Financial Econometrics and AI-First Finance

1121AIFQA06
MBA, IM, NTPU (M5276) (Fall 2023)
Tue 2, 3, 4 (9:10-12:00) (B3F17)

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

Min-Yuh Day, Ph.D,
Associate Professor

Institute of Information Management, National Taipei University

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

2023-11-07; 2023-11--14

Syllabus

Week	Date	Subject/Topics
1	2023/09/12	Introduction to Artificial Intelligence in Finance and Quantitative Analysis
2	2023/09/19	AI in FinTech: Metaverse, Web3, DeFi, NFT, Financial Services Innovation and Applications
3	2023/09/26	Investing Psychology and Behavioral Finance
4	2023/10/03	Event Studies in Finance
5	2023/10/10	National Day (Day off)
6	2023/10/17	Case Study on AI in Finance and Quantitative Analysis I

Syllabus

Week	Date	Subject/Topics
7	2023/10/24	Finance Theory and Data-Driven Finance
8	2023/10/31	Midterm Project Report
9	2023/11/07	Financial Econometrics
10	2023/11/14	AI-First Finance
11	2023/11/21	Industry Practices of AI in Finance and Quantitative Analysis
12	2023/11/28	Case Study on AI in Finance and Quantitative Analysis II

Syllabus

Week	Date	Subject/Topics
13	2023/12/05	Deep Learning in Finance; Reinforcement Learning in Finance
14	2023/12/12	Algorithmic Trading; Risk Management; Trading Bot and Event-Based Backtesting
15	2023/12/19	Final Project Report I
16	2023/12/26	Final Project Report II
17	2024/01/02	Self-study
18	2024/01/09	Self-study

Financial Econometrics and AI-First Finance

Outline

- Financial Econometrics
  - Financial Theories, OLS Regression
- Machine Learning
  - Learning, Evaluation, Bias and variance
  - Cross-validation
- AI-First Finance
  - Efficient Markets
  - Market Prediction Based on Returns Data
  - Market Prediction with More Features

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

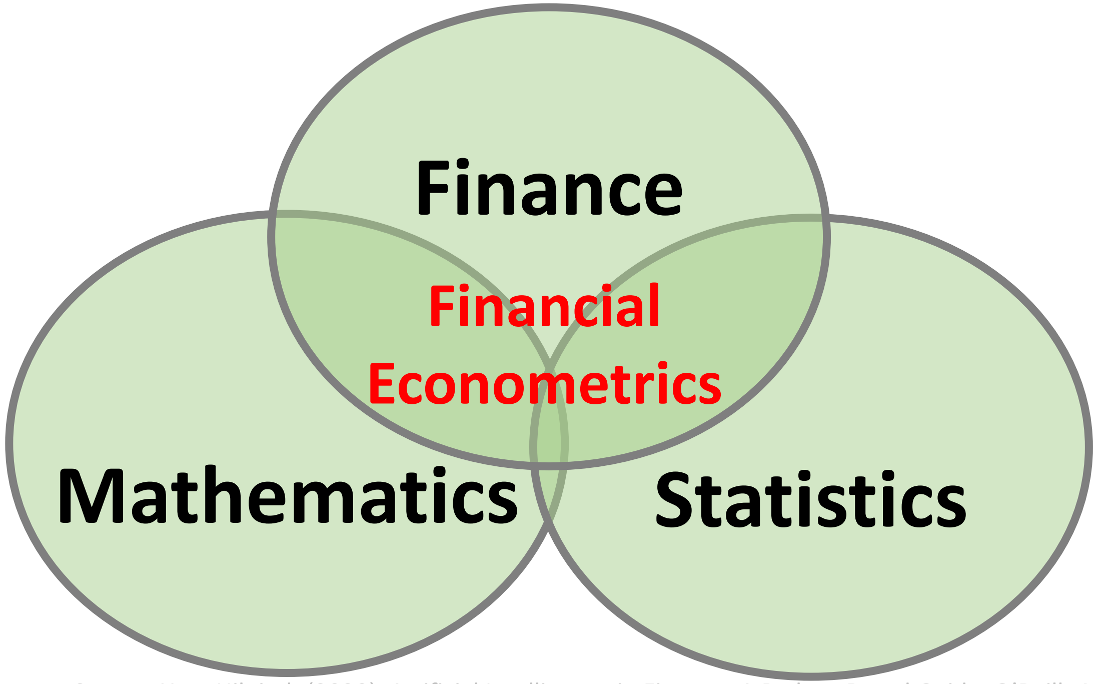
Financial Econometrics

Financial Econometrics

- **Financial Theories**
- **OLS Regression**

Financial Econometrics

- **The discipline at the intersection of mathematics, statistics, and finance that applies such methods to financial market data is typically called financial econometrics.**



Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Financial Econometrics

(Chris Brooks, 2019)

- Financial econometrics
  - the application of statistical techniques to problems in finance
- Financial econometrics can be useful for
  testing theories in finance,
  determining asset prices or returns,
  testing hypotheses concerning the relationships between variables,
  examining the effect on financial markets of changes in economic conditions,
  forecasting future values of financial variables and for financial decision-making.

Source: Chris Brooks (2019), Introductory Econometrics for Finance, 4th Edition, Cambridge University Press

Financial Econometrics

- [Financial] econometrics is the quantitative application of statistical and mathematical models using [financial] data to develop financial theories or test existing hypotheses in finance and to forecast future trends from historical data.
- It subjects real-world [financial] data to statistical trials and then compares and contrasts the results against the [financial] theory or theories being tested.

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Topics of Financial Econometrics

(Oliver Linton, 2019)

1. Econometric
2. Return Predictability and the Efficient Markets Hypothesis
3. Robust Tests and Tests of Nonlinear Predictability of Returns
4. Empirical Market Microstructure
5. Event Study Analysis
6. Portfolio Choice and Testing the Capital Asset Pricing Model
7. Multifactor Pricing Models

Source: Oliver Linton (2019), Financial Econometrics: Models and Methods, Cambridge University Press

Topics of Financial Econometrics

(Oliver Linton, 2019)

8. Present Value Relations
9. Intertemporal Equilibrium Pricing
10. Volatility
11. Continuous Time Processes
12. Yield Curve
13. Risk Management and Tail Estimation

Source: Oliver Linton (2019), Financial Econometrics: Models and Methods, Cambridge University Press

Applications of Financial Econometrics

(Chris Brooks, 2019)

1. **Testing whether financial markets are weak-form informationally efficient**
2. **Testing whether the capital asset pricing model (CAPM) or arbitrage pricing theory (APT) represent superior models for the determination of returns on risky assets**
3. **Measuring and forecasting the volatility of bond returns**
4. **Explaining the determinants of bond credit ratings used by the ratings agencies**
5. **Modelling long-term relationships between prices and exchange rates**

Source: Chris Brooks (2019), Introductory Econometrics for Finance, 4th Edition, Cambridge University Press

Applications of Financial Econometrics

(Chris Brooks, 2019)

6. Determining the optimal hedge ratio for a spot position in oil
7. Testing technical trading rules to determine which makes the most money
8. Testing the hypothesis that earnings or dividend announcements have no effect on stock prices
9. Testing whether spot or futures markets react more rapidly to news
10. Forecasting the correlation between the stock indices of two countries

Source: Chris Brooks (2019), Introductory Econometrics for Finance, 4th Edition, Cambridge University Press

Machine Learning and Financial Econometrics

- **ML and DL methods** are able to **discover statistical inefficiencies** and even **economic inefficiencies** that are not discoverable by **traditional econometric methods**, such as multivariate OLS regression.

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Normative Financial Theories

- **Normative financial theories** mostly rely on assumptions and axioms in combination with deduction as the major analytical method to arrive at their central results.
  - **Expected utility theory (EUT)** assumes that agents have the same utility function no matter what state of the world unfolds and that they maximize expected utility under conditions of uncertainty.
  - **Mean-variance portfolio (MVP)** theory describes how investors should invest under conditions of uncertainty assuming that only the expected return and the expected volatility of a portfolio over one period count.

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Normative Financial Theories

- The capital asset pricing model (CAPM) assumes that only the nondiversifiable market risk explains the expected return and the expected volatility of a stock over one period.
- Arbitrage pricing theory (APT) assumes that a number of identifiable risk factors explains the expected return and the expected volatility of a stock over time; admittedly, compared to the other theories, the formulation of APT is rather broad and allows for wide-ranging interpretations.

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Financial Econometrics and Regression

- One of the major tools in financial econometrics is regression, in both its univariate and multivariate forms
  - $\boldsymbol{y = \alpha + \beta\, x}$
  - $\boldsymbol{y = \alpha + \beta_1\, x_1 + \beta_2\, x_2}$
  - $\boldsymbol{y = \alpha + \beta_1\, x_1 + \beta_2\, x_2 + \beta_3\, x_3}$
- Regression is also a central tool in statistical learning in general

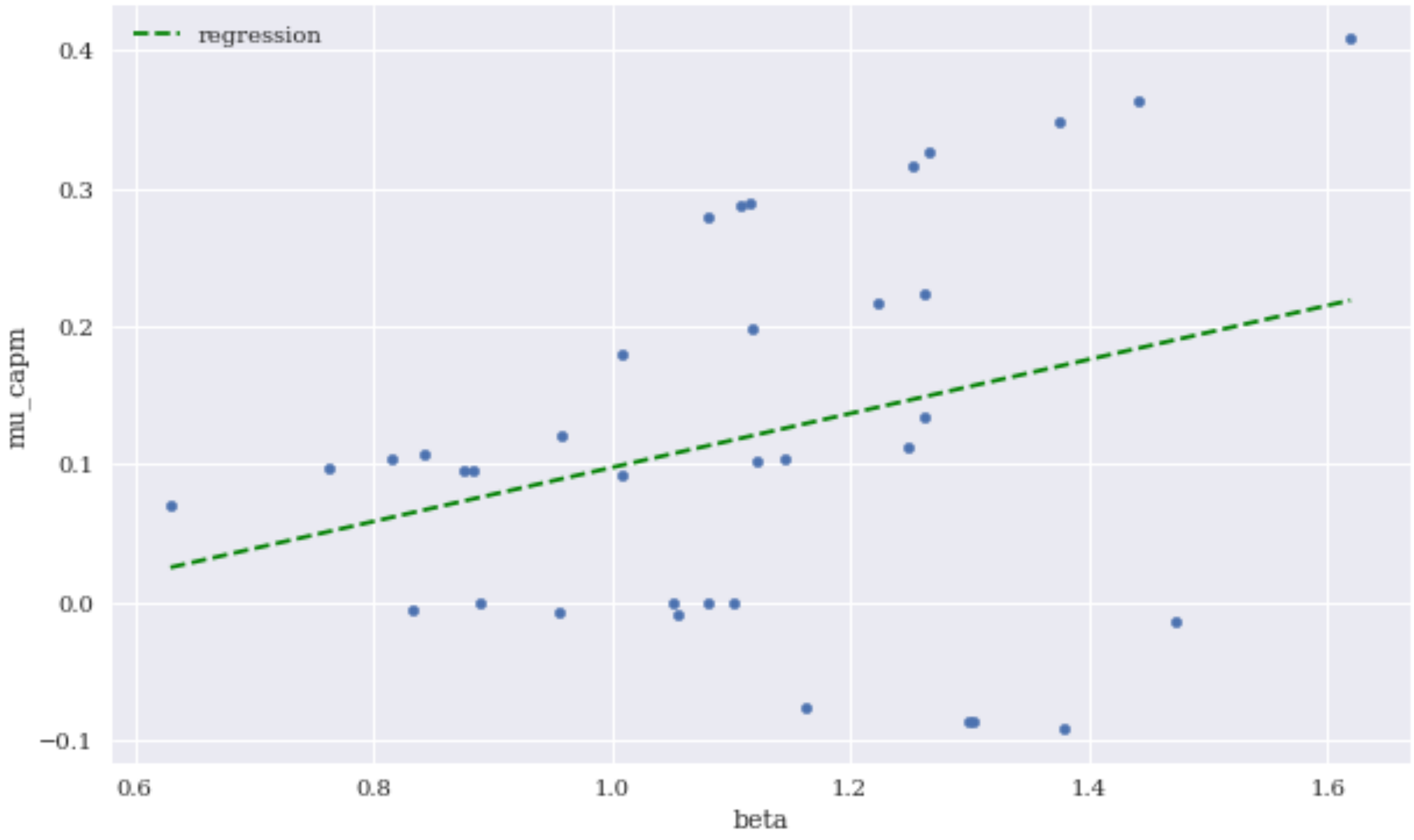
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

CAPM and APT
OLS regression

- Both the CAPM and the APT
  relate the output variables
  with the relevant input factors in linear fashion.
- From an econometric point of view,
  both models are implemented based on
  linear ordinary least-squares (OLS) regression.
- CAPM: univariate linear OLS regression
- APT: multivariate OLS regression

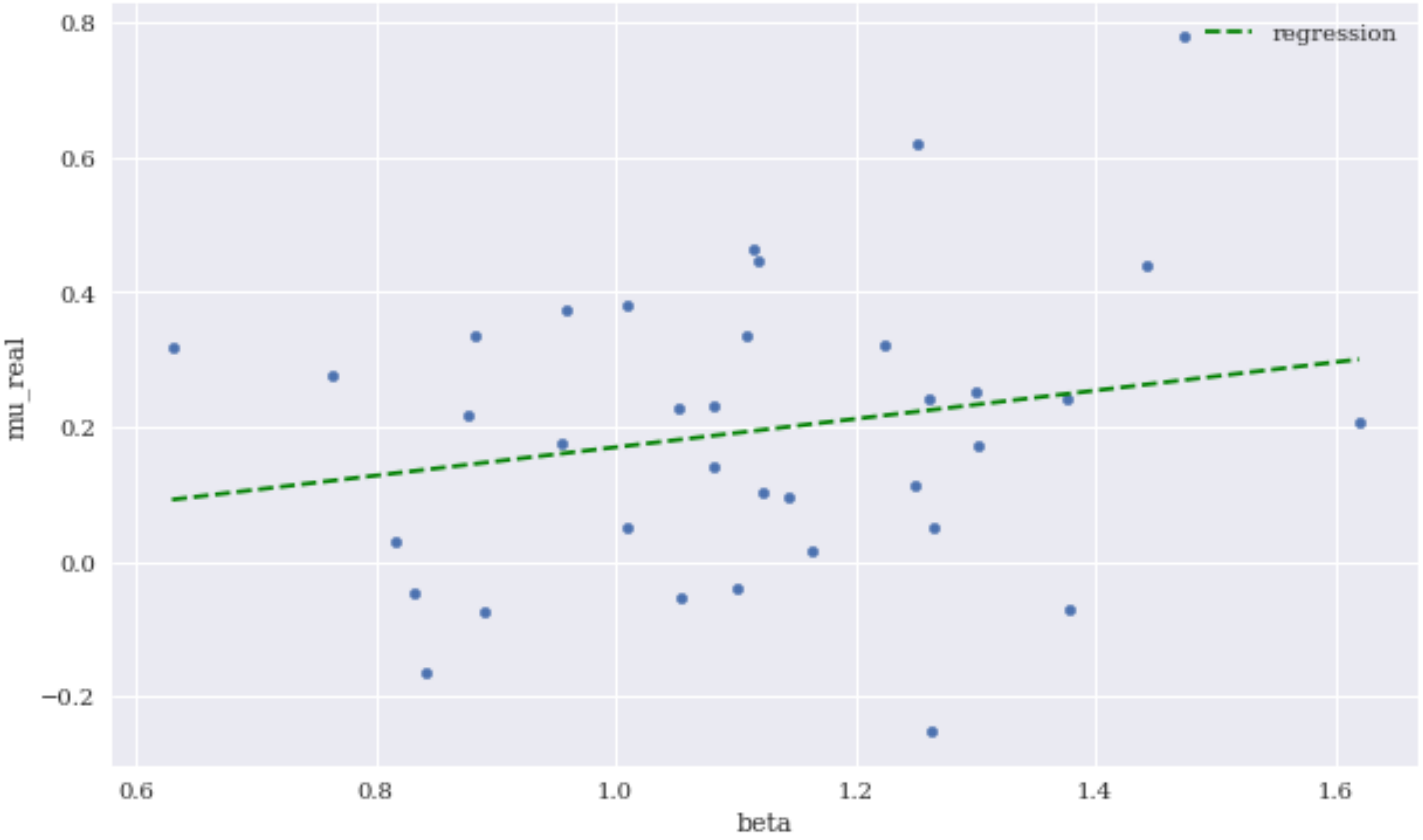
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Expected CAPM return versus beta (including linear regression)



Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Expected CAPM return versus beta (including linear regression)



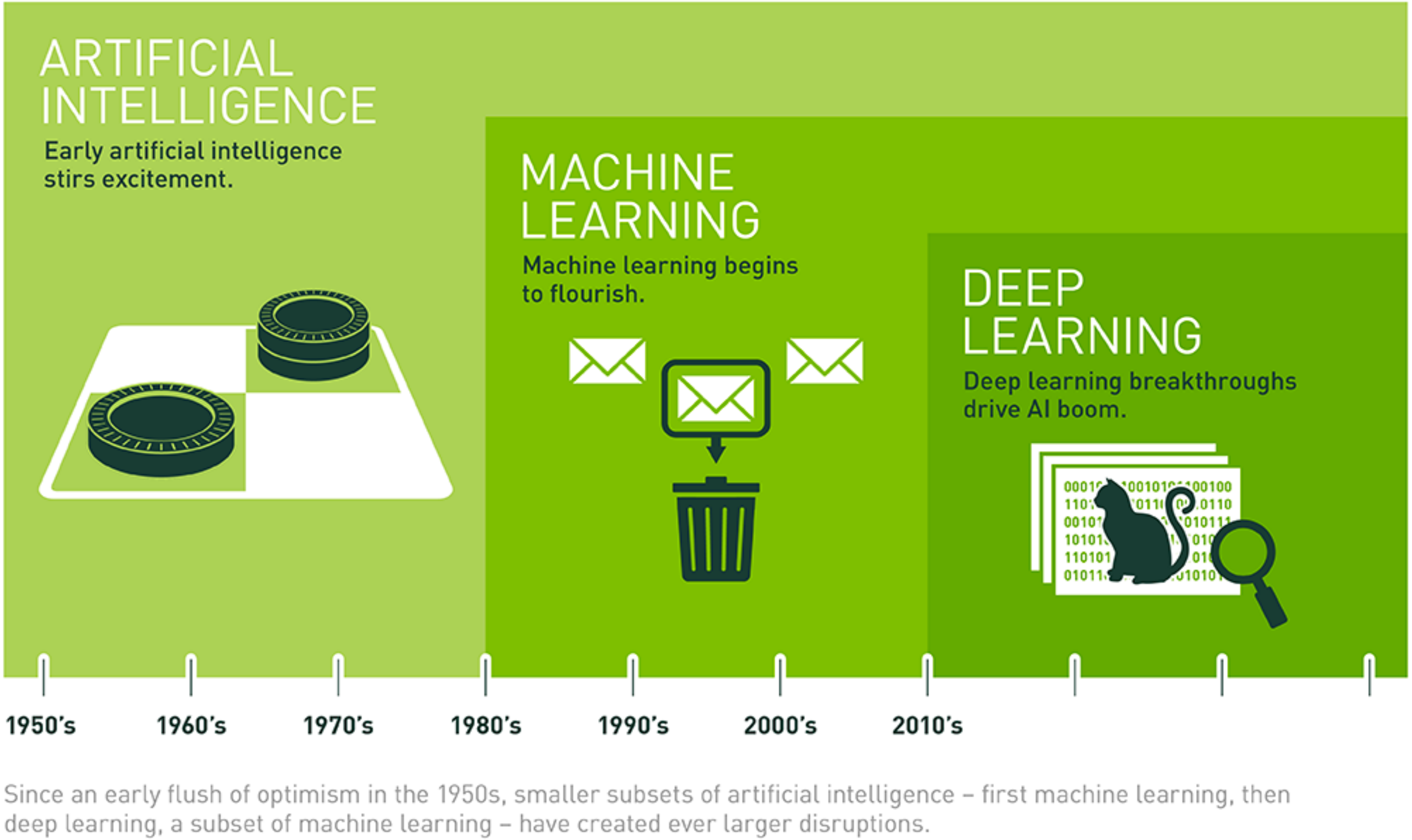
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Machine Learning

Machine Learning

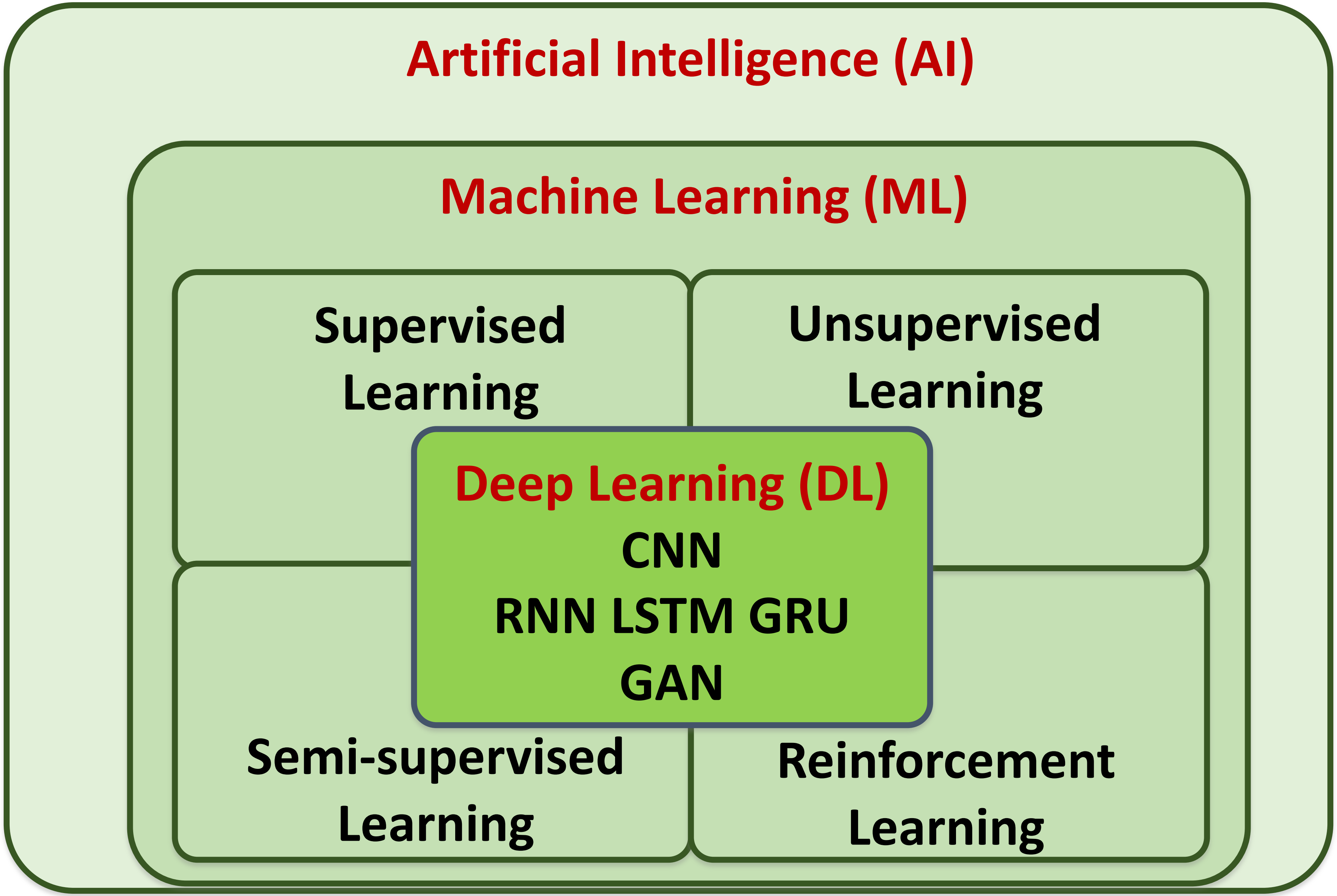
- Learning
- Evaluation
- Bias and variance
- Cross-validation

