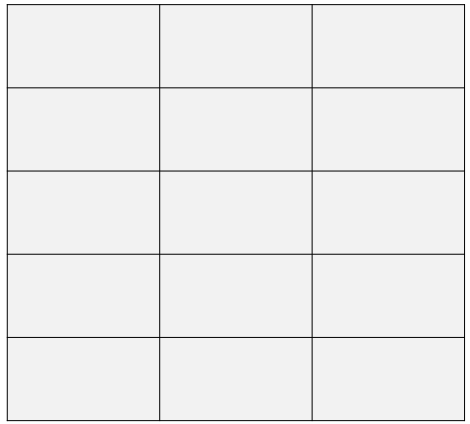
Query

5 x 3

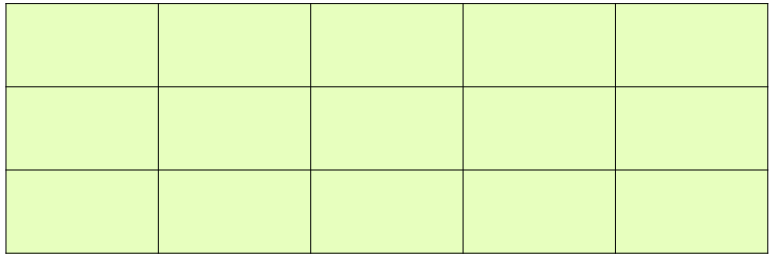
K

Key

Attention



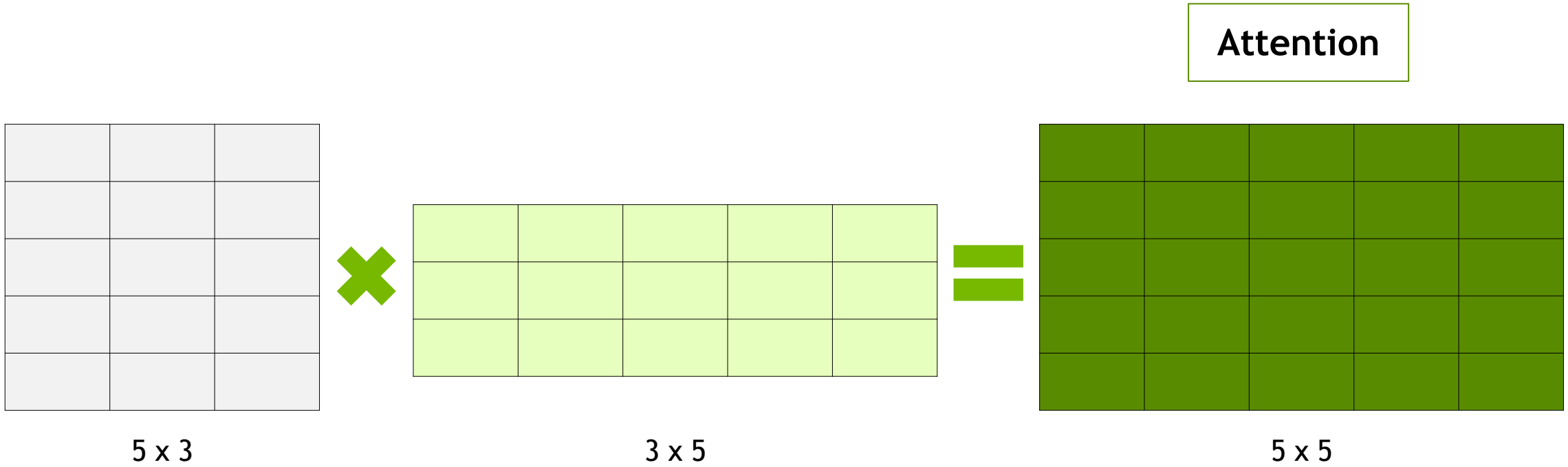
5 x 3



3 x 5

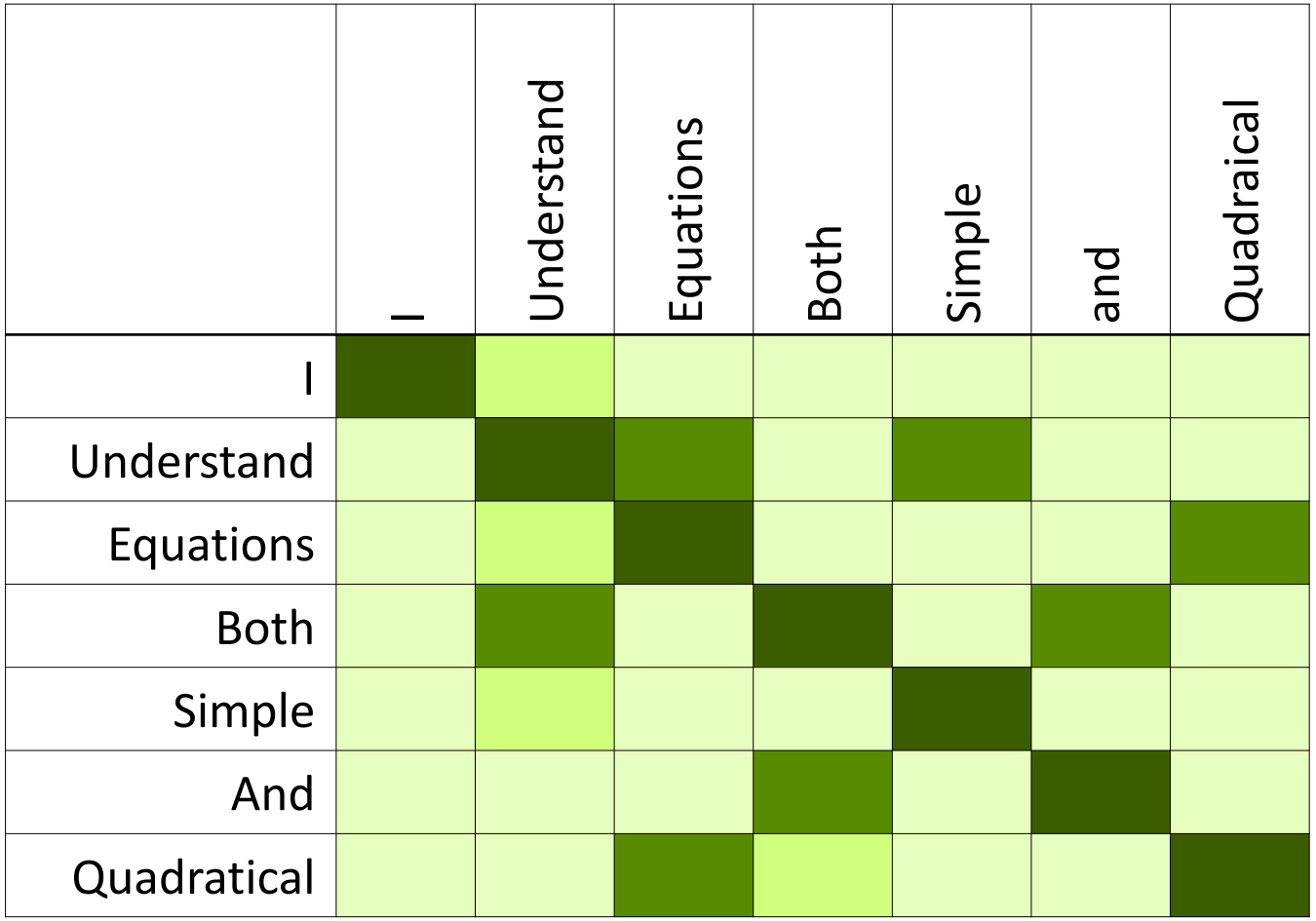
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention



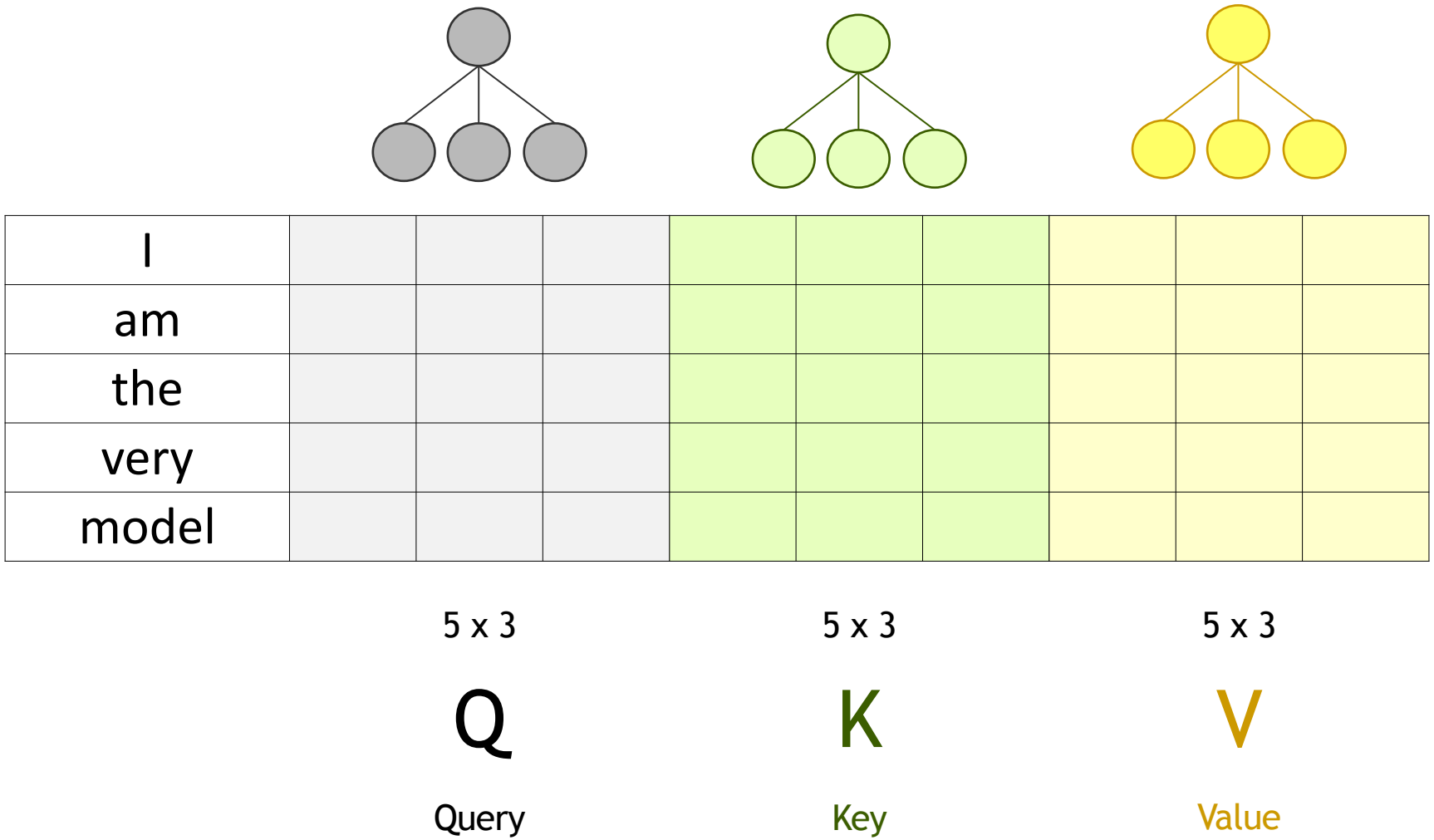
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Attention