Artificial Intelligence
Machine Learning & Deep Learning



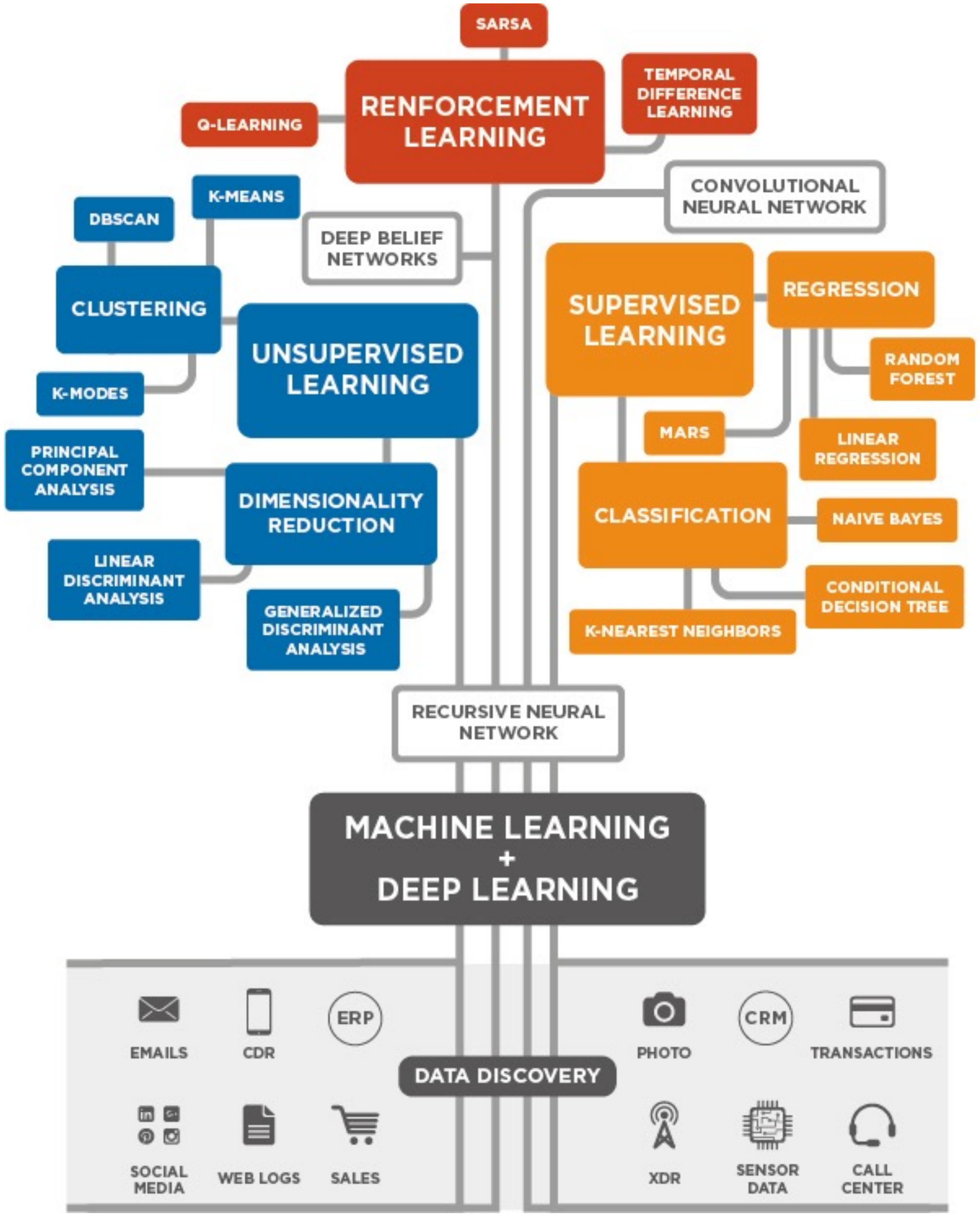
Source: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

AI, ML, DL



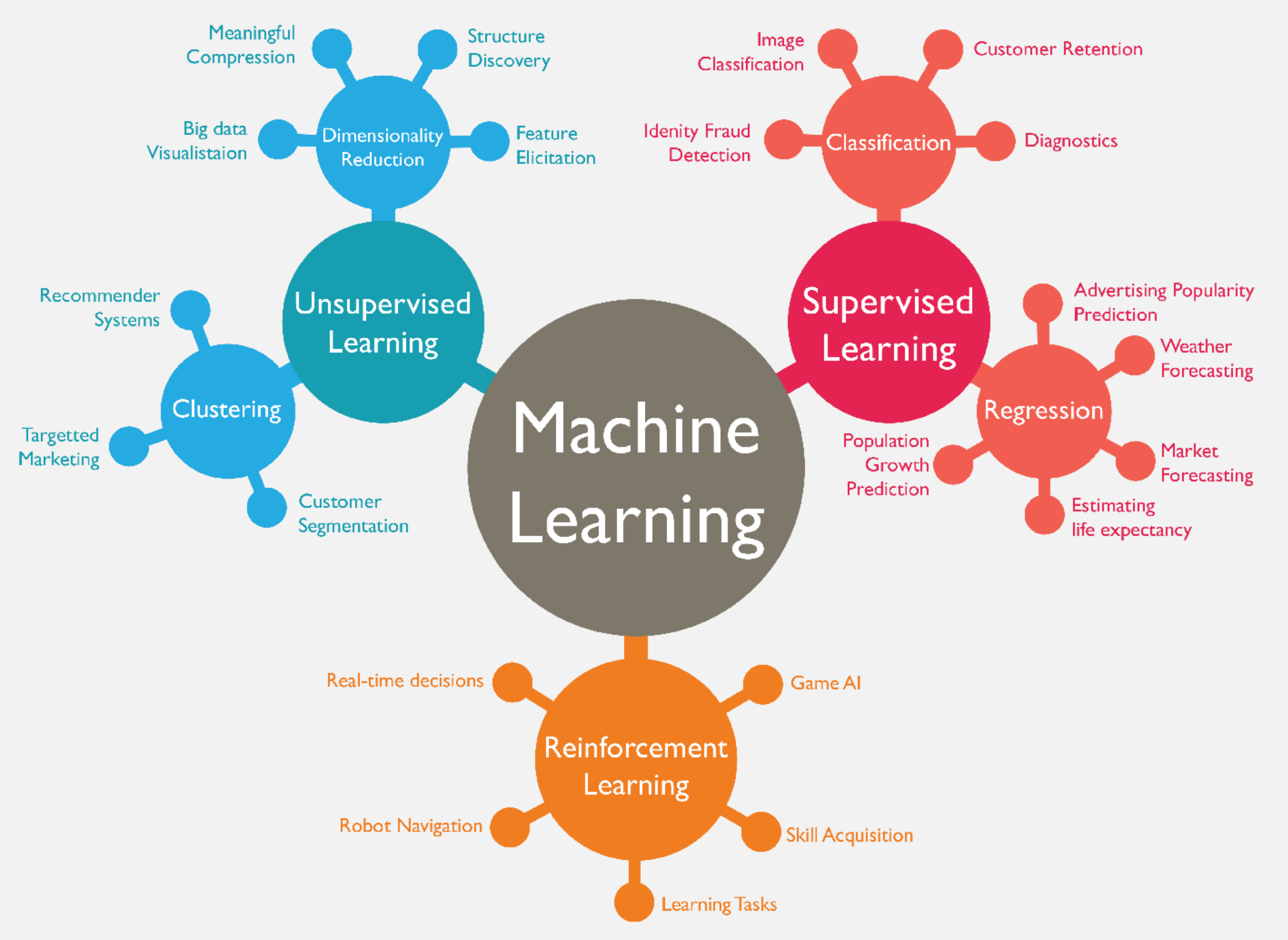
Source: https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/deep_learning.html

3 Machine Learning Algorithms



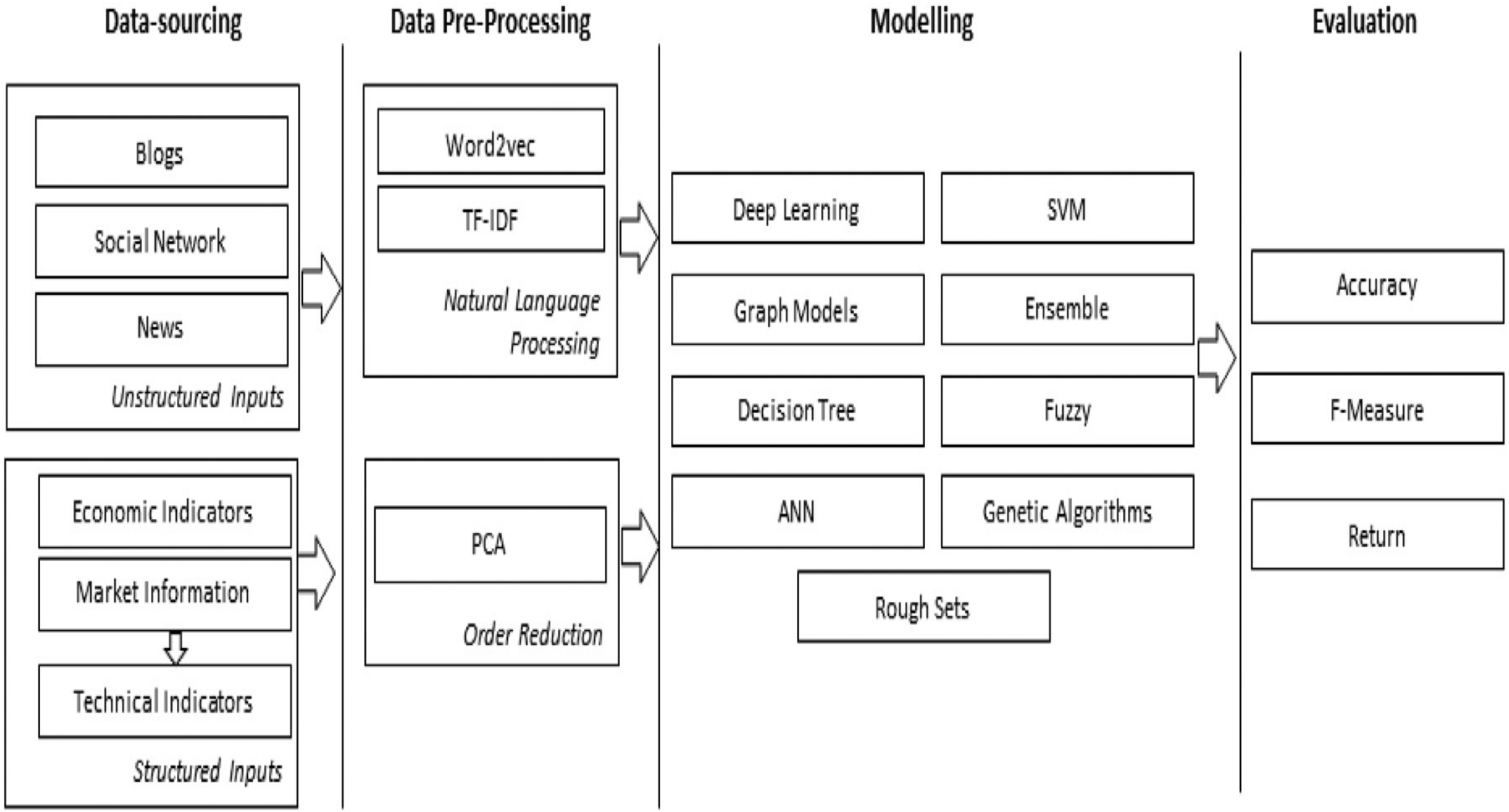
Source: Enrico Galimberti, http://blogs.teradata.com/data-points/tree-machine-learning-algorithms/

Machine Learning (ML)



Source: https://www.mactores.com/services/aws-big-data-machine-learning-cognitive-services/

Stock Market Movement Forecast:
ML Phases of the stock market modeling



Source: O. Bustos and A. Pomares-Quimbaya (2020), "Stock Market Movement Forecast: A Systematic Review." Expert Systems with Applications (2020): 113464.

Machine Learning

- Learning
- Data: Features, Labels
- Success (Loss Function): MSE
- Capacity (Model Fit)
- Evaluation
- Bias and variance
- Cross-validation

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Learning (Mitchell, 1997)

- A computer program is said
  to learn from experience E
  with respect to some class of tasks T
  and performance measure P,
  if its performance at tasks in T,
  as measured by P,
  improves with experience E.

Machine Learning
Supervised Learning (Classification)
Learning from Examples

$$y = f(x)$$

x	y

Machine Learning
Supervised Learning (Classification)
Learning from Examples

$$y = f(x)$$

x	y
input	*Output*
	label

Machine Learning
Supervised Learning (Classification)
Learning from Examples

$$y = f(x)$$

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
6.3,3.3,6.0,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
7.1,3.0,5.9,2.1,Iris-virginica
```

Machine Learning
Supervised Learning (Classification)
Learning from Examples

$$y = f(x)$$

Example

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
6.3,3.3,6.0,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
7.1,3.0,5.9,2.1,Iris-virginica
```

Machine Learning
Supervised Learning (Classification)
Learning from Examples

$$y = f(x)$$

Example

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
6.3,3.3,6.0,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
7.1,3.0,5.9,2.1,Iris-virginica
```

x y

Time Series Data

[10, 20, 30, 40, 50, 60, 70, 80, 90]

X	Y
[10 20 30]	40
[20 30 40]	50
[30 40 50]	60
[40 50 60]	70
[50 60 70]	80
[60 70 80]	90

Time Series Data

[100, 110, 120, 130, 140, 150]

Y

X	Y
[100 110 120 130 140]	150

X_{t1} X_{t2} X_{t3} X_{t4} X_{t5}

The Theory of Learning

- How can we be sure that our learned hypothesis will predict well for previously unseen inputs?
  - How do we know that the hypothesis h is close to the target function f if we don't know what is?
- How many examples do we need to get a good h?
- What hypothesis space should we use?
- If the hypothesis space is very complex,
  can we even find the best h or do we have to settle for a local maximum?
- How complex should h be?
- How do we avoid overfitting?

Linear function

$$y = f(x)$$

$$y = w_1 x + w_0$$

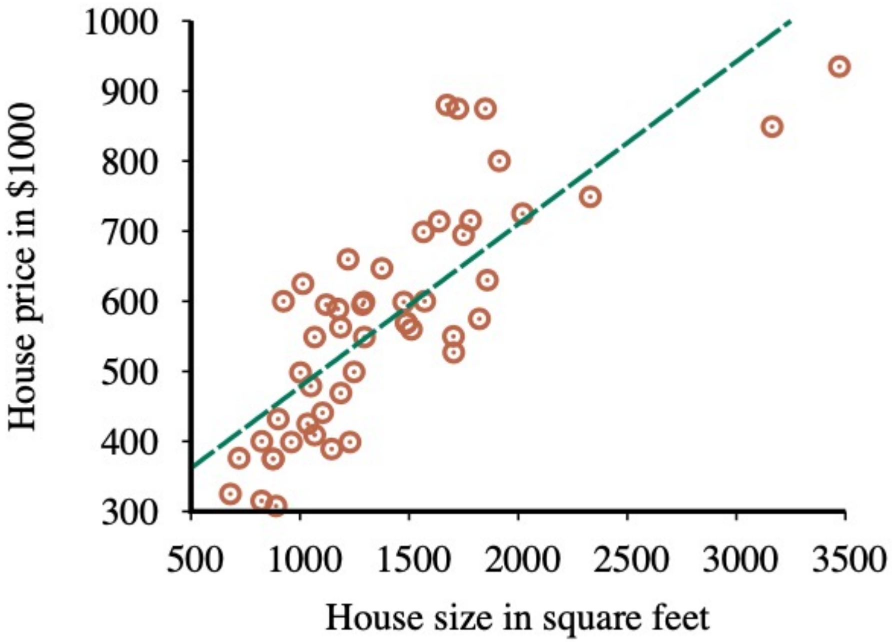
$$h_w(x) = w_1 x + w_0$$

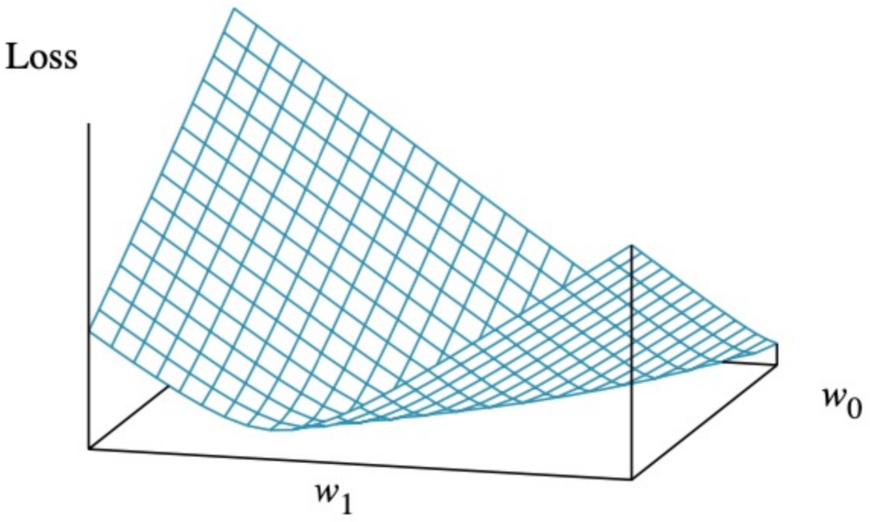
Source: Stuart Russell and Peter Norvig (2020), Artificial Intelligence: A Modern Approach, 4th Edition, Pearson

Linear Regression Weight Space

$$h_w(x) = w_1 x + w_0$$

$$w^* = \mathbf{argmin}_w\, Loss(h_w)$$





$y = 0.232\, x + 246$

Loss function for Weights (w_1, w_0)

Performance Measure

- The measure of success for estimation problems
  - mean-squared error (MSE)
- Classification problems
  - accuracy ratio

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Evaluation

(Accuracy of Classification Model)

Assessing the Classification Model

- **Predictive accuracy**
  - **Hit rate**
- **Speed**
  - **Model building; predicting**
- **Robustness**
- **Scalability**
- **Interpretability**
  - **Transparency, explainability**

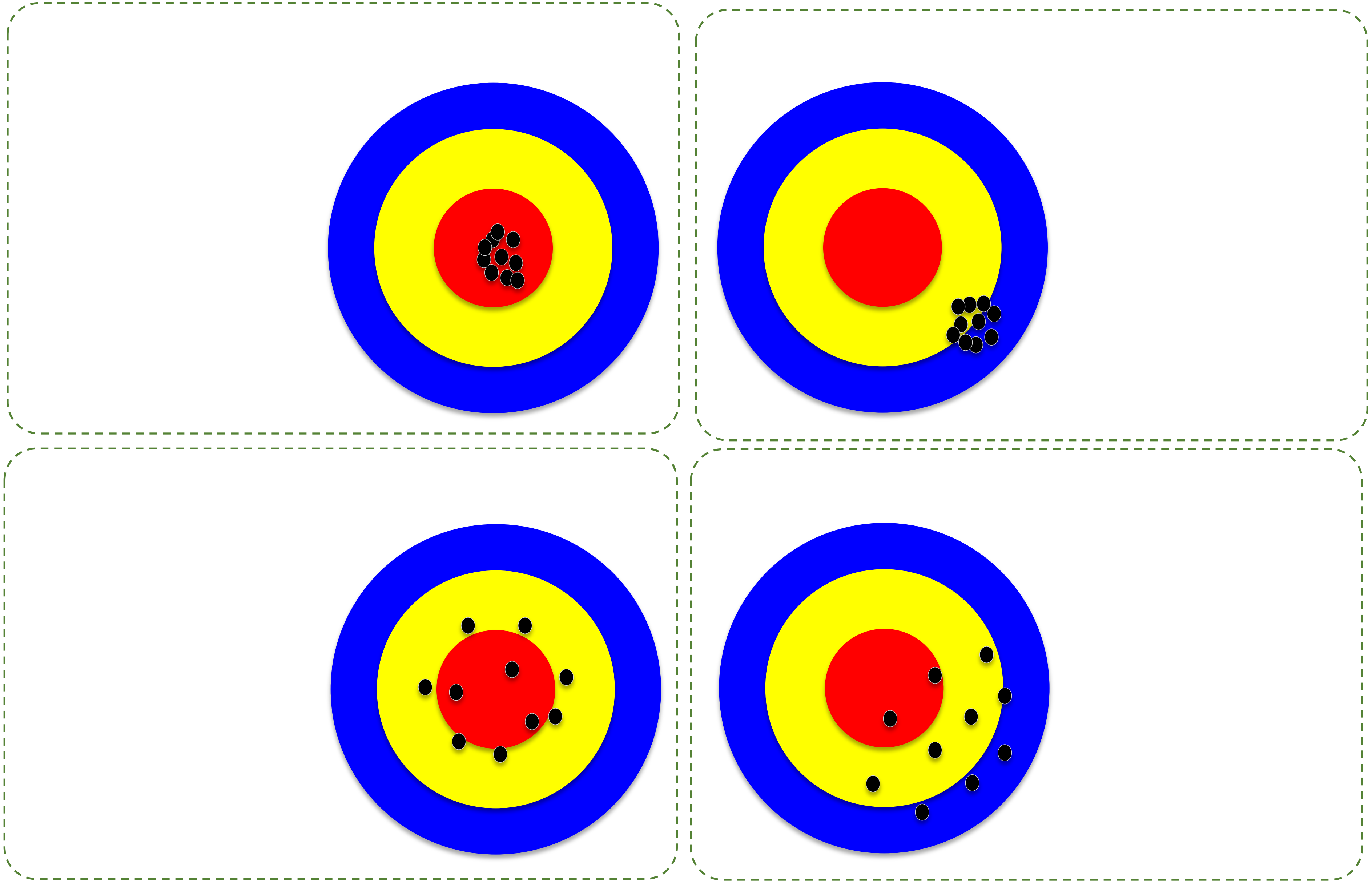
Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Accuracy

Validity

Precision

Reliability



Accuracy vs. Precision

A

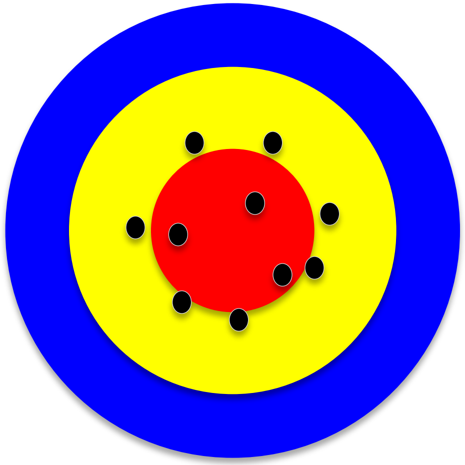
High Accuracy
High Precision

B

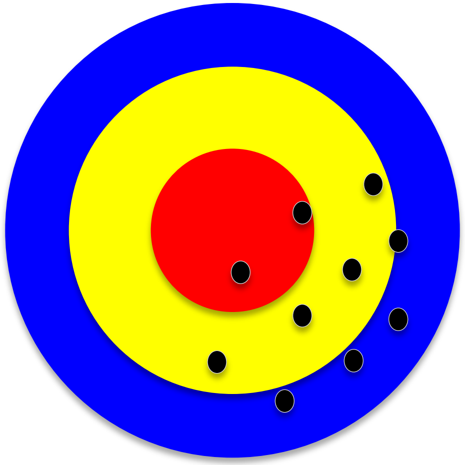
Low Accuracy
High Precision

C

High Accuracy
Low Precision



D



Low Accuracy
Low Precision

Accuracy vs. Precision

A

High Accuracy
High Precision

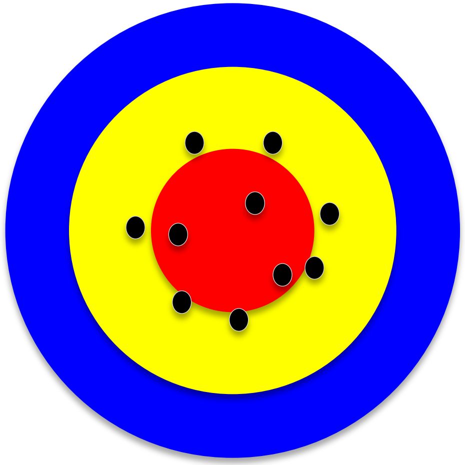
High Validity
High Reliability

B

Low Accuracy
High Precision

Low Validity
High Reliability

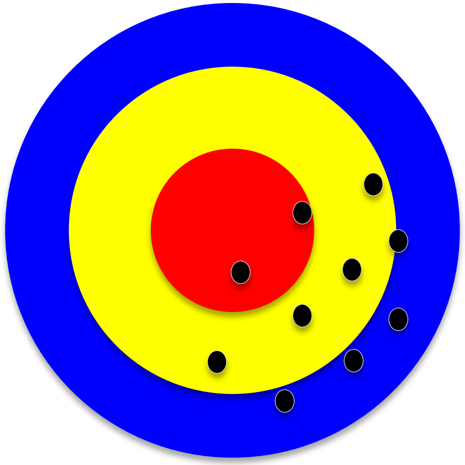
C



High Accuracy
Low Precision

High Validity
Low Reliability

D



Low Accuracy
Low Precision

Low Validity
Low Reliability

Accuracy vs. Precision

A

High Accuracy
High Precision

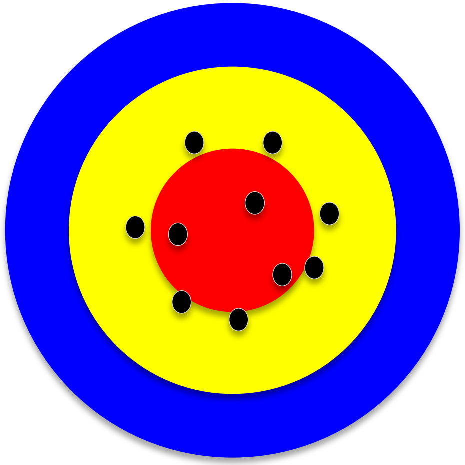
High Validity
High Reliability

B

Low Accuracy
High Precision

Low Validity
High Reliability

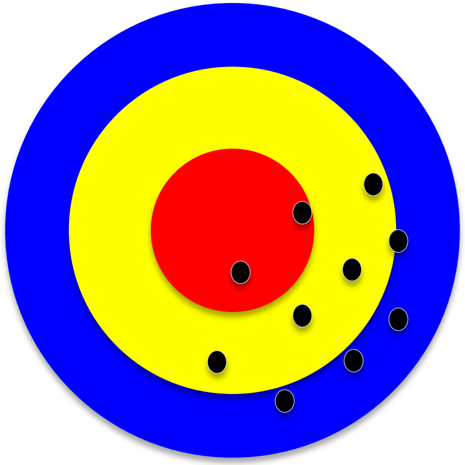
C



High Accuracy
Low Precision

High Validity
Low Reliability

D



Low Accuracy
Low Precision

Low Validity
Low Reliability

Confusion Matrix

for Tabulation of Two-Class Classification Results

		True/Observed Class	
		Positive	**Negative**
Predicted Class	**Positive**	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

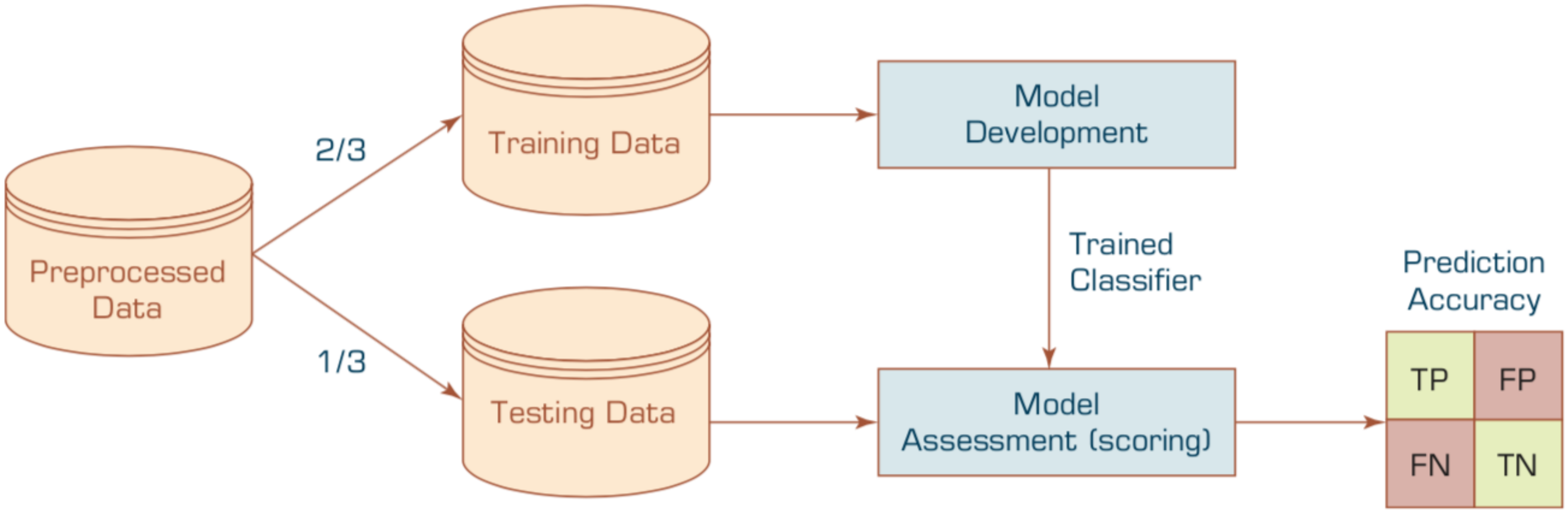
$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

Sensitivity =True Positive Rate

Specificity =True Negative Rate

Estimation Methodologies for Classification

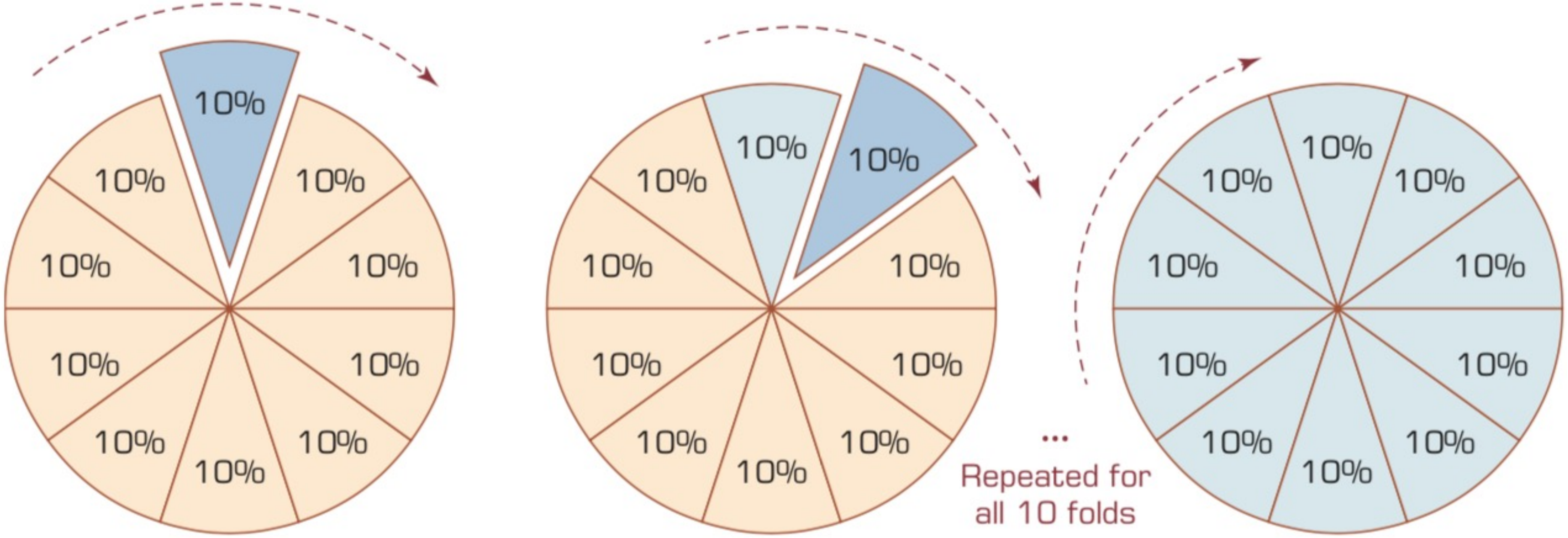
- **Simple split (or holdout or test sample estimation)**
  - **Split the data into 2 mutually exclusive sets training (~70%) and testing (30%)**



  - **For ANN, the data is split into three sub-sets (training [~60%], validation [~20%], testing [~20%])**

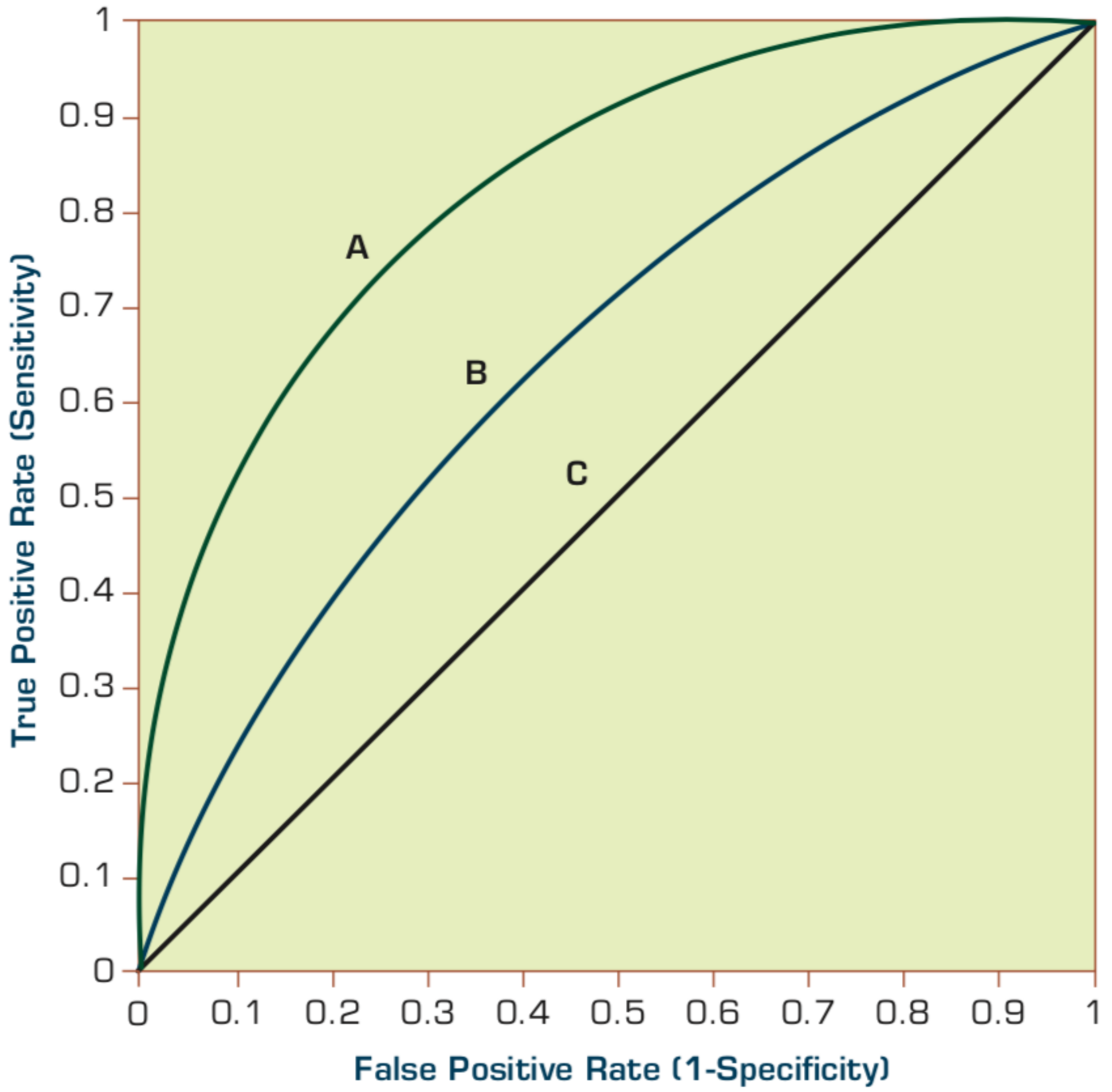
Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

k-Fold Cross-Validation



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