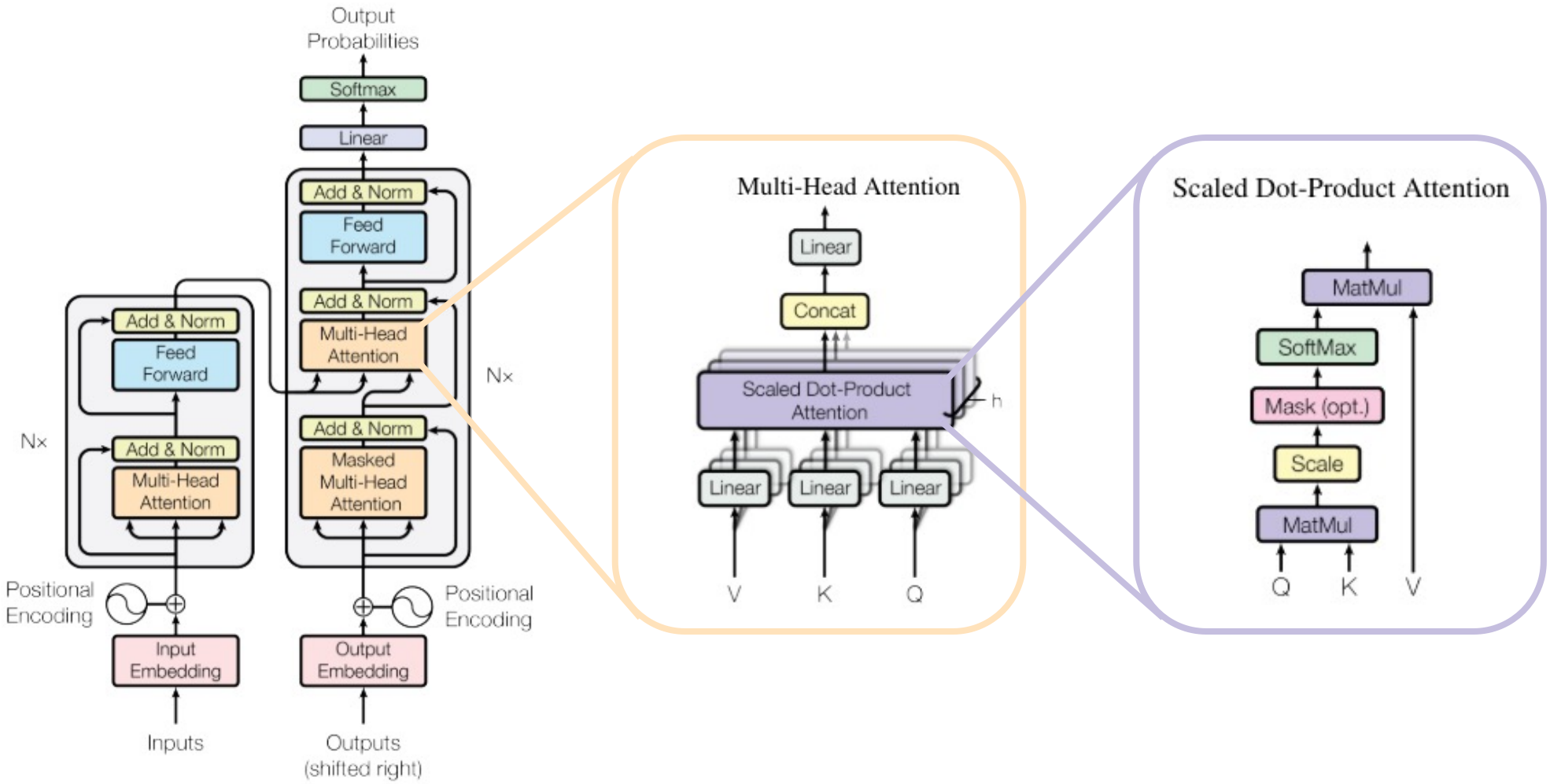
$$Z = softmax\left(\frac{Q \times K^T}{\sqrt{d_k}}\right)V$$

I												
am												
the												
very												
model												

5 x 3	5 x 3	5 x 3	5 x 3
Q	K	V	Z
Query	Key	Value	Attention Score

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Transformers



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

BERT

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin **Ming-Wei Chang** **Kenton Lee** **Kristina Toutanova**
Google AI Language
{jacobdevlin,mingweichang,kentonl,kristout}@google.com

Abstract

We introduce a new language representation model called **BERT**, which stands for **B**idirectional **E**ncoder **R**epresentations from

There are two existing strategies for applying pre-trained language representations to downstream tasks: *feature-based* and *fine-tuning*. The feature-based approach, such as ELMo (Peters

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_m | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

6. Natural Language Processing

In this tutorial, we'll take a detour away from stand-alone pieces of data such as still images, to data that is dependent on other data items in a sequence. For our example, we'll use text sentences. Language is naturally composed of sequence data, in the form of characters in words, and words in sentences. Other examples of sequence data include stock prices and weather data over time. Videos, while containing still images, are also sequences. Elements in the data have a relationship with what comes before and what comes after, and this fact requires a different approach.

6.1 Objectives

- Use a tokenizer to prepare text for a neural network
- See how embeddings are used to identify numerical features for text data