Estimation Methodologies for Classification
Area under the ROC curve



Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson

		True Class (actual value)		
		Positive	Negative	total
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

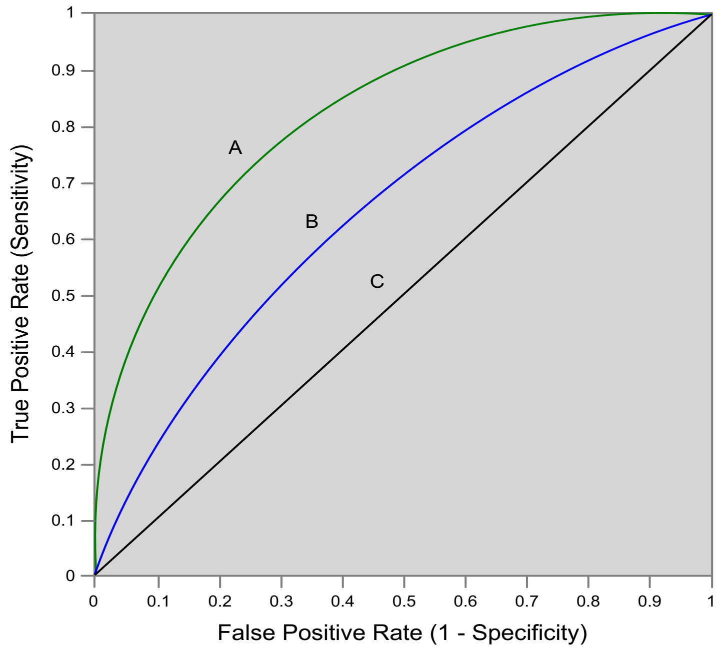
$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$True\ Positive\ Rate\ (\text{Sensitivity}) = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate\ (\text{Specificity}) = \frac{TN}{TN + FP}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

$$False\ Positive\ Rate\ (1\text{-Specificity}) = \frac{FP}{FP + TN}$$



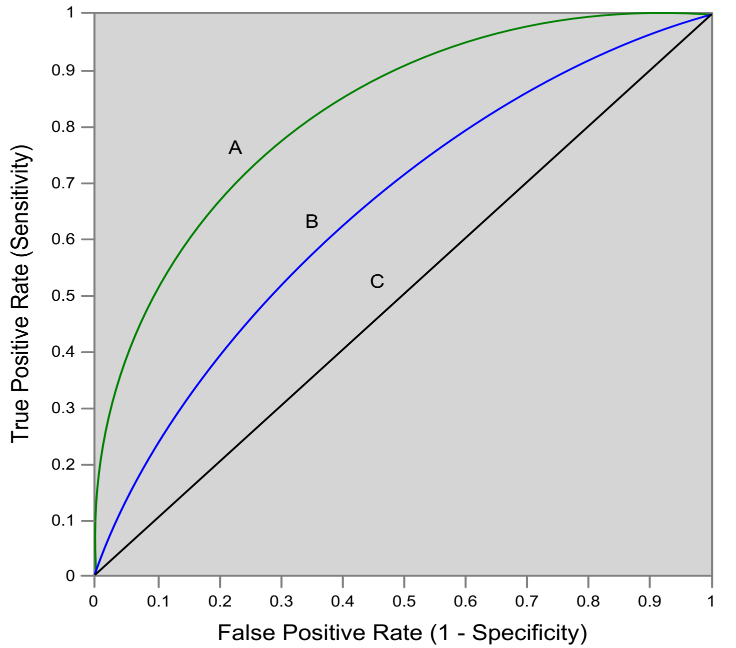
Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

		True Class (actual value)		
		Positive	Negative	**total**
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$True\ Positive\ Rate\ (\text{Sensitivity}) = \frac{TP}{TP + FN}$$



Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

Sensitivity
= True Positive Rate
= Recall
= Hit rate
= TP / (TP + FN)

		True Class (actual value)		
		Positive	Negative	total
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

Specificity

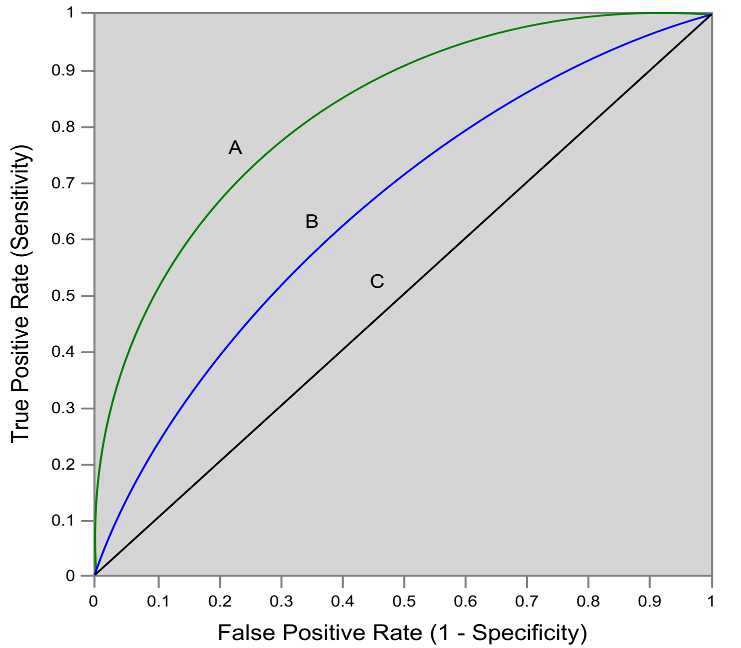
= True Negative Rate

= TN / N

= TN / (TN+ FP)

$$True\ Negative\ Rate\ (\text{Specificity}) = \frac{TN}{TN + FP}$$

$$False\ Positive\ Rate\ (1\text{-}\text{Specificity}) = \frac{FP}{FP + TN}$$



Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

		True Class (actual value)		
		Positive	Negative	total
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P’
	Negative	False Negative (FN)	True Negative (TN)	N’
total		P	N	

Precision

= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

Recall

= True Positive Rate (TPR)

= Sensitivity

= Hit Rate

$$Recall = \frac{TP}{TP + FN}$$

F1 score (F-score)(F-measure)

is the harmonic mean of precision and recall

= 2TP / (P + P’)

= 2TP / (2TP + FP + FN)

$$F = 2 * \frac{precision * recall}{precision + recall}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

A

63 (TP)	28 (FP)	91
37 (FN)	72 (TN)	109
100	100	200

Recall
= True Positive Rate (TPR)
= Sensitivity
= Hit Rate
= TP / (TP + FN)

Specificity
= True Negative Rate
= TN / N
= TN / (TN + FP)

TPR = 0.63

$$Recall = \frac{TP}{TP+FN}$$

$$True\ Negative\ Rate\ (\text{Specificity}) = \frac{TN}{TN+FP}$$

FPR = 0.28

$$False\ Positive\ Rate\ (1\text{-Specificity}) = \frac{FP}{FP+TN}$$

PPV = 0.69
=63/(63+28)
=63/91

$$Precision = \frac{TP}{TP+FP}$$

Precision
= Positive Predictive Value (PPV)

F1 = 0.66
= 2*(0.63*0.69)/(0.63+0.69)
= (2 * 63) /(100 + 91)
= (0.63 + 0.69) / 2 =1.32 / 2 =0.66

$$F = 2 * \frac{precision * recall}{precision + recall}$$

F1 score (F-score)
(F-measure)
is the harmonic mean of precision and recall
= 2TP / (P + P')
= 2TP / (2TP + FP + FN)

ACC = 0.68
= (63 + 72) / 200
= 135/200 = 67.5

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

A

63 (TP)	28 (FP)	91
37 (FN)	72 (TN)	109
100	100	200

TPR = 0.63

FPR = 0.28

PPV = 0.69
=63/(63+28)
=63/91

F1 = 0.66
= 2*(0.63*0.69)/(0.63+0.69)
= (2 * 63) /(100 + 91)
= (0.63 + 0.69) / 2 =1.32 / 2 =0.66

ACC = 0.68
= (63 + 72) / 200
= 135/200 = 67.5

B

77 (TP)	77 (FP)	154
23 (FN)	23 (TN)	46
100	100	200

TPR = 0.77
FPR = 0.77
PPV = 0.50
F1 = 0.61
ACC = 0.50

Recall
= True Positive Rate (TPR)
= Sensitivity
= Hit Rate

$$Recall = \frac{TP}{TP + FN}$$

Precision
= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

C

24 (TP)	88 (FP)	112
76 (FN)	12 (TN)	88
100	100	200

TPR = 0.24
FPR = 0.88
PPV = 0.21
F1 = 0.22
ACC = 0.18

C'

76 (TP)	12 (FP)	88
24 (FN)	88 (TN)	112
100	100	200

TPR = 0.76
FPR = 0.12
PPV = 0.86
F1 = 0.81
ACC = 0.82

Recall
= True Positive Rate (TPR)
= Sensitivity
= Hit Rate

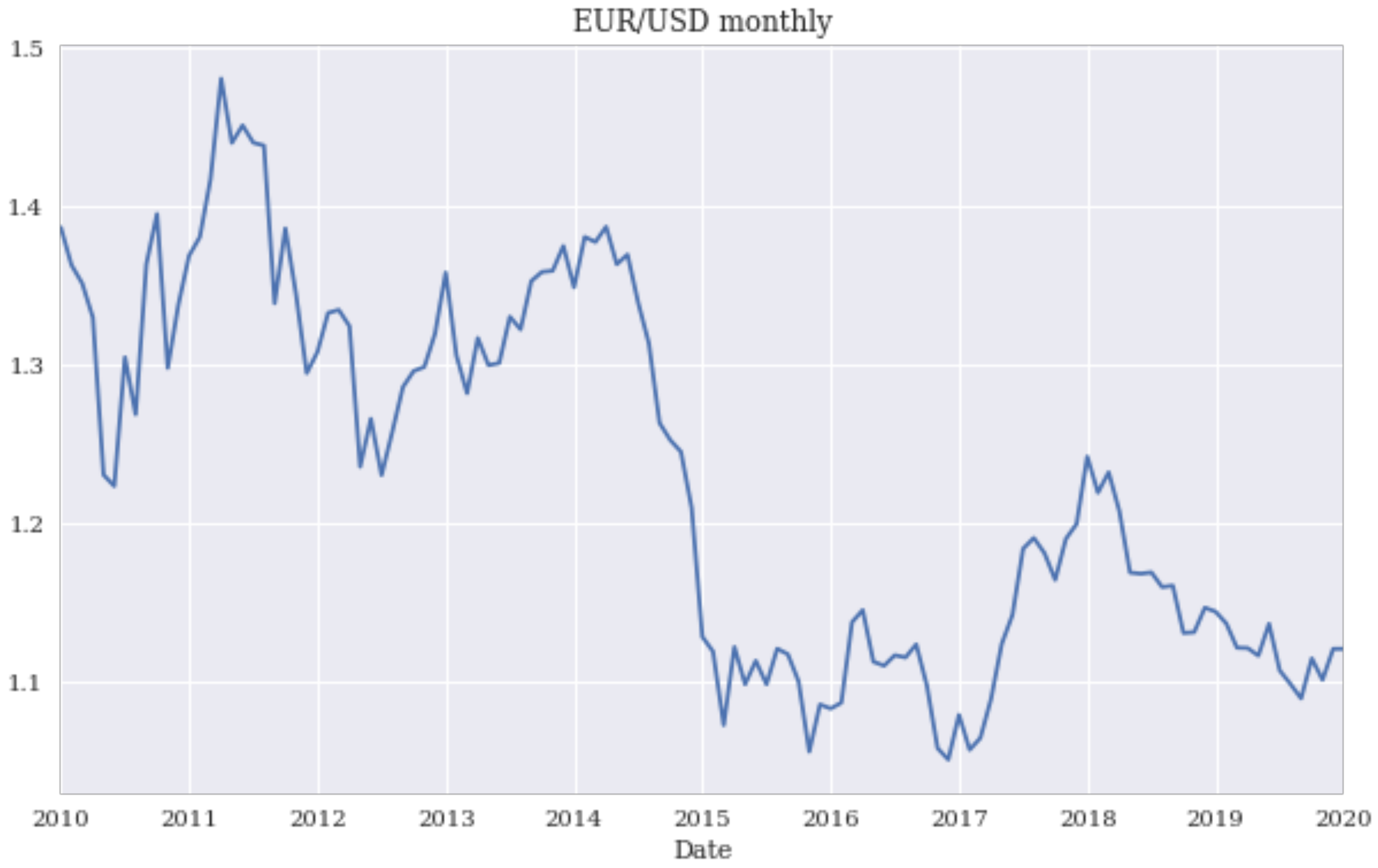
$$Recall = \frac{TP}{TP + FN}$$

Precision
= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

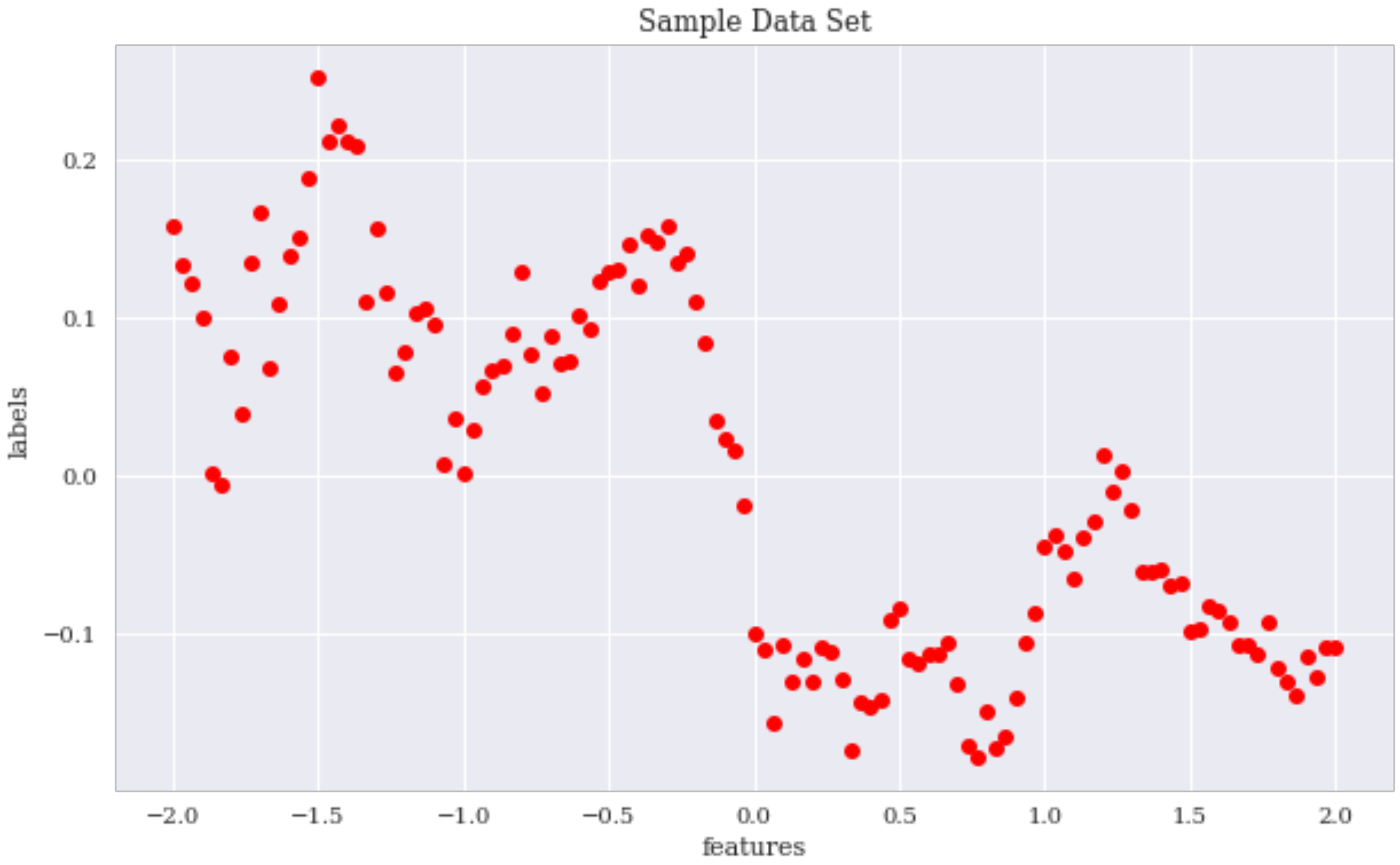
Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

EUR/USD exchange rate as time series (monthly)



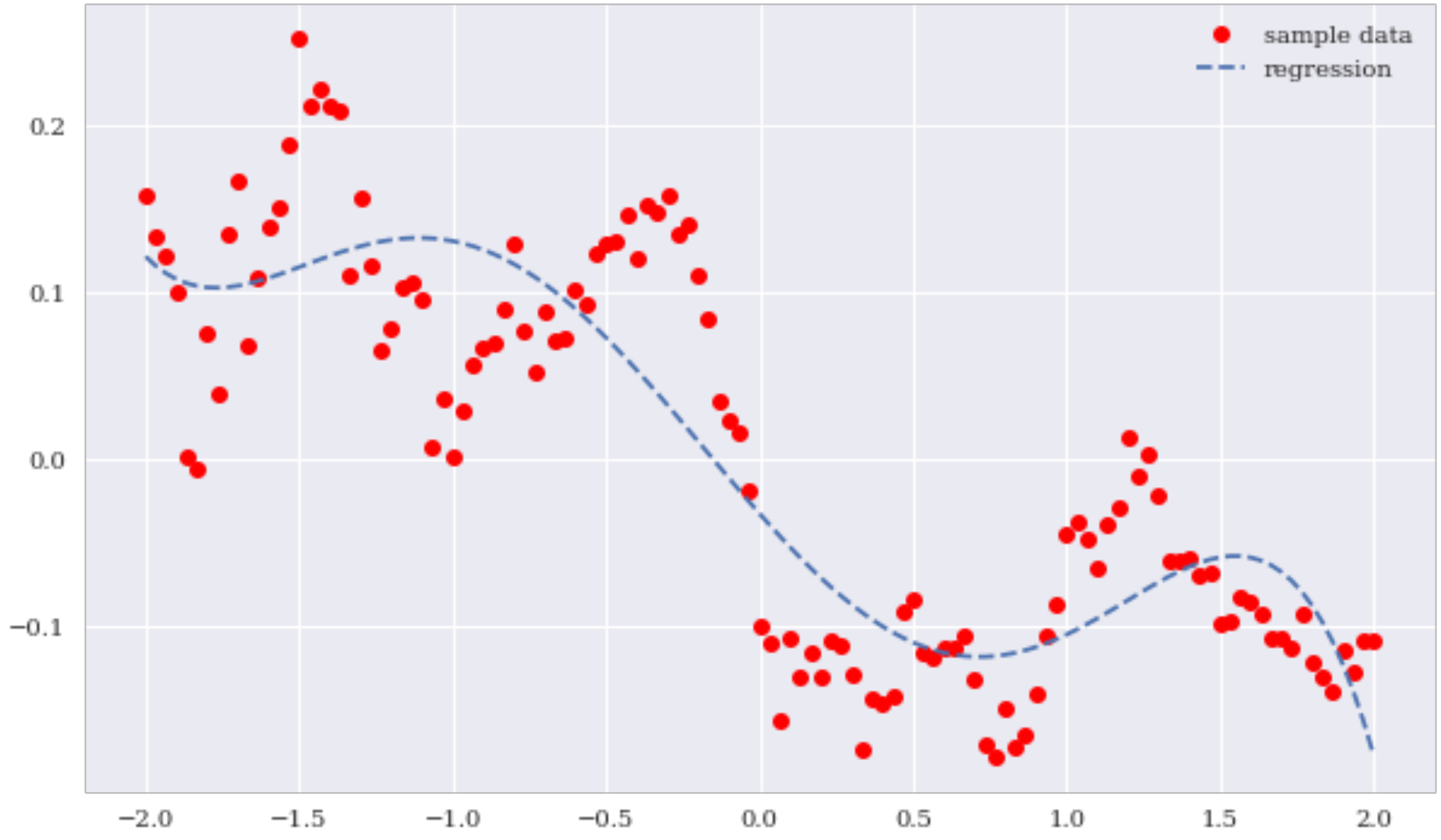
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Sample data set



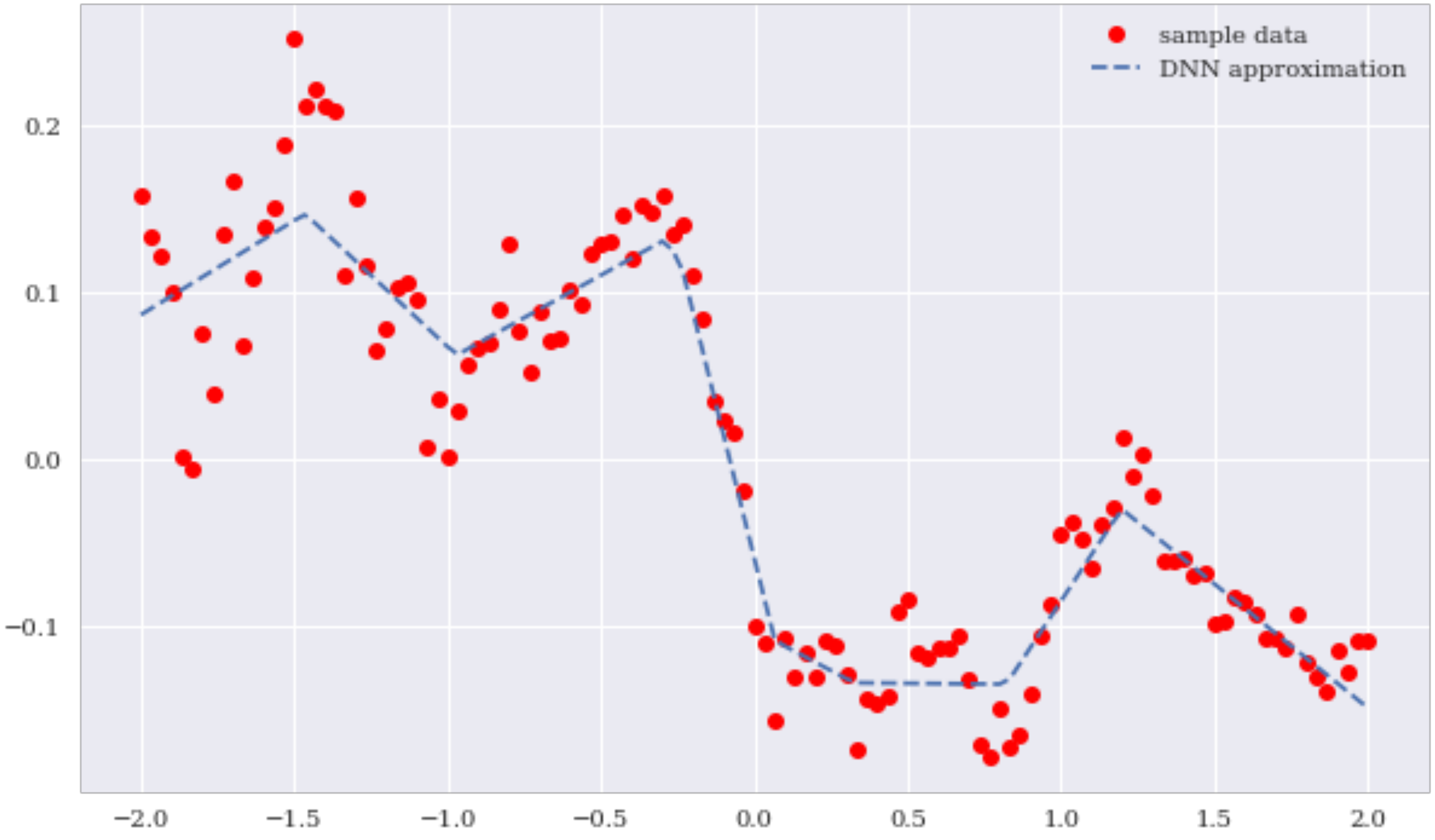
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Sample data and cubic regression line



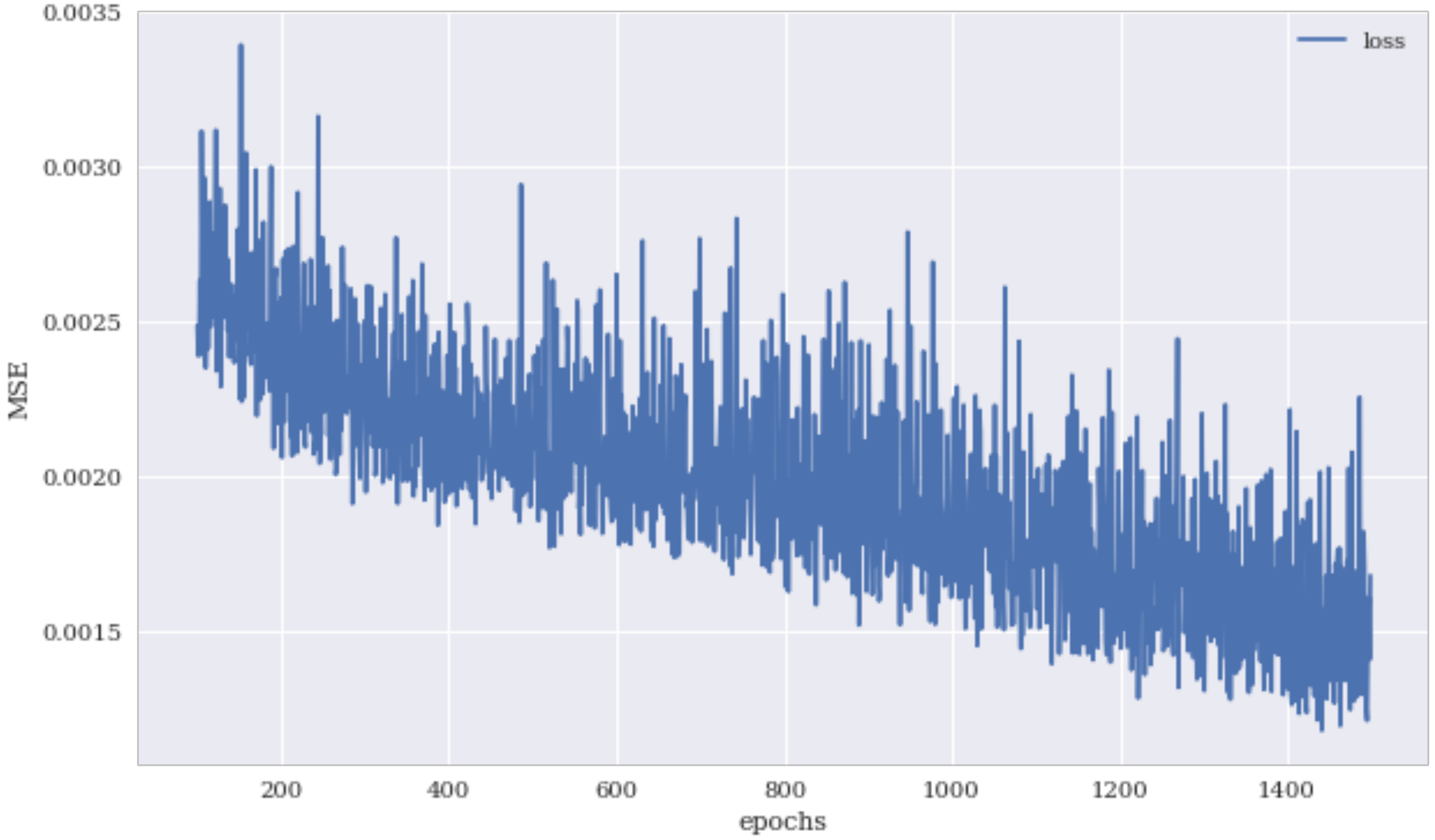
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Sample data and neural network approximation



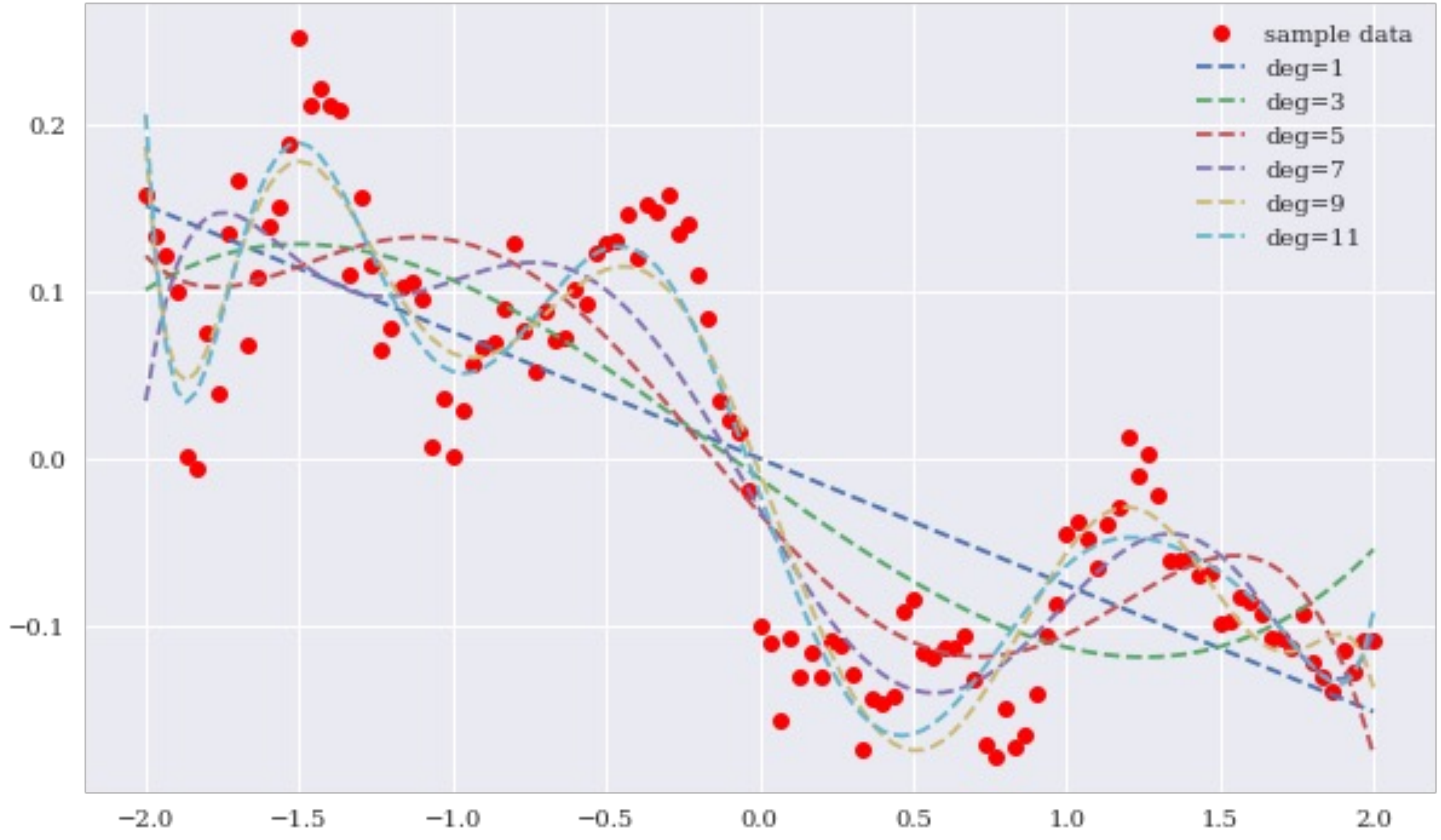
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

MSE values against number of training epochs



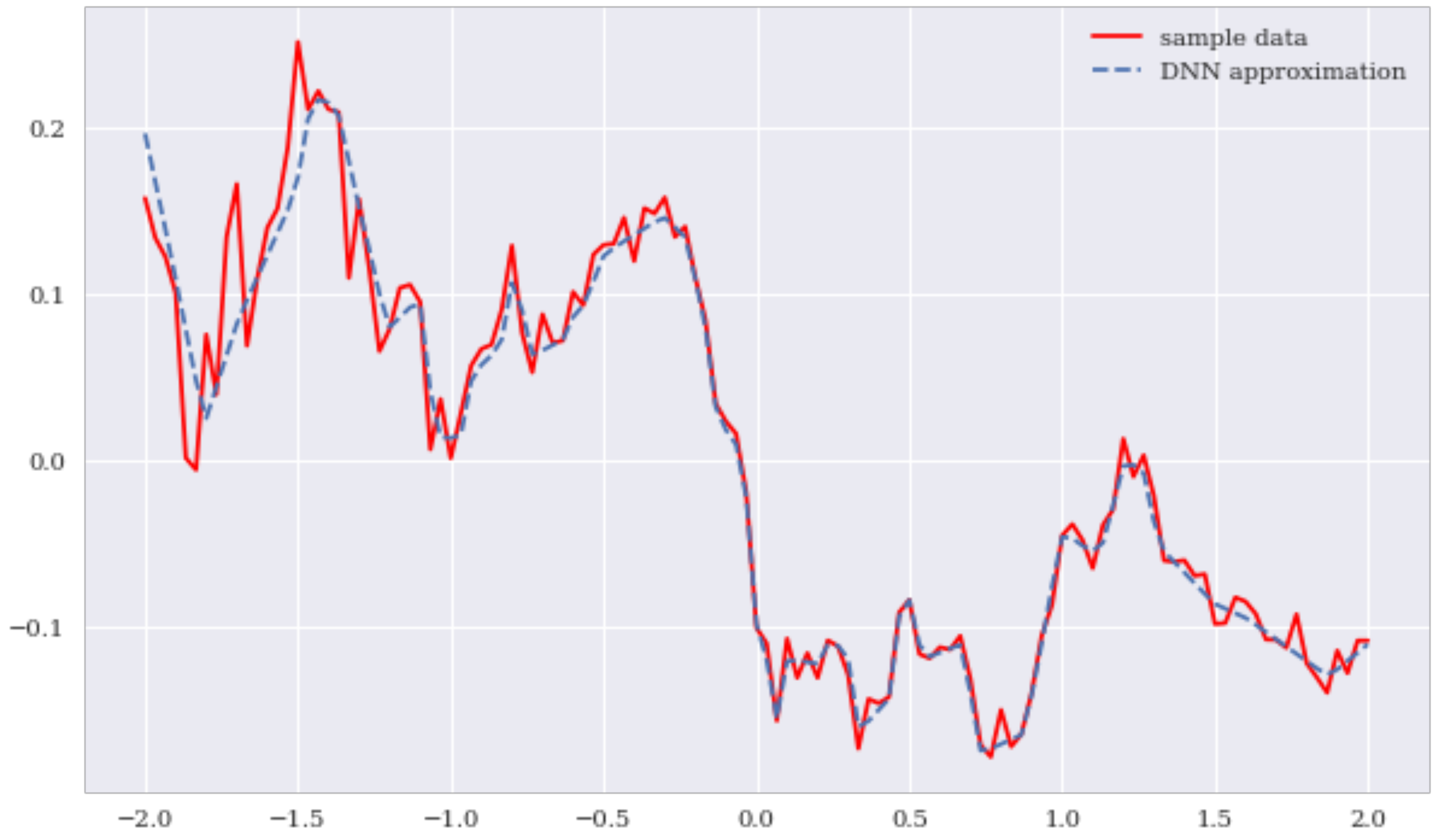
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Regression lines for different highest degrees



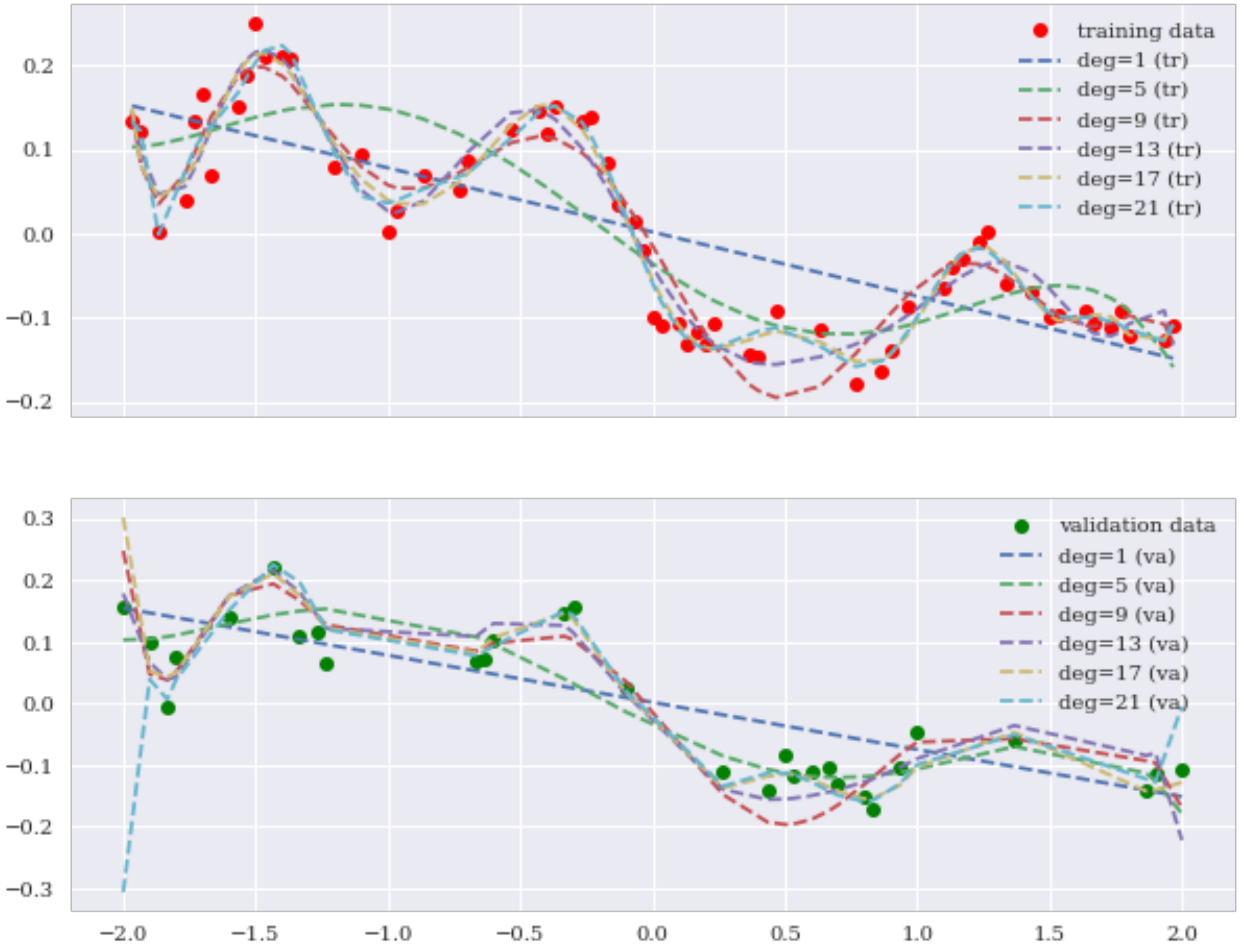
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Sample data and DNN approximation (higher capacity)



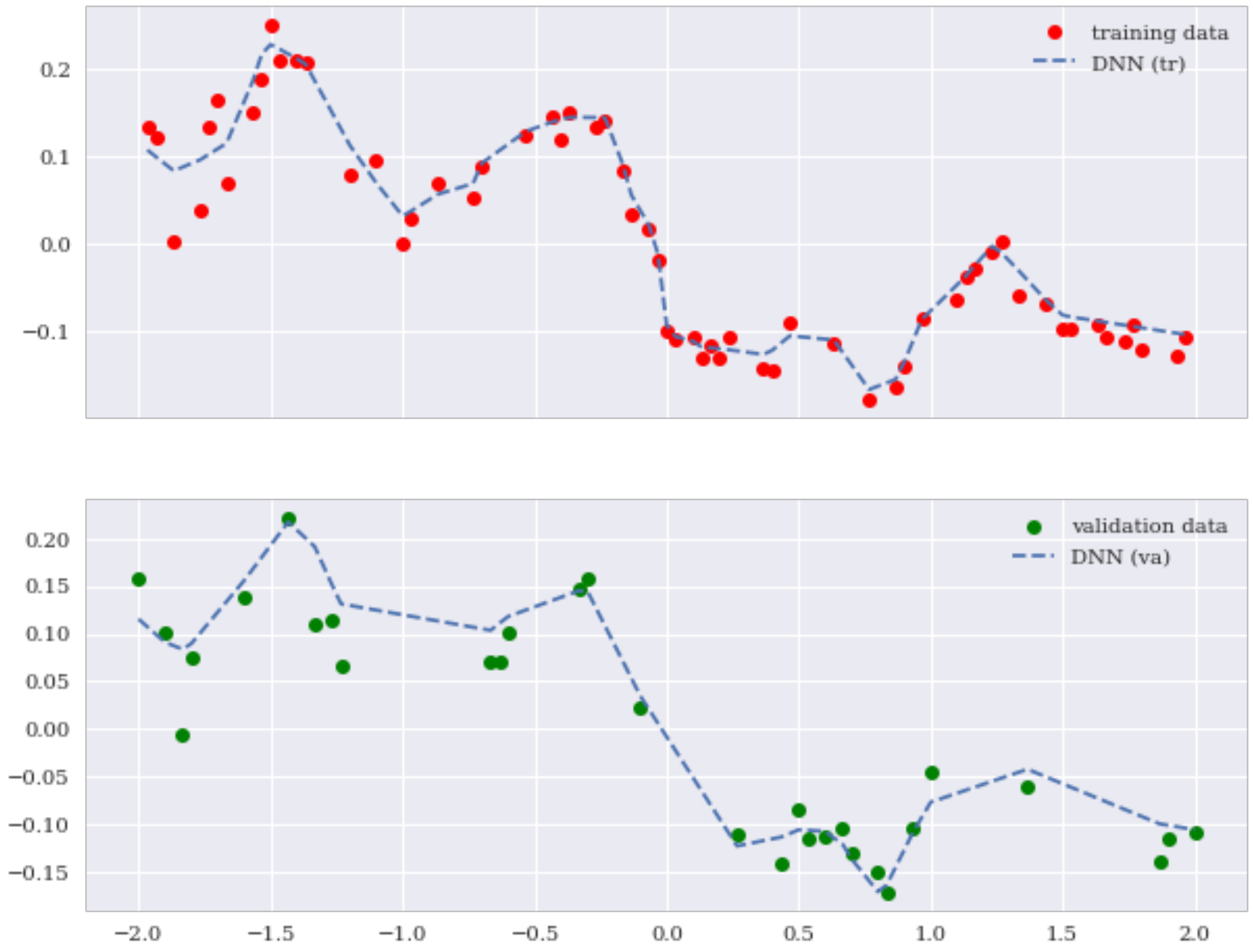
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Training and validation data including regression fits



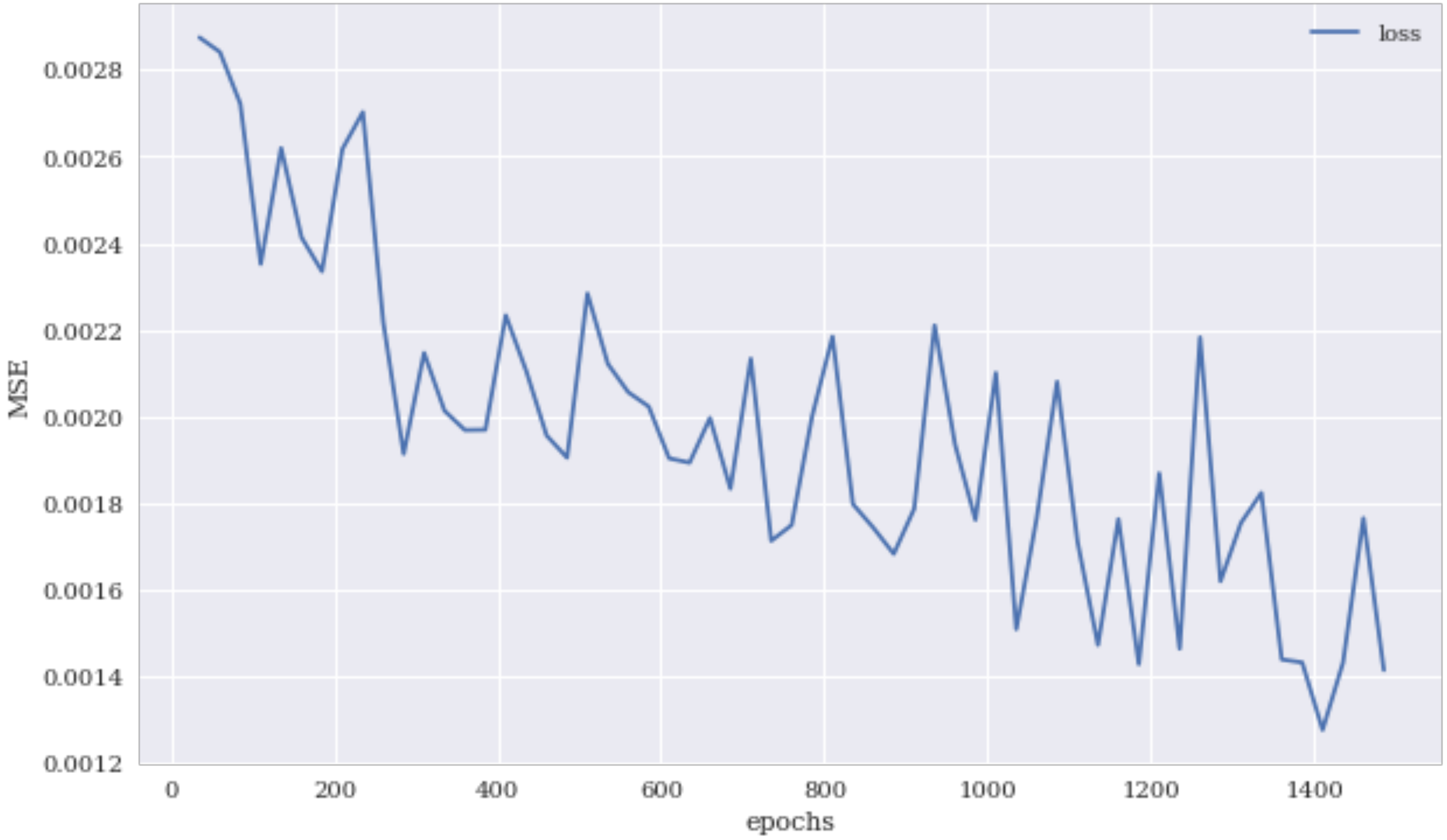
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Training and validation data including DNN predictions



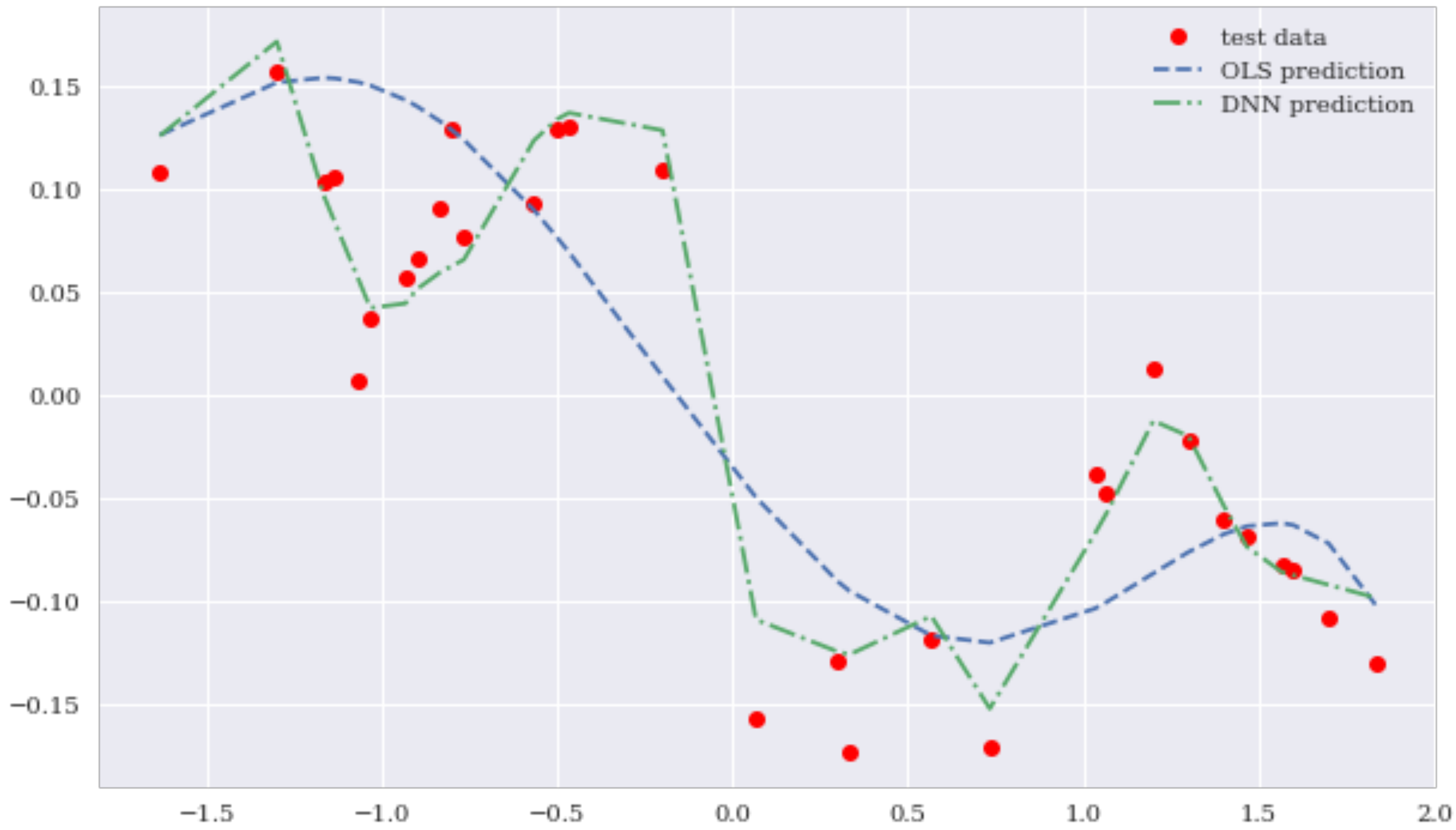
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

MSE values for DNN model on the training and validation data sets



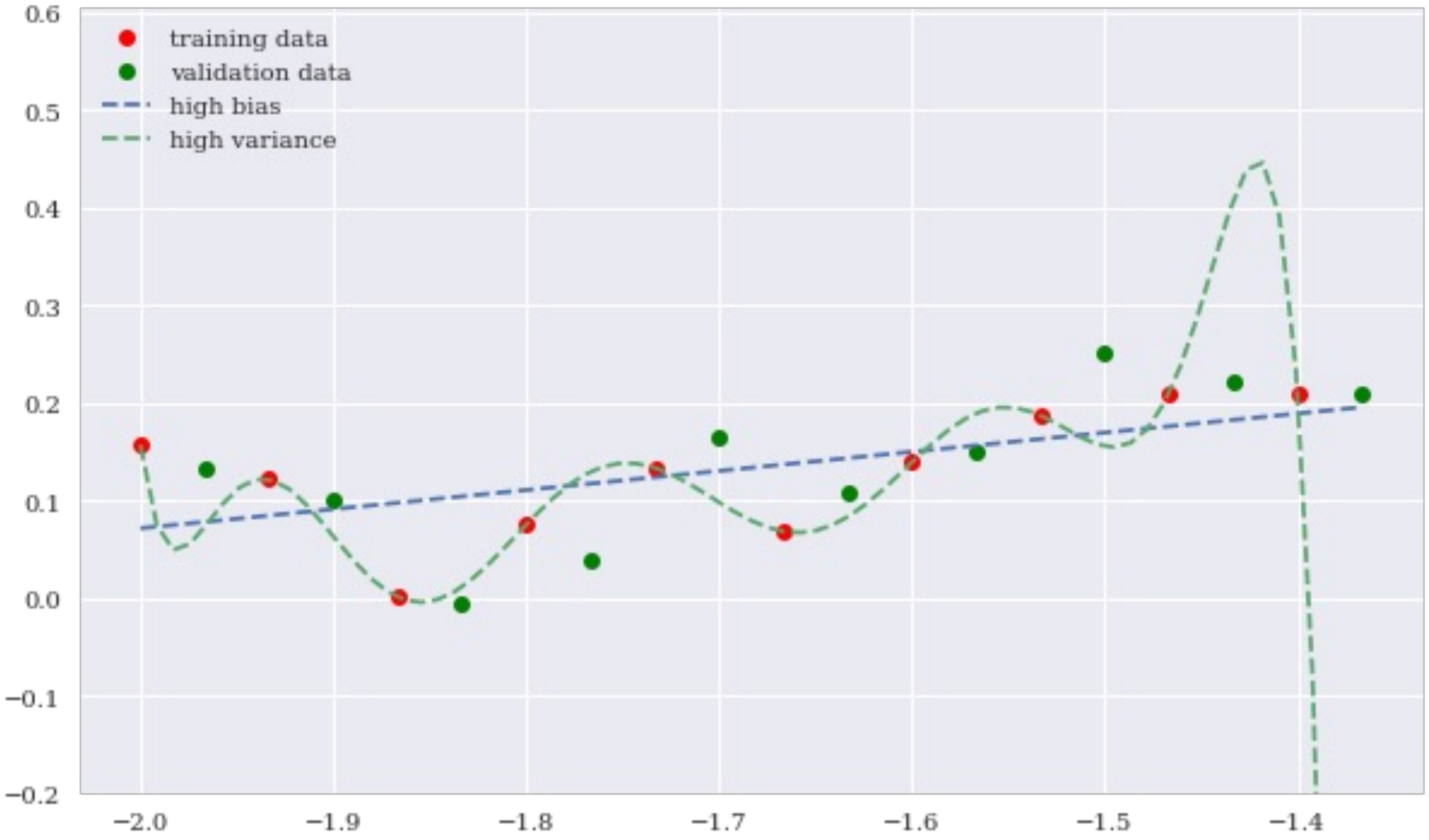
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Test data and predictions from OLS regression and the DNN model



Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

High bias and high variance OLS regression fits



Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

AI-First Finance

AI-First Finance

- **Efficient Markets**
- **Market Prediction Based on Returns Data**
- **Market Prediction with More Features**
- **Market Prediction Intraday**

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Life 3.0:

Being human in the age of artificial intelligence

Max Tegmark (2017)

A computation takes information and transforms it, implementing what mathematicians call a function.... If you're in possession of a function that inputs all the world's financial data and outputs the best stocks to buy, you'll soon be extremely rich.

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Efficient Markets

- Efficient Market Hypothesis (EMH)
  - Random Walk Hypothesis (RWH)
- Weak form of EMH
  - The information set θ_t only encompasses the past price and return history of the market.
- Semi-strong form of EMH
  - The information set θ_t is taken to be all publicly available information, including not only the past price and return history but also financial reports, news articles, weather data, and so on.
- Strong form of EMH
  - The information set θ_t includes all information available to anyone (that is, even private information).

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
import numpy as np
import pandas as pd
from pylab import plt, mpl
plt.style.use('seaborn')
mpl.rcParams['savefig.dpi'] = 300
mpl.rcParams['font.family'] = 'serif'
pd.set_option('precision', 4)
np.set_printoptions(suppress=True, precision=4)

url = 'http://hilpisch.com/aiif_eikon_eod_data.csv'

data = pd.read_csv(url, index_col=0, parse_dates=True).dropna()

(data / data.iloc[0]).plot(figsize=(10, 6), cmap='coolwarm')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Normalized time series data (end-of-day)

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
lags = 7

def add_lags(data, ric, lags):
    cols = []
    df = pd.DataFrame(data[ric])
    for lag in range(1, lags + 1):
        col = 'lag_{}'.format(lag)
        df[col] = df[ric].shift(lag)
        cols.append(col)
    df.dropna(inplace=True)
    return df, cols

dfs = {}
for sym in data.columns:
    df, cols = add_lags(data, sym, lags)
    dfs[sym] = df
dfs[sym].head(7)
```

lagged prices

Date	GLD	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7
2010-01-13	111.54	110.49	112.85	111.37	110.82	111.51	109.70	109.80
2010-01-14	112.03	111.54	110.49	112.85	111.37	110.82	111.51	109.70
2010-01-15	110.86	112.03	111.54	110.49	112.85	111.37	110.82	111.51
2010-01-19	111.52	110.86	112.03	111.54	110.49	112.85	111.37	110.82
2010-01-20	108.94	111.52	110.86	112.03	111.54	110.49	112.85	111.37
2010-01-21	107.37	108.94	111.52	110.86	112.03	111.54	110.49	112.85
2010-01-22	107.17	107.37	108.94	111.52	110.86	112.03	111.54	110.49

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```
regs = {}
for sym in data.columns:
df = dfs[sym]
reg = np.linalg.lstsq(df[cols], df[sym], rcond=-1)[0]
#Return the least-squares solution to a linear matrix equation
regs[sym] = reg

rega = np.stack(tuple(regs.values()))
regd = pd.DataFrame(rega, columns=cols, index=data.columns)
regd
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

regression analysis

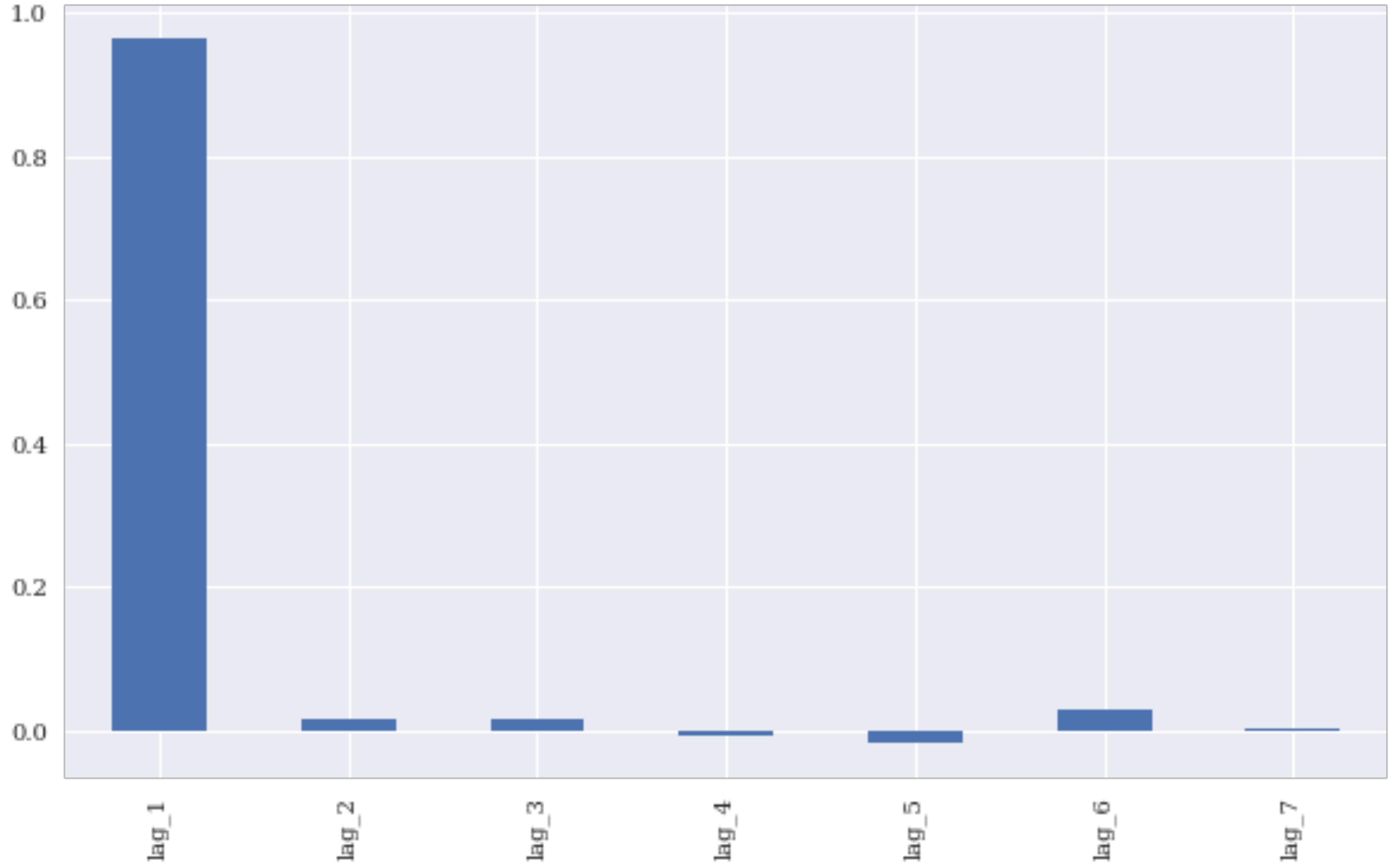
```
reg = np.linalg.lstsq(df[cols], df[sym], rcond=-1)[0]
```

	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7
AAPL.O	1.0106	-0.0592	0.0258	0.0535	-0.0172	0.0060	-0.0184
MSFT.O	0.8928	0.0112	0.1175	-0.0832	-0.0258	0.0567	0.0323
INTC.O	0.9519	0.0579	0.0490	-0.0772	-0.0373	0.0449	0.0112
AMZN.O	0.9799	-0.0134	0.0206	0.0007	0.0525	-0.0452	0.0056
GS.N	0.9806	0.0342	-0.0172	0.0042	-0.0387	0.0585	-0.0215
SPY	0.9692	0.0067	0.0228	-0.0244	-0.0237	0.0379	0.0121
.SPX	0.9672	0.0106	0.0219	-0.0252	-0.0318	0.0515	0.0063
.VIX	0.8823	0.0591	-0.0289	0.0284	-0.0256	0.0511	0.0306
EUR=	0.9859	0.0239	-0.0484	0.0508	-0.0217	0.0149	-0.0055
XAU=	0.9864	0.0069	0.0166	-0.0215	0.0044	0.0198	-0.0125
GDX	0.9765	0.0096	-0.0039	0.0223	-0.0364	0.0379	-0.0065
GLD	0.9766	0.0246	0.0060	-0.0142	-0.0047	0.0223	-0.0106

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Average optimal regression parameters for the lagged prices

```python
regd.mean().plot(kind='bar', figsize=(10, 6))
```



Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Correlations between the lagged time series

```
dfs[sym].corr()
```

	GLD	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7
GLD	1.0000	0.9972	0.9946	0.9920	0.9893	0.9867	0.9841	0.9815
lag_1	0.9972	1.0000	0.9972	0.9946	0.9920	0.9893	0.9867	0.9842
lag_2	0.9946	0.9972	1.0000	0.9972	0.9946	0.9920	0.9893	0.9867
lag_3	0.9920	0.9946	0.9972	1.0000	0.9972	0.9946	0.9920	0.9893
lag_4	0.9893	0.9920	0.9946	0.9972	1.0000	0.9972	0.9946	0.9920
lag_5	0.9867	0.9893	0.9920	0.9946	0.9972	1.0000	0.9972	0.9946
lag_6	0.9841	0.9867	0.9893	0.9920	0.9946	0.9972	1.0000	0.9972
lag_7	0.9815	0.9842	0.9867	0.9893	0.9920	0.9946	0.9972	1.0000

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
from statsmodels.tsa.stattools import adfuller
#Tests for stationarity using the Augmented Dickey-Fuller (ADF) test

adfuller(data[sym].dropna())
```

```
(-1.9488969577009954,
 0.3094193074034718,
 0,
 2515,
 {'1%': -3.4329527780962255,
  '10%': -2.567382133955709,
  '5%': -2.8626898965523724},
 8446.683102944744)
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Market Prediction Based on Returns Data

- **Statistical inefficiencies**
  - **are given when a model is able to predict the direction of the future price movement with a certain edge (say, the prediction is correct in 55% or 60% of the cases)**
- **Economic inefficiencies**
  - **would only be given if the statistical inefficiencies can be exploited profitably through a trading strategy that takes into account, for example, transaction costs.**

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Market Prediction Based on Returns Data

- Create data sets with lagged log returns data
- The normalized lagged log returns data is also tested for stationarity (given)
- The features are tested for correlation (not correlated)
- Time-series-related data
  - weak form market efficiency

```
rets = np.log(data / data.shift(1))
rets.dropna(inplace=True)
rets
```

log returns

Date	AAPL.O	MSFT.O	INTC.O	AMZN.O	GS.N	SPY	.SPX	.VIX	EUR=	XAU=	GDX	GLD
2010-01-05	0.0017	0.0003	-0.0005	0.0059	0.0175	2.6436e-03	3.1108e-03	-0.0350	-2.9883e-03	-0.0012	0.0096	-0.0009
2010-01-06	-0.0160	-0.0062	-0.0034	-0.0183	-0.0107	7.0379e-04	5.4538e-04	-0.0099	3.0577e-03	0.0176	0.0240	0.0164
2010-01-07	-0.0019	-0.0104	-0.0097	-0.0172	0.0194	4.2124e-03	3.9933e-03	-0.0052	-6.5437e-03	-0.0058	-0.0049	-0.0062
2010-01-08	0.0066	0.0068	0.0111	0.0267	-0.0191	3.3223e-03	2.8775e-03	-0.0500	6.5437e-03	0.0037	0.0150	0.0050
2010-01-11	-0.0089	-0.0128	0.0057	-0.0244	-0.0159	1.3956e-03	1.7452e-03	-0.0325	6.9836e-03	0.0144	0.0066	0.0132
...	...	...	...	...	...	...	...	...	...	...	...	...
2019-12-24	0.0010	-0.0002	0.0030	-0.0021	0.0036	3.1131e-05	-1.9543e-04	0.0047	9.0200e-05	0.0091	0.0315	0.0094
2019-12-26	0.0196	0.0082	0.0069	0.0435	0.0056	5.3092e-03	5.1151e-03	-0.0016	8.1143e-04	0.0083	0.0145	0.0078
2019-12-27	-0.0004	0.0018	0.0043	0.0006	-0.0024	-2.4775e-04	3.3951e-05	0.0598	7.0945e-03	-0.0006	-0.0072	-0.0004
2019-12-30	0.0059	-0.0087	-0.0077	-0.0123	-0.0037	-5.5285e-03	-5.7976e-03	0.0985	1.9667e-03	0.0031	0.0212	0.0021
2019-12-31	0.0073	0.0007	0.0039	0.0005	0.0006	2.4264e-03	2.9417e-03	-0.0728	1.1604e-03	0.0012	-0.0071	0.0019

2515 rows × 12 columns

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
dfs = {}
for sym in data:
    df, cols = add_lags(rets, sym, lags)
    mu, std = df[cols].mean(), df[cols].std()
    df[cols] = (df[cols] - mu) / std
    dfs[sym] = df
dfs[sym].head()
```

Date	GLD	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7
2010-01-14	0.0044	0.9570	-2.1692	1.3386	0.4959	-0.6434	1.6613	-0.1028
2010-01-15	-0.0105	0.4379	0.9571	-2.1689	1.3388	0.4966	-0.6436	1.6614
2010-01-19	0.0059	-1.0842	0.4385	0.9562	-2.1690	1.3395	0.4958	-0.6435
2010-01-20	-0.0234	0.5967	-1.0823	0.4378	0.9564	-2.1686	1.3383	0.4958
2010-01-21	-0.0145	-2.4045	0.5971	-1.0825	0.4379	0.9571	-2.1680	1.3384

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Augmented Dickey-Fuller (ADF)

Tests for stationarity of the time series data

```
adfuller(dfs[sym]['lag_1'])
```

```
(-51.568251505825536,
 0.0,
 0,
 2507,
 {'1%': -3.4329610922579095,
  '10%': -2.567384088736619,
  '5%': -2.8626935681060375},
 7017.165474260225)
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

Shows the correlation data for the features

```
dfs[sym].corr()
```

	GLD	lag_1	lag_2	lag_3	lag_4	lag_5	lag_6	lag_7
GLD	1.0000	-0.0297	0.0003	1.2635e-02	-0.0026	-5.9392e-03	0.0099	-0.0013
lag_1	-0.0297	1.0000	-0.0305	8.1418e-04	0.0128	-2.8765e-03	-0.0053	0.0098
lag_2	0.0003	-0.0305	1.0000	-3.1617e-02	0.0003	1.3234e-02	-0.0043	-0.0052
lag_3	0.0126	0.0008	-0.0316	1.0000e+00	-0.0313	-6.8542e-06	0.0141	-0.0044
lag_4	-0.0026	0.0128	0.0003	-3.1329e-02	1.0000	-3.1761e-02	0.0002	0.0141
lag_5	-0.0059	-0.0029	0.0132	-6.8542e-06	-0.0318	1.0000e+00	-0.0323	0.0002
lag_6	0.0099	-0.0053	-0.0043	1.4115e-02	0.0002	-3.2289e-02	1.0000	-0.0324
lag_7	-0.0013	0.0098	-0.0052	-4.3869e-03	0.0141	2.1707e-04	-0.0324	1.0000

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

OLS Regression

```
from sklearn.metrics import accuracy_score
```

```
%%time
for sym in data:
    df = dfs[sym]
    reg = np.linalg.lstsq(df[cols], df[sym], rcond=-1)[0]
    pred = np.dot(df[cols], reg)
    acc = accuracy_score(np.sign(df[sym]), np.sign(pred))
    print(f'OLS | {sym:10s} | acc={acc:.4f}')
```

OLS Regression Accuracy

```
OLS | AAPL.O     | acc=0.5056
OLS | MSFT.O     | acc=0.5088
OLS | INTC.O     | acc=0.5040
OLS | AMZN.O     | acc=0.5048
OLS | GS.N       | acc=0.5080
OLS | SPY        | acc=0.5080
OLS | .SPX       | acc=0.5167
OLS | .VIX       | acc=0.5291
OLS | EUR=       | acc=0.4984
OLS | XAU=       | acc=0.5207
OLS | GDX        | acc=0.5307
OLS | GLD        | acc=0.5072
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

```python
from sklearn.neural_network import MLPRegressor
```

```python
%%time
for sym in data.columns:
   df = dfs[sym]
   model = MLPRegressor(hidden_layer_sizes=[512],
                        random_state=100,
                        max_iter=1000,
                        early_stopping=True,
                        validation_fraction=0.15,
                        shuffle=False)
   model.fit(df[cols].values, df[sym].values)
   pred = model.predict(df[cols].values)
   acc = accuracy_score(np.sign(df[sym].values),
   np.sign(pred))
   print(f'MLP | {sym:10s} | acc={acc:.4f}')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Scikit-learn MLPRegressor Accuracy

```
MLP | AAPL.O       | acc=0.6005
MLP | MSFT.O       | acc=0.5853
MLP | INTC.O       | acc=0.5766
MLP | AMZN.O       | acc=0.5510
MLP | GS.N         | acc=0.6527
MLP | SPY          | acc=0.5419
MLP | .SPX         | acc=0.5399
MLP | .VIX         | acc=0.6579
MLP | EUR=         | acc=0.5642
MLP | XAU=         | acc=0.5522
MLP | GDX          | acc=0.6029
MLP | GLD          | acc=0.5259
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
import tensorflow as tf
from keras.layers import Dense
from keras.models import Sequential

np.random.seed(100)
tf.random.set_seed(100)

def create_model(problem='regression'):
    model = Sequential()
    model.add(Dense(512, input_dim=len(cols), activation='relu'))
    if problem == 'regression’:
        model.add(Dense(1, activation='linear’))
        model.compile(loss='mse', optimizer='adam')
    else:
        model.add(Dense(1, activation='sigmoid’))
        model.compile(loss='binary_crossentropy', optimizer='adam')
    return model
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

```python
%%time
for sym in data.columns[:]:
    df = dfs[sym]
    model = create_model()
    model.fit(df[cols], df[sym], epochs=25, verbose=False)
    pred = model.predict(df[cols])
    acc = accuracy_score(np.sign(df[sym]), np.sign(pred))
    print(f'DNN | {sym:10s} | acc={acc:.4f}')
```

TF Keras DNN Accuracy

```
DNN | AAPL.O       | acc=0.6069
DNN | MSFT.O       | acc=0.6260
DNN | INTC.O       | acc=0.6344
DNN | AMZN.O       | acc=0.6316
DNN | GS.N         | acc=0.6045
DNN | SPY          | acc=0.5610
DNN | .SPX         | acc=0.5435
DNN | .VIX         | acc=0.6096
DNN | EUR=         | acc=0.5817
DNN | XAU=         | acc=0.6017
DNN | GDX          | acc=0.6164
DNN | GLD          | acc=0.5973
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Train Data (0.8): In-Sample
Test Data (0.2): Out-of-Sample

```
split = int(len(dfs[sym]) * 0.8)
```

```
%%time
for sym in data.columns:
   df = dfs[sym]
   train = df.iloc[:split]
   reg = np.linalg.lstsq(train[cols], train[sym], rcond=-1)[0]
   test = df.iloc[split:]
   pred = np.dot(test[cols], reg)
   acc = accuracy_score(np.sign(test[sym]), np.sign(pred))
   print(f'OLS | {sym:10s} | acc={acc:.4f}')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

OLS Out-of-Sample Accuracy

```
OLS | AAPL.O        | acc=0.5219
OLS | MSFT.O        | acc=0.4960
OLS | INTC.O        | acc=0.5418
OLS | AMZN.O        | acc=0.4841
OLS | GS.N          | acc=0.4980
OLS | SPY           | acc=0.5020
OLS | .SPX          | acc=0.5120
OLS | .VIX          | acc=0.5458
OLS | EUR=          | acc=0.4482
OLS | XAU=          | acc=0.5299
OLS | GDX           | acc=0.5159
OLS | GLD           | acc=0.5100
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

```python
%%time
for sym in data.columns:
    df = dfs[sym]
    train = df.iloc[:split]
    model = MLPRegressor(hidden_layer_sizes=[512],
                         random_state=100,
                         max_iter=1000,
                         early_stopping=True,
                         validation_fraction=0.15,
                         shuffle=False)
    model.fit(train[cols].values, train[sym].values)
    test = df.iloc[split:]
    pred = model.predict(test[cols].values)
    acc = accuracy_score(np.sign(test[sym].values), np.sign(pred))
    print(f'MLP | {sym:10s} | acc={acc:.4f}')
```

MLP Out-of-Sample Accuracy

```
MLP | AAPL.O      | acc=0.4920
MLP | MSFT.O      | acc=0.5279
MLP | INTC.O      | acc=0.5279
MLP | AMZN.O      | acc=0.4641
MLP | GS.N        | acc=0.5040
MLP | SPY         | acc=0.5259
MLP | .SPX        | acc=0.5478
MLP | .VIX        | acc=0.5279
MLP | EUR=        | acc=0.4980
MLP | XAU=        | acc=0.5239
MLP | GDX         | acc=0.4880
MLP | GLD         | acc=0.5000
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

```python
%%time
for sym in data.columns:
   df = dfs[sym]
   train = df.iloc[:split]
   model = create_model()
   model.fit(train[cols], train[sym], epochs=50,
   verbose=False)
   test = df.iloc[split:]
   pred = model.predict(test[cols])
   acc = accuracy_score(np.sign(test[sym]), np.sign(pred))
   print(f'DNN | {sym:10s} | acc={acc:.4f}')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

DNN Out-of-Sample Accuracy

```
DNN | AAPL.O      | acc=0.4701
DNN | MSFT.O      | acc=0.4960
DNN | INTC.O      | acc=0.5040
DNN | AMZN.O      | acc=0.4920
DNN | GS.N        | acc=0.5538
DNN | SPY         | acc=0.5299
DNN | .SPX        | acc=0.5458
DNN | .VIX        | acc=0.5020
DNN | EUR=        | acc=0.5100
DNN | XAU=        | acc=0.4940
DNN | GDX         | acc=0.4661
DNN | GLD         | acc=0.4880
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

Market Prediction with More Features

- In trading, there is a long tradition of using technical indicators to generate, based on observed patterns, buy or sell signals.
- Such technical indicators, basically of any kind, can also be used as features for the training of neural networks.
- SMA, rolling minimum and maximum values, momentum, and rolling volatility as features

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
url = 'http://hilpisch.com/aiif_eikon_eod_data.csv'

data = pd.read_csv(url, index_col=0, parse_dates=True).dropna()
data
```

	AAPL.O	MSFT.O	INTC.O	AMZN.O	GS.N	SPY	.SPX	.VIX	EUR=	XAU=	GDX	GLD
Date												
2010-01-04	30.5728	30.950	20.88	133.90	173.08	113.33	1132.99	20.04	1.4411	1120.0000	47.71	109.80
2010-01-05	30.6257	30.960	20.87	134.69	176.14	113.63	1136.52	19.35	1.4368	1118.6500	48.17	109.70
2010-01-06	30.1385	30.770	20.80	132.25	174.26	113.71	1137.14	19.16	1.4412	1138.5000	49.34	111.51
2010-01-07	30.0828	30.452	20.60	130.00	177.67	114.19	1141.69	19.06	1.4318	1131.9000	49.10	110.82
2010-01-08	30.2828	30.660	20.83	133.52	174.31	114.57	1144.98	18.13	1.4412	1136.1000	49.84	111.37
...	...	...	...	...	...	...	...	...	...	...	...	...
2019-12-24	284.2700	157.380	59.41	1789.21	229.91	321.23	3223.38	12.67	1.1087	1498.8100	28.66	141.27
2019-12-26	289.9100	158.670	59.82	1868.77	231.21	322.94	3239.91	12.65	1.1096	1511.2979	29.08	142.38
2019-12-27	289.8000	158.960	60.08	1869.80	230.66	322.86	3240.02	13.43	1.1175	1510.4167	28.87	142.33
2019-12-30	291.5200	157.590	59.62	1846.89	229.80	321.08	3221.29	14.82	1.1197	1515.1230	29.49	142.63
2019-12-31	293.6500	157.700	59.85	1847.84	229.93	321.86	3230.78	13.78	1.1210	1517.0100	29.28	142.90

2516 rows × 12 columns

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
def add_lags(data, ric, lags, window=50):
    cols = []
    df = pd.DataFrame(data[ric])
    df.dropna(inplace=True)
    df['r'] = np.log(df / df.shift())
    df['sma'] = df[ric].rolling(window).mean()
    df['min'] = df[ric].rolling(window).min()
    df['max'] = df[ric].rolling(window).max()
    df['mom'] = df['r'].rolling(window).mean()
    df['vol'] = df['r'].rolling(window).std()
    df.dropna(inplace=True)
    df['d'] = np.where(df['r'] > 0, 1, 0)
    features = [ric, 'r', 'd', 'sma', 'min', 'max', 'mom', 'vol']
    for f in features:
        for lag in range(1, lags + 1):
            col = f'{f}_lag_{lag}'
            df[col] = df[f].shift(lag)
            cols.append(col)
    df.dropna(inplace=True)
    return df, cols
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
lags = 5

dfs = {}
for ric in data:
    df, cols = add_lags(data, ric, lags)
    dfs[ric] = df.dropna(), cols

len(cols)
```

```
40
```

```python
from sklearn.neural_network import MLPClassifier
```

```python
%%time
for ric in data:
    model = MLPClassifier(hidden_layer_sizes=[512],
                          random_state=100,
                          max_iter=1000,
                          early_stopping=True,
                          validation_fraction=0.15,
                          shuffle=False)
    df, cols = dfs[ric]
    df[cols] = (df[cols] - df[cols].mean()) / df[cols].std()
    model.fit(df[cols].values, df['d'].values)
    pred = model.predict(df[cols].values)
    acc = accuracy_score(df['d'].values, pred)
    print(f'IN-SAMPLE | {ric:7s} | acc={acc:.4f}')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