6.2 BERT

Trash

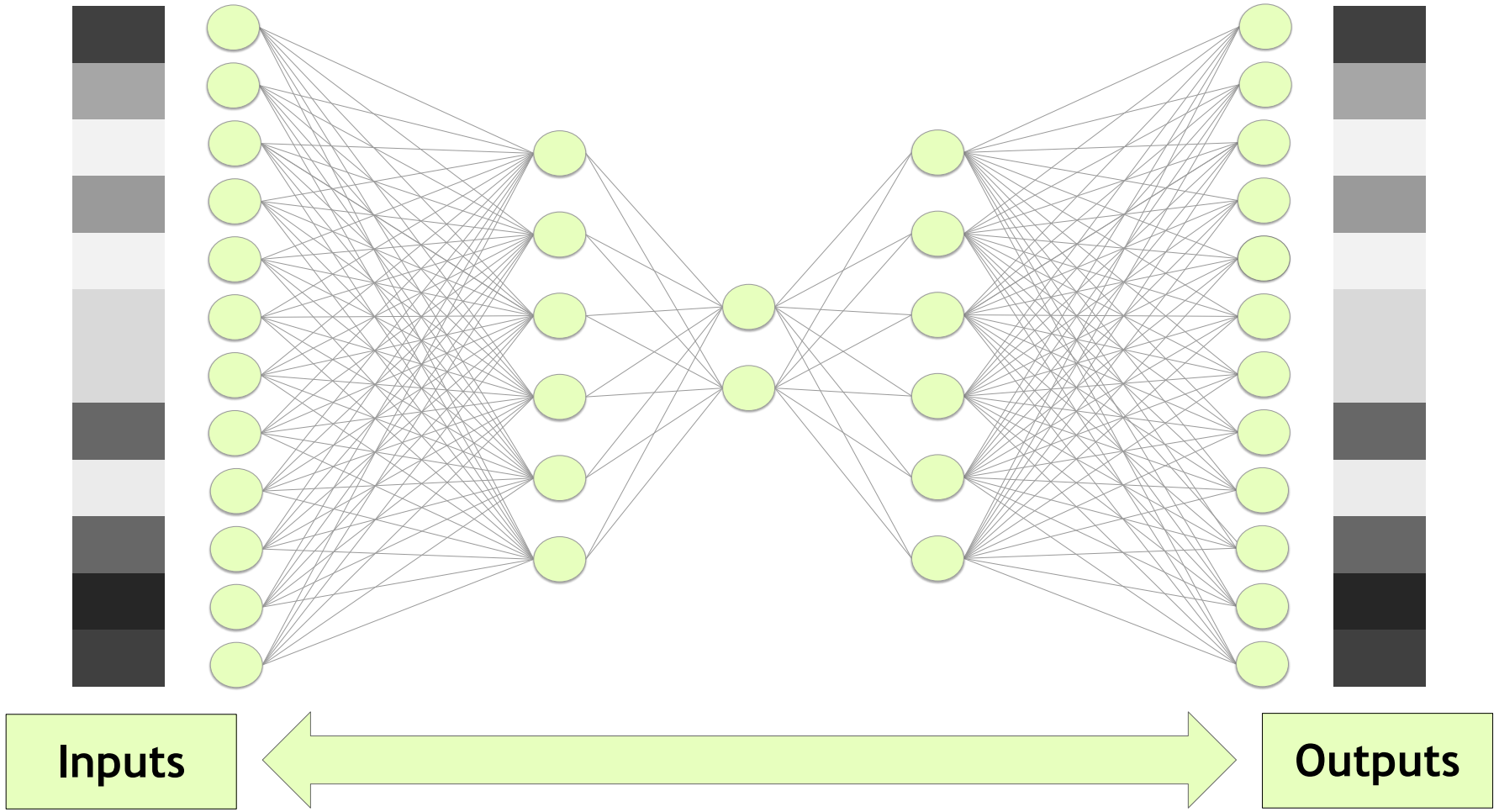
Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 06_nlp.ipynb

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Other Architectures

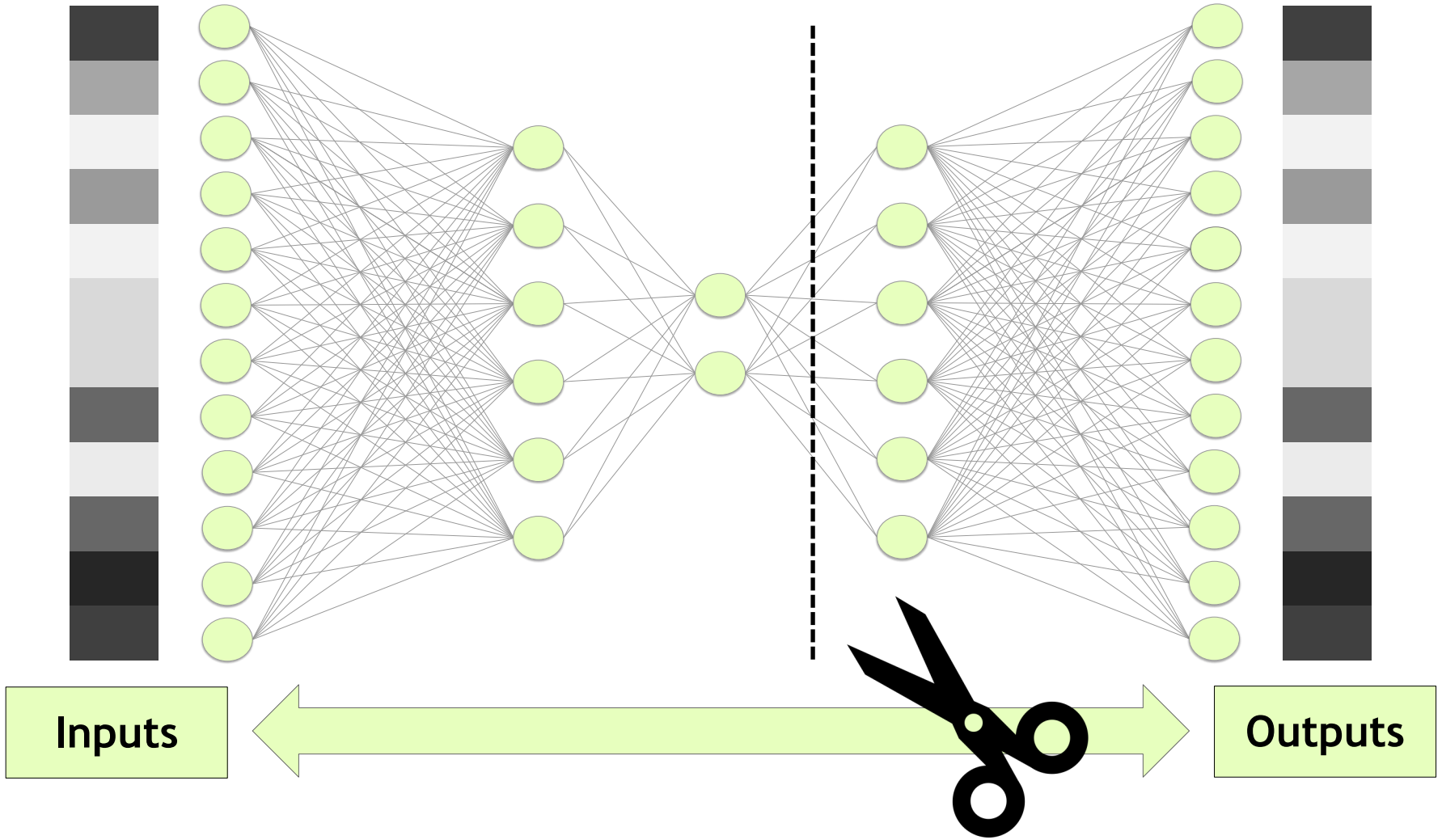
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Autoencoders