MLP In-Sample Accuracy

```
IN-SAMPLE | AAPL.O   | acc=0.5510
IN-SAMPLE | MSFT.O   | acc=0.5376
IN-SAMPLE | INTC.O   | acc=0.5607
IN-SAMPLE | AMZN.O   | acc=0.5559
IN-SAMPLE | GS.N     | acc=0.5794
IN-SAMPLE | SPY      | acc=0.5729
IN-SAMPLE | .SPX     | acc=0.5941
IN-SAMPLE | .VIX     | acc=0.6940
IN-SAMPLE | EUR=     | acc=0.5766
IN-SAMPLE | XAU=     | acc=0.5672
IN-SAMPLE | GDX      | acc=0.5847
IN-SAMPLE | GLD      | acc=0.5567
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
%%time
for ric in data:
    model = create_model('classification')
    df, cols = dfs[ric]
    df[cols] = (df[cols] - df[cols].mean()) / df[cols].std()
    model.fit(df[cols], df['d'], epochs=50, verbose=False)
    pred = np.where(model.predict(df[cols]) > 0.5, 1, 0)
    acc = accuracy_score(df['d'], pred)
    print(f'IN-SAMPLE | {ric:7s} | acc={acc:.4f}')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media.

TF Keras DNN In-Sample Accuracy

```
IN-SAMPLE | AAPL.O   | acc=0.7042
IN-SAMPLE | MSFT.O   | acc=0.6928
IN-SAMPLE | INTC.O   | acc=0.6969
IN-SAMPLE | AMZN.O   | acc=0.6713
IN-SAMPLE | GS.N     | acc=0.6924
IN-SAMPLE | SPY      | acc=0.6806
IN-SAMPLE | .SPX     | acc=0.6920
IN-SAMPLE | .VIX     | acc=0.7347
IN-SAMPLE | EUR=     | acc=0.6766
IN-SAMPLE | XAU=     | acc=0.7038
IN-SAMPLE | GDX      | acc=0.6806
IN-SAMPLE | GLD      | acc=0.6936
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
def train_test_model(model):
    for ric in data:
        df, cols = dfs[ric]
        split = int(len(df) * 0.85)
        train = df.iloc[:split].copy()
        mu, std = train[cols].mean(), train[cols].std()
        train[cols] = (train[cols] - mu) / std
        model.fit(train[cols].values, train['d'].values)
        test = df.iloc[split:].copy()
        test[cols] = (test[cols] - mu) / std
        pred = model.predict(test[cols].values)
        acc = accuracy_score(test['d'].values, pred)
        print(f'OUT-OF-SAMPLE | {ric:7s} | acc={acc:.4f}')
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
model_mlp = MLPClassifier(hidden_layer_sizes=[512],
                          random_state=100,
                          max_iter=1000,
                          early_stopping=True,
                          validation_fraction=0.15,
                          shuffle=False)
```

```python
train_test_model(model_mlp)
```

```
train_test_model(model_mlp)
```

```
OUT-OF-SAMPLE | AAPL.O   | acc=0.4432
OUT-OF-SAMPLE | MSFT.O   | acc=0.4595
OUT-OF-SAMPLE | INTC.O   | acc=0.5000
OUT-OF-SAMPLE | AMZN.O   | acc=0.5270
OUT-OF-SAMPLE | GS.N     | acc=0.4838
OUT-OF-SAMPLE | SPY      | acc=0.4811
OUT-OF-SAMPLE | .SPX     | acc=0.5027
OUT-OF-SAMPLE | .VIX     | acc=0.5676
OUT-OF-SAMPLE | EUR=     | acc=0.4649
OUT-OF-SAMPLE | XAU=     | acc=0.5514
OUT-OF-SAMPLE | GDX      | acc=0.5162
OUT-OF-SAMPLE | GLD      | acc=0.4946
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```
from sklearn.ensemble import BaggingClassifier

base_estimator = MLPClassifier(hidden_layer_sizes=[256],
                      random_state=100,
                      max_iter=1000,
                      early_stopping=True,
                      validation_fraction=0.15,
                      shuffle=False)

model_bag = BaggingClassifier(base_estimator=base_estimator,
                      n_estimators=35,
                      max_samples=0.25,
                      max_features=0.5,
                      bootstrap=False,
                      bootstrap_features=True,
                      n_jobs=8,
                      random_state=100
                      )
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```
train_test_model(model_bag)
```

```
OUT-OF-SAMPLE | AAPL.O   | acc=0.5000
OUT-OF-SAMPLE | MSFT.O   | acc=0.5703
OUT-OF-SAMPLE | INTC.O   | acc=0.5054
OUT-OF-SAMPLE | AMZN.O   | acc=0.5270
OUT-OF-SAMPLE | GS.N     | acc=0.5135
OUT-OF-SAMPLE | SPY      | acc=0.5568
OUT-OF-SAMPLE | .SPX     | acc=0.5514
OUT-OF-SAMPLE | .VIX     | acc=0.5432
OUT-OF-SAMPLE | EUR=     | acc=0.5054
OUT-OF-SAMPLE | XAU=     | acc=0.5351
OUT-OF-SAMPLE | GDX      | acc=0.5054
OUT-OF-SAMPLE | GLD      | acc=0.5189
```

Market Prediction Intraday

- Weakly efficient on an end-of-day basis
- Weakly inefficient intraday
  - Intraday Data
  - hourly data

Intraday Data

```
url = 'http://hilpisch.com/aiif_eikon_id_data.csv'
data = pd.read_csv(url, index_col=0, parse_dates=True) # .dropna()
data.tail()
```

Date	AAPL.O	MSFT.O	INTC.O	AMZN.O	GS.N	SPY	.SPX	.VIX	EUR=	XAU=	GDX	GLD
2019-12-31 20:00:00	292.36	157.2845	59.575	1845.22	228.92	320.94	3219.75	14.16	1.1215	1519.6451	29.40	143.12
2019-12-31 21:00:00	293.37	157.4900	59.820	1846.95	229.89	321.89	3230.56	13.92	1.1216	1517.3600	29.29	142.93
2019-12-31 22:00:00	293.82	157.9000	59.990	1850.20	229.93	322.39	3230.78	13.78	1.1210	1517.0100	29.30	142.90
2019-12-31 23:00:00	293.75	157.8300	59.910	1851.00	NaN	322.22	NaN	NaN	1.1211	1516.8900	29.40	142.88
2020-01-01 00:00:00	293.81	157.8800	59.870	1850.10	NaN	322.32	NaN	NaN	1.1211	NaN	29.34	143.00

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```python
lags = 5

dfs = {}
for ric in data:
   df, cols = add_lags(data, ric, lags)
   dfs[ric] = df, cols
```

```
train_test_model(model_mlp)
```

```
OUT-OF-SAMPLE | AAPL.O   | acc=0.5420
OUT-OF-SAMPLE | MSFT.O   | acc=0.4930
OUT-OF-SAMPLE | INTC.O   | acc=0.5549
OUT-OF-SAMPLE | AMZN.O   | acc=0.4709
OUT-OF-SAMPLE | GS.N     | acc=0.5184
OUT-OF-SAMPLE | SPY      | acc=0.4860
OUT-OF-SAMPLE | .SPX     | acc=0.5019
OUT-OF-SAMPLE | .VIX     | acc=0.4885
OUT-OF-SAMPLE | EUR=     | acc=0.5130
OUT-OF-SAMPLE | XAU=     | acc=0.4824
OUT-OF-SAMPLE | GDX      | acc=0.4765
OUT-OF-SAMPLE | GLD      | acc=0.5455
```

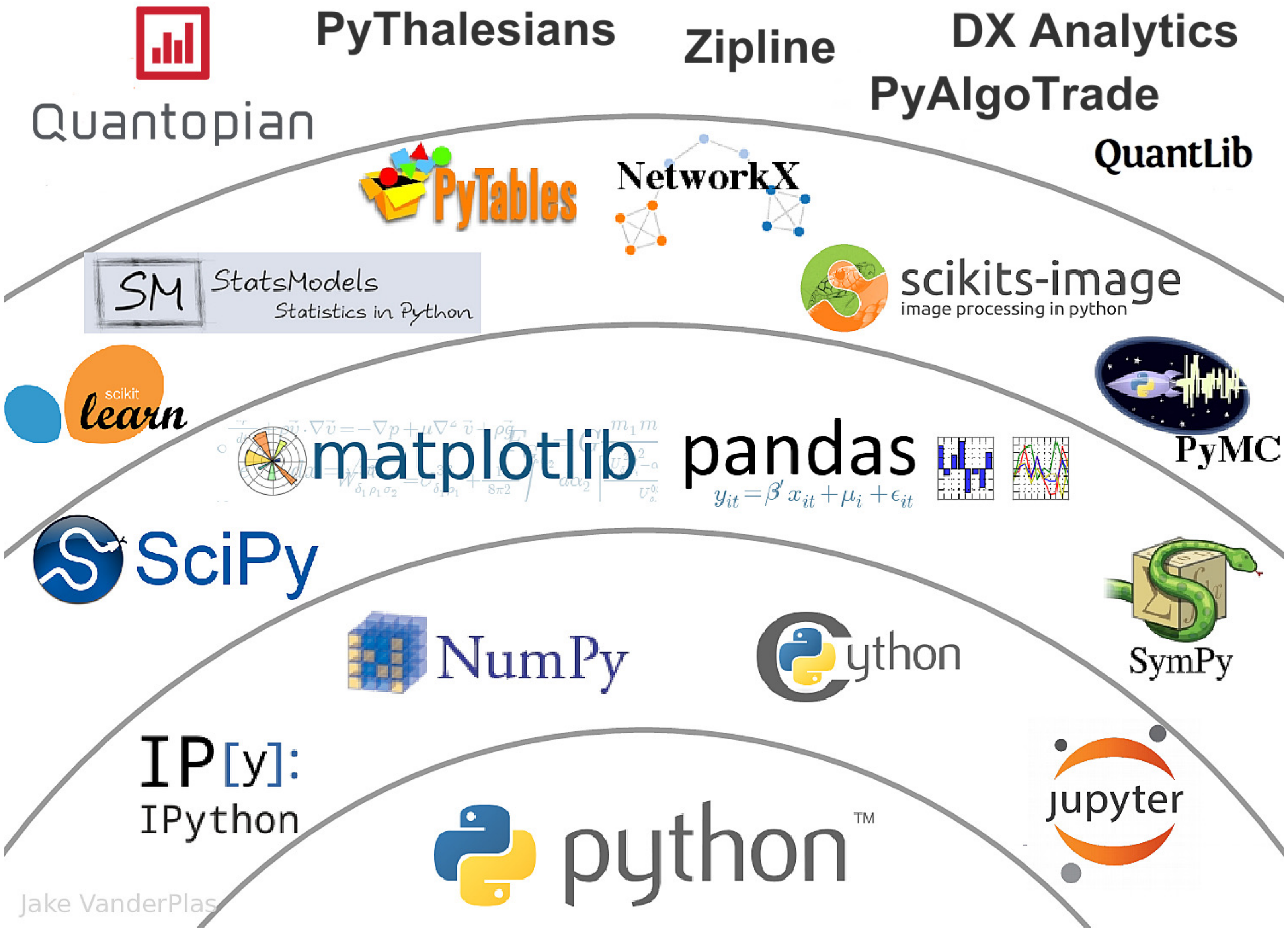
Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

```
train_test_model(model_bag)
```

```
OUT-OF-SAMPLE | AAPL.O   | acc=0.5660
OUT-OF-SAMPLE | MSFT.O   | acc=0.5551
OUT-OF-SAMPLE | INTC.O   | acc=0.5072
OUT-OF-SAMPLE | AMZN.O   | acc=0.4830
OUT-OF-SAMPLE | GS.N     | acc=0.5020
OUT-OF-SAMPLE | SPY      | acc=0.4680
OUT-OF-SAMPLE | .SPX     | acc=0.4677
OUT-OF-SAMPLE | .VIX     | acc=0.5161
OUT-OF-SAMPLE | EUR=     | acc=0.5242
OUT-OF-SAMPLE | XAU=     | acc=0.5229
OUT-OF-SAMPLE | GDX      | acc=0.5107
OUT-OF-SAMPLE | GLD      | acc=0.5475
```

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

The Quant Finance PyData Stack

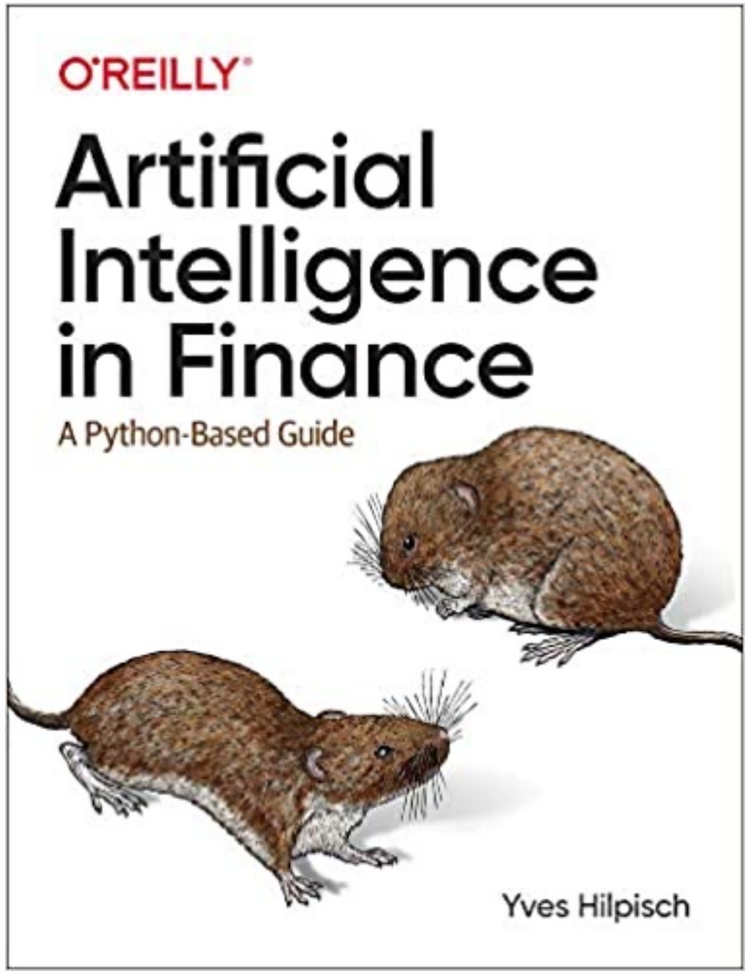


Source: http://nbviewer.jupyter.org/format/slides/github/quantopian/pyfolio/blob/master/pyfolio/examples/overview_slides.ipynb#/5

Yves Hilpisch (2020),

Artificial Intelligence in Finance: A Python-Based Guide,

O'Reilly



Source: https://www.amazon.com/Artificial-Intelligence-Finance-Python-Based-Guide/dp/1492055433

Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly

yhilpisch / aiif Public

https://github.com/yhilpisch/aiif

Notifications | Star 98 | Fork 77

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights

main | 1 branch | 0 tags | Go to file | Code

yves Code updates for TF 2.3. e334251 on Dec 8, 2020 | 4 commits

code	Code updates for TF 2.3.	11 months ago
.gitignore	Code updates for TF 2.3.	11 months ago
LICENSE.txt	Code updates.	11 months ago
README.md	Code updates.	11 months ago

README.md

Artificial Intelligence in Finance

About this Repository

This repository provides Python code and Jupyter Notebooks accompanying the **Artificial Intelligence in Finance** book published by O'Reilly.

O'REILLY

About

Jupyter Notebooks and code for the book Artificial Intelligence in Finance (O'Reilly) by Yves Hilpisch.

home.tpq.io/books/aiif

Readme

View license

Releases

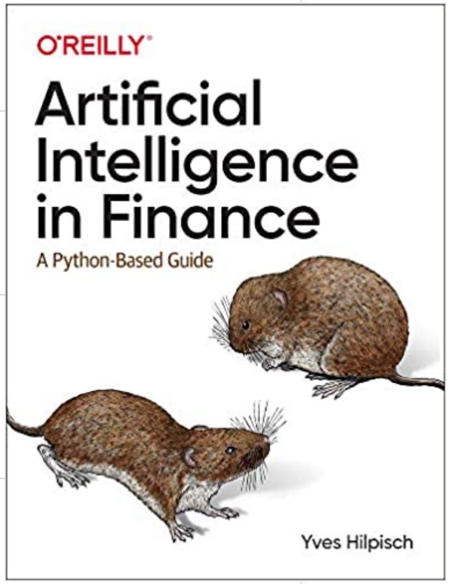
No releases published

Packages

No packages published

Languages

Jupyter Notebook 97.4% • Python 2.6%



Source: https://github.com/yhilpisch/aiif

Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly

yhilpisch / aiif Public

Notifications | Star 98 | Fork 77

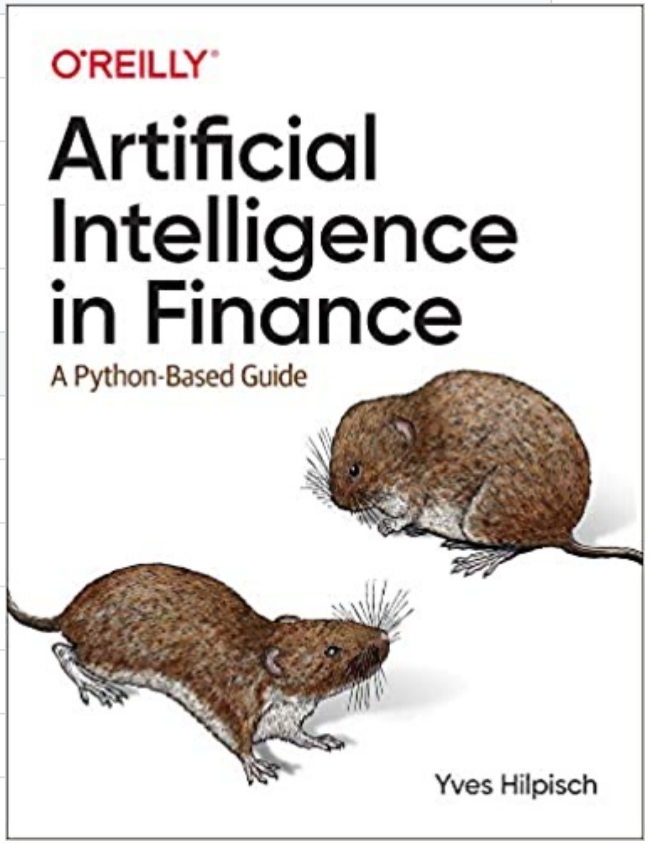
Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights

main | aiif / code /

https://github.com/yhilpisch/aiif/tree/main/code

Go to file

yves Code updates for TF 2.3.	e334251 on Dec 8, 2020 History
..	
oanda	Code updates for TF 2.3.
01_artificial_intelligence.ipynb	Code updates for TF 2.3.
02_superintelligence.ipynb	Code updates for TF 2.3.
03_normative_finance.ipynb	Code updates for TF 2.3.
04_data_driven_finance_a.ipynb	Initial commit.
04_data_driven_finance_b.ipynb	Initial commit.
05_machine_learning.ipynb	Code updates for TF 2.3.
06_ai_first_finance.ipynb	Code updates for TF 2.3.
07_dense_networks.ipynb	Code updates for TF 2.3.
08_recurrent_networks.ipynb	Code updates for TF 2.3.
09_reinforcement_learning_a.ipynb	Code updates.
09_reinforcement_learning_b.ipynb	Code updates for TF 2.3.



Source: https://github.com/yhilpisch/aiif/tree/main/code

Python in Google Colab (Python101)

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT

python101.ipynb - Colaboratory

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT?authuser=2#scrollTo=wsh36fLxDKC3

python101.ipynb

File Edit View Insert Runtime Tools Help

COMMENT SHARE

CODE TEXT CELL CELL CONNECTED EDITING

```python
# Future Value
pv = 100
r = 0.1
n = 7
fv = pv * ((1 + (r)) ** n)
print(round(fv, 2))
```

```
194.87
```

[11]

```python
amount = 100
interest = 10 #10% = 0.01 * 10
years = 7

future_value = amount * ((1 + (0.01 * interest)) ** years)
print(round(future_value, 2))
```

```
194.87
```

[12]

```python
# Python Function def
def getfv(pv, r, n):
    fv = pv * ((1 + (r)) ** n)
    return fv
fv = getfv(100, 0.1, 7)
print(round(fv, 2))
```

```
194.87
```

[13]

```python
# Python if else
score = 80
if score >=60 :
    print("Pass")
else:
    print("Fail")
```

```
Pass
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM
Disk

Editing

Table of contents

AI in Finance

AI in Finance

- Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.
- Github: https://github.com/yhilpisch/aiif/

Normative Finance and Financial Theories

Uncertainty and Risk

```python
import numpy as np

#The prices of the stock and bond today.
S0 = 10
B0 = 10
print('S0', S0)
print('S0', S0)

#The uncertain payoff of the stock and bond tomorrow.
S1 = np.array((20, 5))
B1 = np.array((11, 11))
print('S1', S1)
print('B1', B1)

#The market price vector
M0 = np.array((S0, B0))
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text RAM Disk Editing

Table of contents

Data Driven Finance

Data Driven Finance

Financial Econometrics and Regression

[18]

```python
import numpy as np

def f(x):
    return 2 + 1 / 2 * x

x = np.arange(-4, 5)
x
```

```
array([-4, -3, -2, -1,  0,  1,  2,  3,  4])
```

```python
y = f(x)
y
```

```
array([ 0.00,  0.50,  1.00,  1.50,  2.00,  2.50,  3.00,  3.50,  4.00])
```

```python
print('x', x)

print('y', y)

beta = np.cov(x, y, ddof=0)[0, 1] / x.var()
print('beta', beta)
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text RAM Disk Editing

Table of contents

Machine Learning

Machine Learning

Data

```python
import numpy as np
import pandas as pd
from pylab import plt, mpl
np.random.seed(100)
plt.style.use('seaborn')
mpl.rcParams['savefig.dpi'] = 300
mpl.rcParams['font.family'] = 'serif'

url = 'http://hilpisch.com/aiif_eikon_eod_data.csv'

raw = pd.read_csv(url, index_col=0, parse_dates=True)['EUR=']
raw.head()
```

```
Date
2010-01-01    1.4323
2010-01-04    1.4411
2010-01-05    1.4368
2010-01-06    1.4412
2010-01-07    1.4318
Name: EUR=, dtype: float64
```

[2]

```python
raw.tail()
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help Saving...