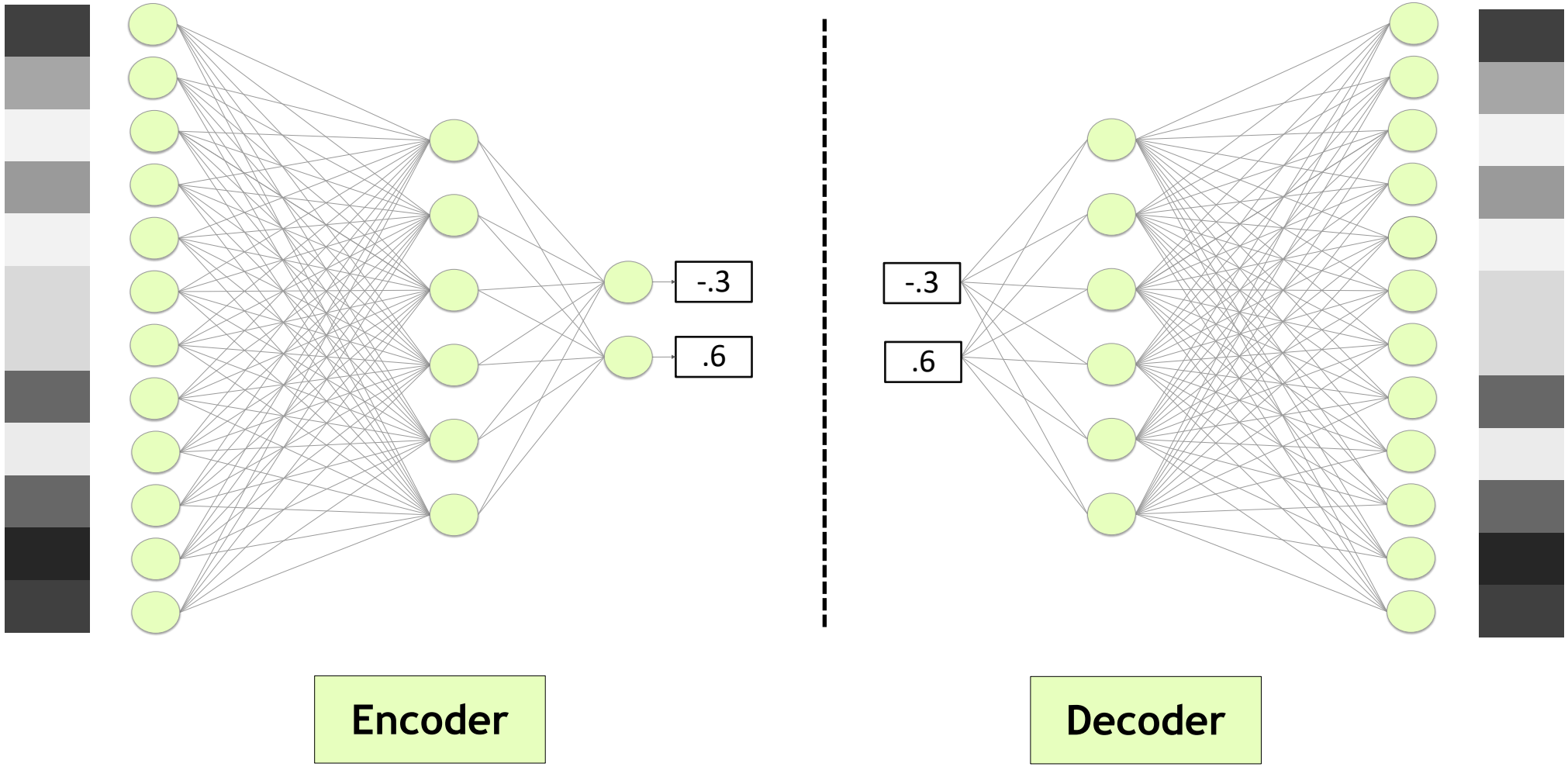
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Autoencoders



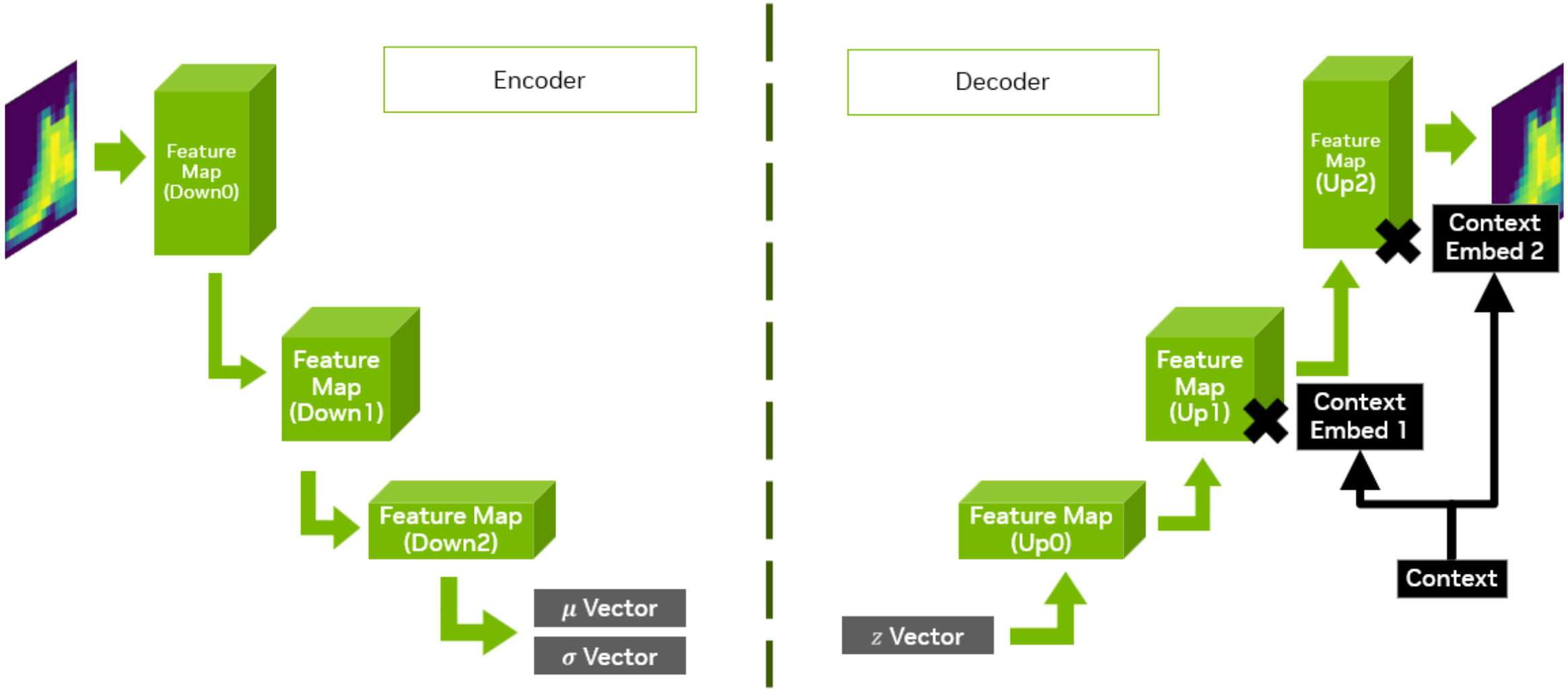
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Autoencoders



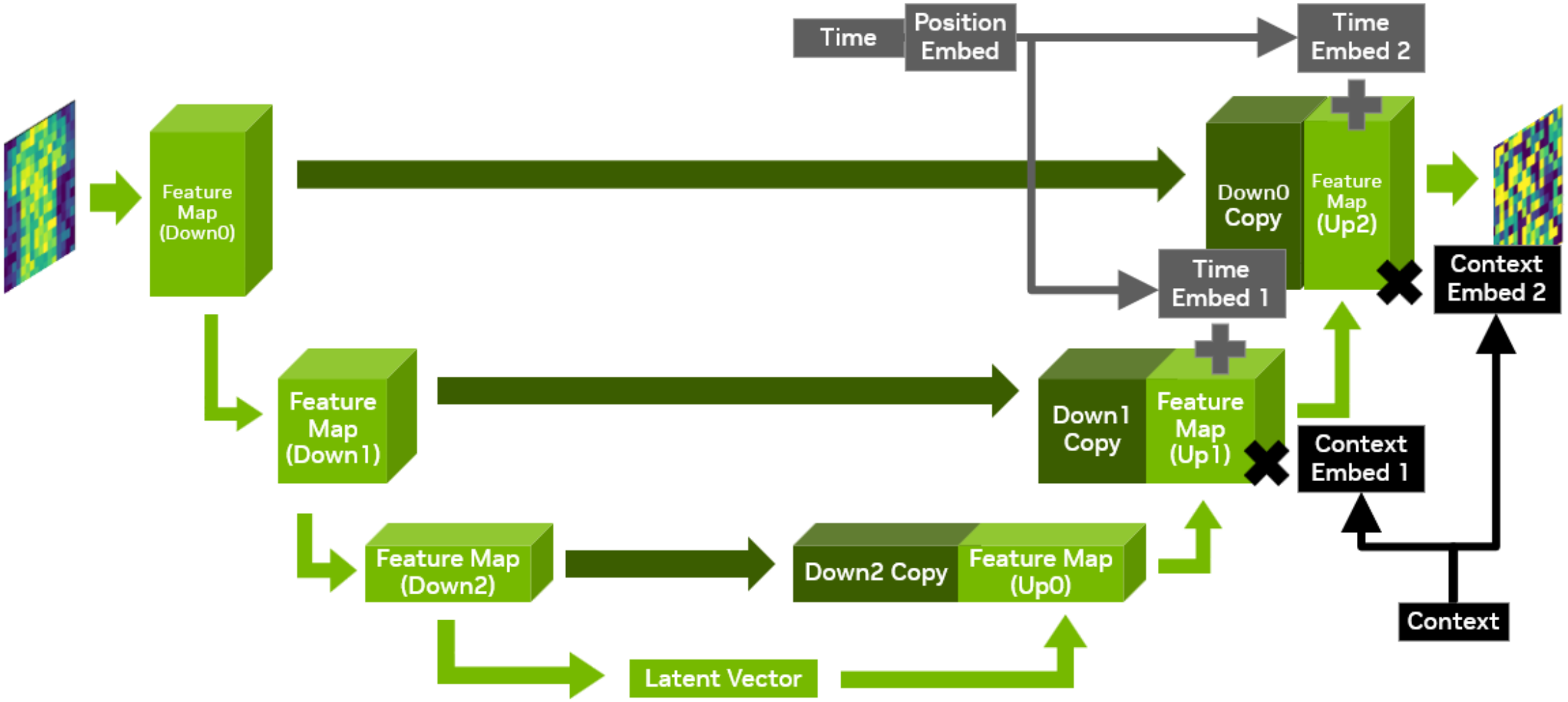
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Variational Autoencoder



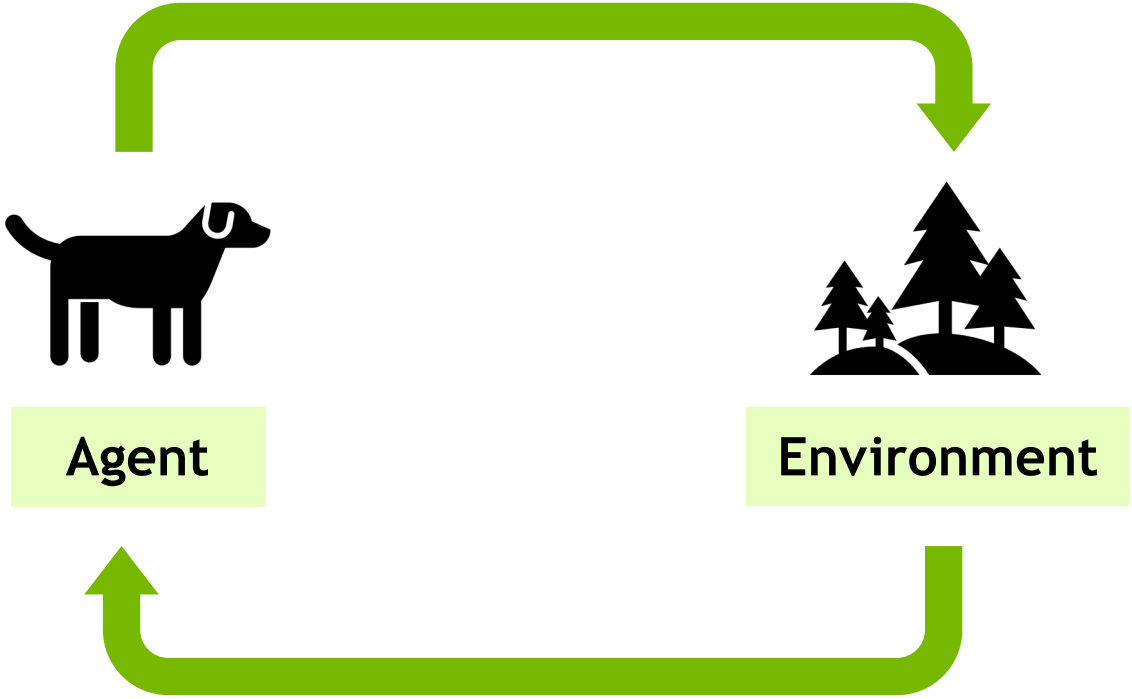
Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Diffusion Models



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Reinforcement Learning



Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Next Steps

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

ENABLING PORTABILITY WITH NGC CONTAINERS

Extensive

- Diverse range of workloads and industry specific use cases

Optimized

- DL containers updated monthly
- Packed with latest features and superior performance

Secure & Reliable

- Scanned for vulnerabilities and crypto
- Tested on workstations, servers, & cloud instances

Scalable

- Supports multi-GPU & multi-node systems

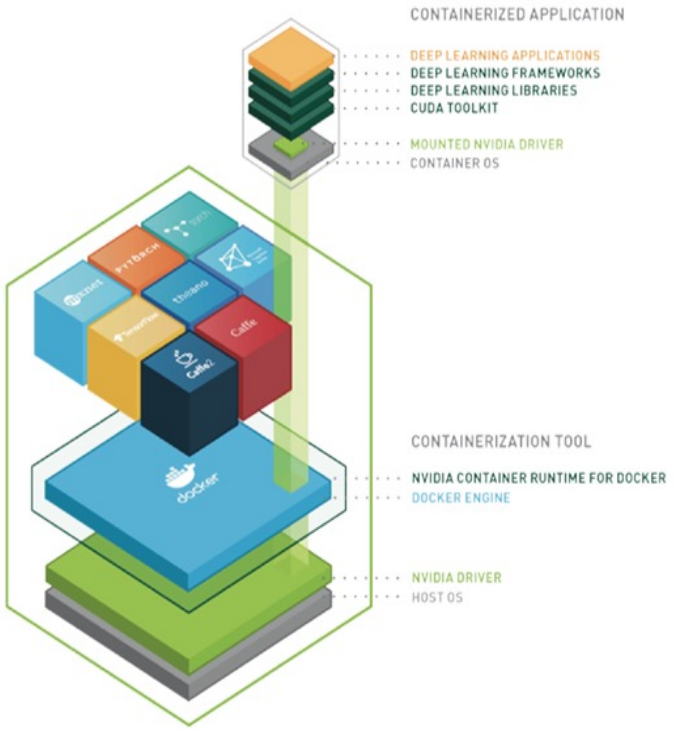
Designed for Enterprise & HPC

- Supports Docker, Singularity & other runtimes

Run Anywhere

- Bare metal, VMs, Kubernetes
- x86, ARM, POWER
- Multi-cloud, on-prem, hybrid, edge

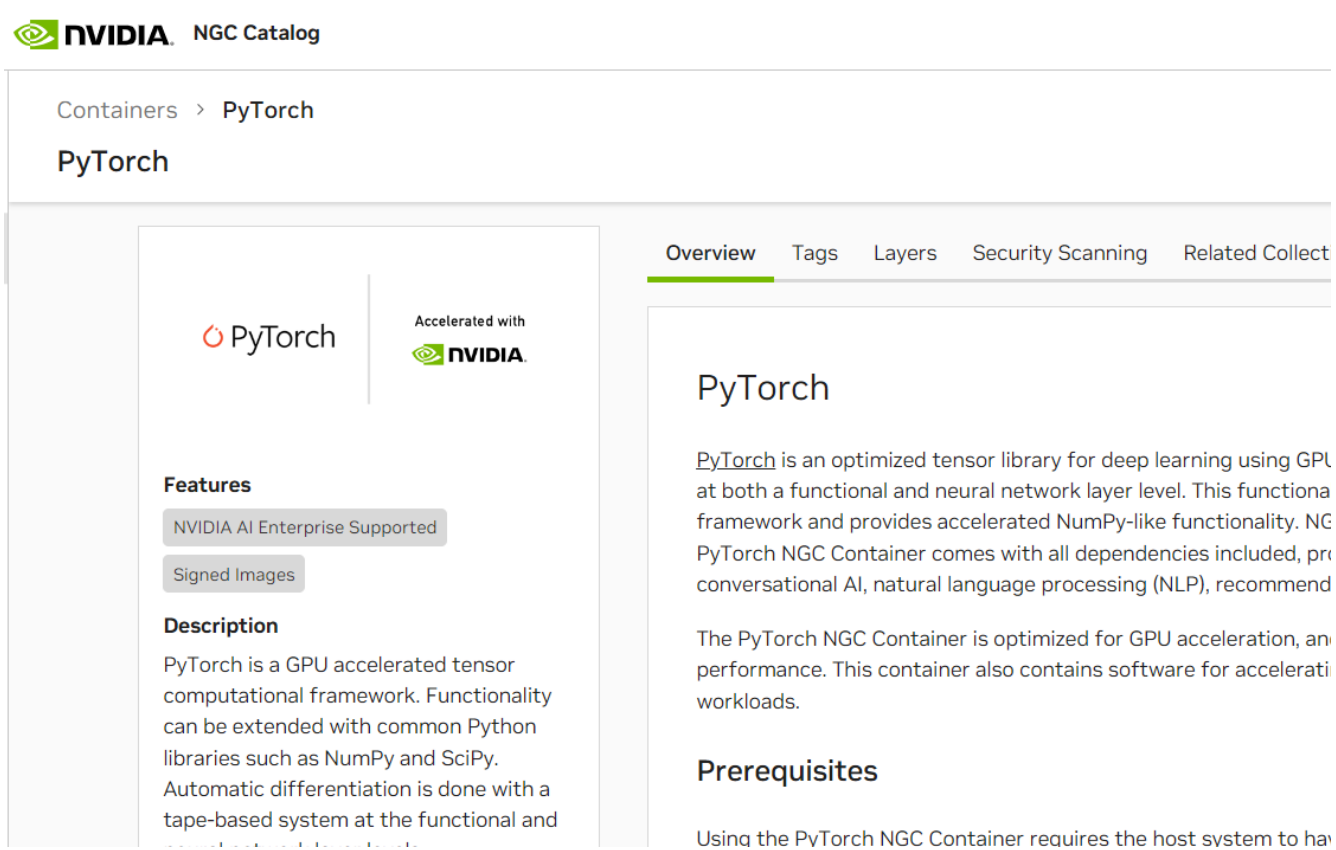
NGC Deep Learning Containers



Learn more about NGC Containers

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Next Steps for This Class



Step 1 Sign up for NGC

https://docs.nvidia.com/dgx/ngc-registry-for-dgx-user-guide/index.html

Step 2 Visit NGC Catalog

https://catalog.ngc.nvidia.com/orgs/nvidia/containers/pytorch

Step 3 Pull and Run Container

Visit localhost:8888 to check out a JupyterLab environment

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Closing Thoughts

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Copying Rocket Science

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

Let’s get Started!

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name

- data
- images
- 00_jupyterlab.ipynb
- 01_mnist.ipynb
- 02_asl.ipynb
- 03_asl_cnn.ipynb
- 04a_asl_augmentation.ipynb
- 04b_asl_predictions.ipynb
- 05a_doggy_door.ipynb
- 05b_presidential_doggy_door.ipynb
- 06_nlp.ipynb
- 07_assessment.ipynb
- utils.py

Laun | 00_j | 01_n | 02_a | 03_a | 04a_ | 04b_ | 05a_ | 05b_ | 06_r | 07_a | utils.

Markdown | Python 3 (ipykernel)

NVIDIA. DEEP LEARNING INSTITUTE

7. Assessment

https://tinyurl.com/nvdlifdls

Congratulations on going through today's course! Hopefully, you've learned some valuable skills along the way and had fun doing it. Now it's time to put those skills to the test. In this assessment, you will train a new model that is able to recognize fresh and rotten fruit. You will need to get the model to a validation accuracy of `92%` in order to pass the assessment, though we challenge you to do even better if you can. You will have the use the skills that you learned in the previous exercises. Specifically, we suggest using some combination of transfer learning, data augmentation, and fine tuning. Once you have trained the model to be at least 92% accurate on the validation dataset, save your model, and then assess its accuracy. Let's get started!

[ ]:
```
import torch
import torch.nn as nn
from torch.optim import Adam
from torch.utils.data import Dataset, DataLoader
import torchvision.transforms.v2 as transforms
import torchvision.io as tv_io

import glob
```

Simple 0 10 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 07_assessment.ipynb 0

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

NVIDIA Getting Started with Deep Learning

Fruits fresh and rotten for classification

Kaggle Datasets: https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification

Source: NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1

References

- NVIDIA DLI (2025), Getting Started with Deep Learning, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-01+V1
- NVIDIA DLI (2025), Building RAG Agents with LLMs, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-15+V1
- NVIDIA DLI (2025), Generative AI with Diffusion Models, https://learn.nvidia.com/courses/course-detail?course_id=course-v1:DLI+S-FX-14+V1