Comment Share

+ Code + Text RAM Disk Editing

Table of contents

Success

```
def MSE(l, p):
    return np.mean((l - p) ** 2)
```

```
[9]
reg = np.polyfit(f, l, deg=5)
reg
```

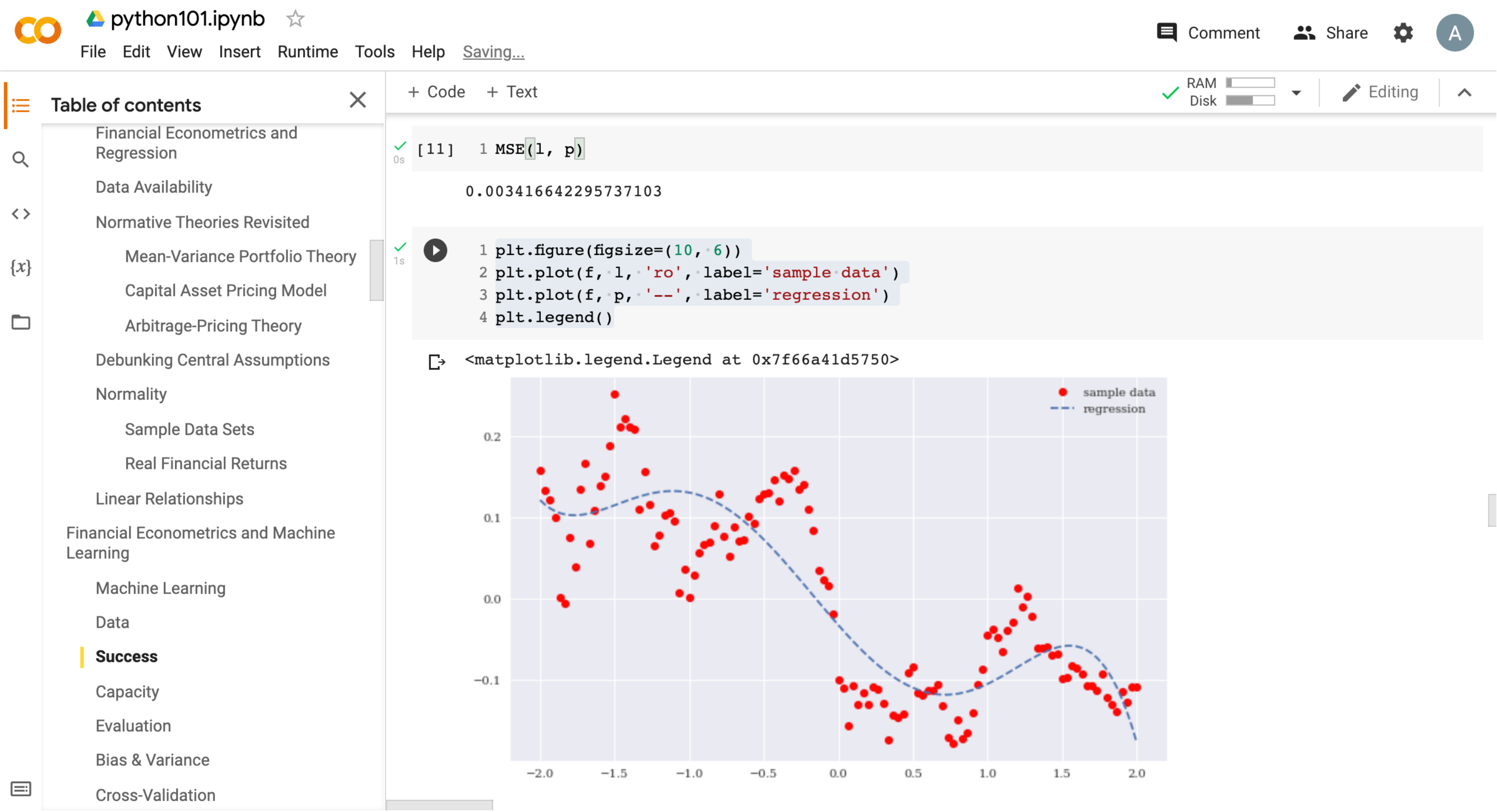
```
array([-0.01910626, -0.0147182 ,  0.10990388,  0.06007211, -0.20833598,
       -0.03275423])
```

```
[10]
p = np.polyval(reg, f)
p
```

```
array([ 0.12088427,  0.11526131,  0.11080193,  0.10739461,  0.10493286,
        0.10331514,  0.10244475,  0.10222973,  0.10258281,  0.10342126,
        0.10466683,  0.10624564,  0.1080881 ,  0.1101288 ,  0.11230643,
        0.11456366,  0.11684709,  0.11910711,  0.12129784,  0.123377  ,
        0.12530587,  0.12704913,  0.12857481,  0.1298542 ,  0.1308617 ,
        0.1315748 ,  0.13197395,  0.13204243,  0.13176634,  0.13113443,
        0.13013803,  0.12877097,  0.12702948,  0.12491207,  0.12241947,
        0.11955452,  0.11632208,  0.11272891,  0.10878364,  0.1044966 ,
        0.09987977,  0.09494668,  0.0897123 ,  0.08419296,  0.07840627,
        0.07237098,  0.06610693,  0.05963494,  0.05297671,  0.04615473,
        0.03919218,  0.03211286,  0.02494106,  0.01770149,  0.01041918,
        0.00311939, -0.00417251, -0.0114311 , -0.01863101, -0.02574704,
       -0.03275423, -0.03962796, -0.04634406, -0.05287887, -0.05920936,
       -0.06531322, -0.07116897, -0.07675602, -0.08205478, -0.08704677,
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)



https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Table of contents

+ Code + Text RAM Disk Editing

Capacity

[21]

```
reg = {}
for d in range(1, 12, 2):
    reg[d] = np.polyfit(f, l, deg=d)
    p = np.polyval(reg[d], f)
    mse = MSE(l, p)
    print(f'{d:2d} | MSE={mse}')
```

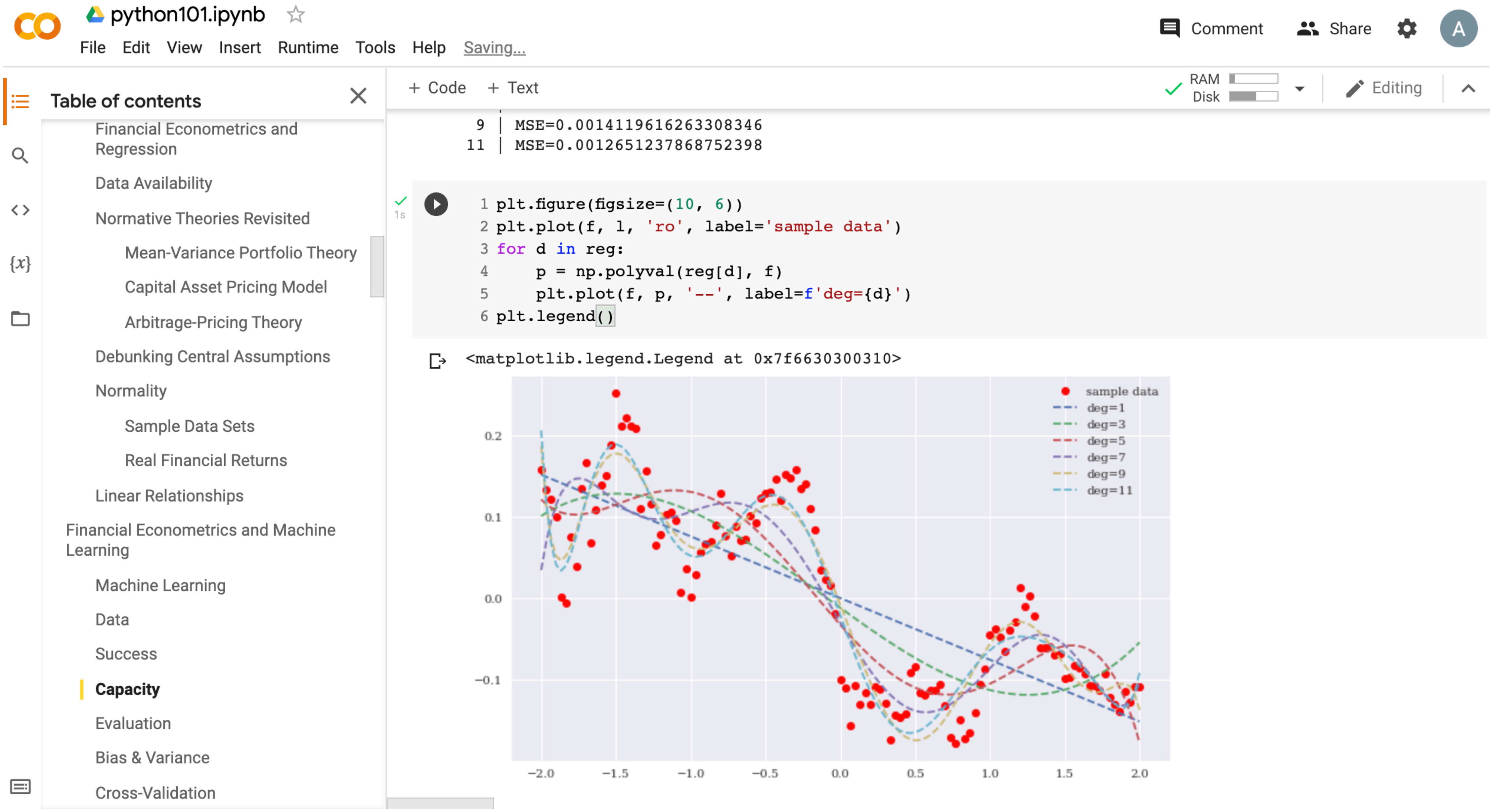
```
 1 | MSE=0.005322474034260403
 3 | MSE=0.004353110724143185
 5 | MSE=0.0034166422957371O3
 7 | MSE=0.002738950177235401
 9 | MSE=0.001411961626330 8346
11 | MSE=0.0012651237868752398
```

```
plt.figure(figsize=(10, 6))
plt.plot(f, l, 'ro', label='sample data')
for d in reg:
    p = np.polyval(reg[d], f)
    plt.plot(f, p, '--', label=f'deg={d}')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x7f6630300310>
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)



https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

```
def create_dnn_model(hl=1, hu=256):
    ''' Function to create Keras DNN model.
    Parameters
    ==========
    hl: int
    number of hidden layers
    hu: int
    number of hidden units (per layer)
    '''
    model = Sequential()
    for _ in range(hl):
        model.add(Dense(hu, activation='relu', input_dim=1))
    model.add(Dense(1, activation='linear'))
    model.compile(loss='mse', optimizer='rmsprop')
    return model

model = create_dnn_model(3)

model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_2 (Dense)             (None, 256)               512

 dense_3 (Dense)             (None, 256)               65792

 dense_4 (Dense)             (None, 256)               65792

 dense_5 (Dense)             (None, 1)                 257

=================================================================
Total params: 132,353
Trainable params: 132,353
Non-trainable params: 0
_________________________________________________________________
```

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

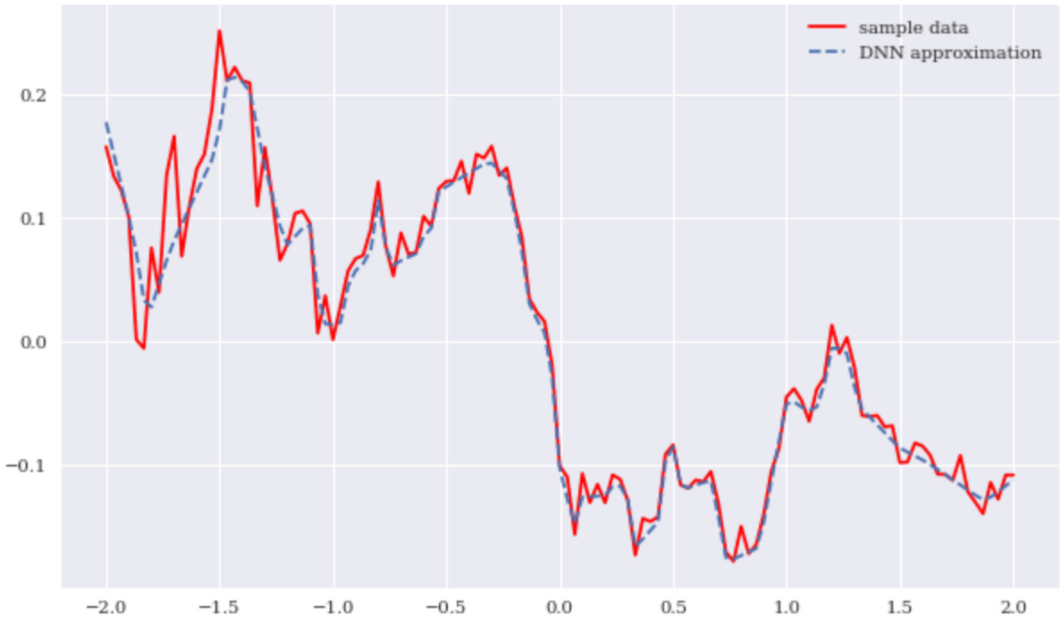
RAM Disk

Editing

Table of contents

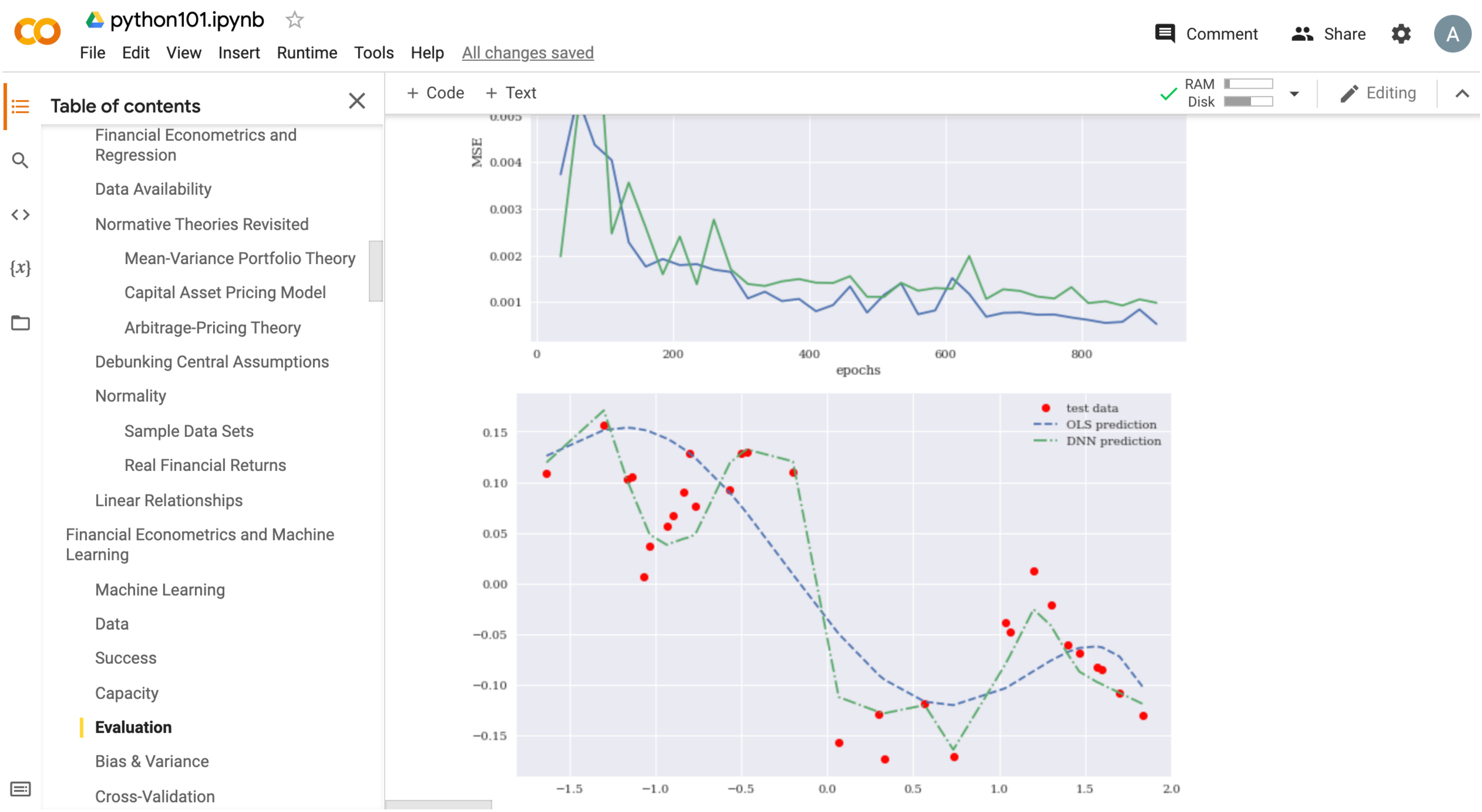

```
<keras.callbacks.History at 0x7f66301596f0>
```

```
p = model.predict(f).flatten()

MSE(l, p)

plt.figure(figsize=(10, 6))
plt.plot(f, l, 'r', label='sample data')
plt.plot(f, p, '--', label='DNN approximation')
plt.legend();
```



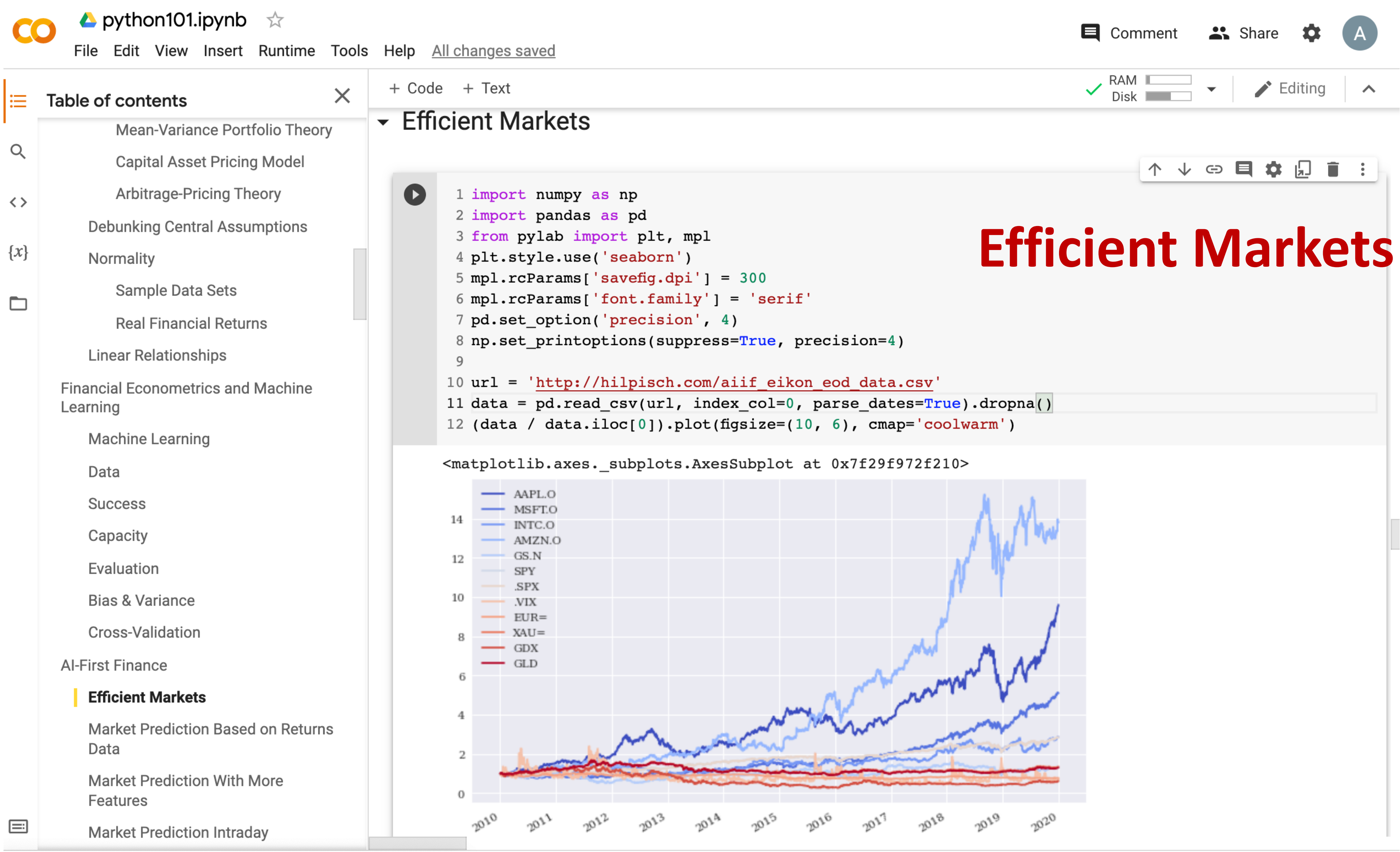
https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)



https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)



https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text RAM Disk Editing

Table of contents

Market Prediction Based on Returns Data

```
rets = np.log(data / data.shift(1))

rets.dropna(inplace=True)
rets
```

Market Prediction Based on Returns Data

Date	AAPL.O	MSFT.O	INTC.O	AMZN.O	GS.N	SPY	.SPX	.VIX	EUR=	XAU=	GDX	GLD
2010-01-05	0.0017	0.0003	-0.0005	0.0059	0.0175	2.6436e-03	3.1108e-03	-0.0350	-2.9883e-03	-0.0012	0.0096	-0.0009
2010-01-06	-0.0160	-0.0062	-0.0034	-0.0183	-0.0107	7.0379e-04	5.4538e-04	-0.0099	3.0577e-03	0.0176	0.0240	0.0164
2010-01-07	-0.0019	-0.0104	-0.0097	-0.0172	0.0194	4.2124e-03	3.9933e-03	-0.0052	-6.5437e-03	-0.0058	-0.0049	-0.0062
2010-01-08	0.0066	0.0068	0.0111	0.0267	-0.0191	3.3223e-03	2.8775e-03	-0.0500	6.5437e-03	0.0037	0.0150	0.0050
2010-01-11	-0.0089	-0.0128	0.0057	-0.0244	-0.0159	1.3956e-03	1.7452e-03	-0.0325	6.9836e-03	0.0144	0.0066	0.0132
...	...	...	...	...	...	...	...	...	...	...	...	...
2019-12-24	0.0010	-0.0002	0.0030	-0.0021	0.0036	3.1131e-05	-1.9543e-04	0.0047	9.0200e-05	0.0091	0.0315	0.0094
2019-12-26	0.0196	0.0082	0.0069	0.0435	0.0056	5.3092e-03	5.1151e-03	-0.0016	8.1143e-04	0.0083	0.0145	0.0078
2019-12-27	-0.0004	0.0018	0.0043	0.0006	-0.0024	-2.4775e-04	3.3951e-05	0.0598	7.0945e-03	-0.0006	-0.0072	-0.0004

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text RAM Disk Editing

Market Prediction with More Features

Market Prediction With More Features

```python
url = 'http://hilpisch.com/aiif_eikon_eod_data.csv'

data = pd.read_csv(url, index_col=0, parse_dates=True).dropna()
data
```

Date	AAPL.O	MSFT.O	INTC.O	AMZN.O	GS.N	SPY	.SPX	.VIX	EUR=	XAU=	GDX	GLD
2010-01-04	30.5728	30.950	20.88	133.90	173.08	113.33	1132.99	20.04	1.4411	1120.0000	47.71	109.80
2010-01-05	30.6257	30.960	20.87	134.69	176.14	113.63	1136.52	19.35	1.4368	1118.6500	48.17	109.70
2010-01-06	30.1385	30.770	20.80	132.25	174.26	113.71	1137.14	19.16	1.4412	1138.5000	49.34	111.51
2010-01-07	30.0828	30.452	20.60	130.00	177.67	114.19	1141.69	19.06	1.4318	1131.9000	49.10	110.82
2010-01-08	30.2828	30.660	20.83	133.52	174.31	114.57	1144.98	18.13	1.4412	1136.1000	49.84	111.37
...	...	...	...	...	...	...	...	...	...	...	...	...
2019-12-24	284.2700	157.380	59.41	1789.21	229.91	321.23	3223.38	12.67	1.1087	1498.8100	28.66	141.27
2019-12-26	289.9100	158.670	59.82	1868.77	231.21	322.94	3239.91	12.65	1.1096	1511.2979	29.08	142.38
2019-12-27	289.8000	158.960	60.08	1869.80	230.66	322.86	3240.02	13.43	1.1175	1510.4167	28.87	142.33
2019-12-30	291.5200	157.590	59.62	1846.89	229.80	321.08	3221.29	14.82	1.1197	1515.1230	29.49	142.63
2019-12-31	293.6500	157.700	59.85	1847.84	229.93	321.86	3230.78	13.78	1.1210	1517.0100	29.28	142.90

2516 rows × 12 columns

https://tinyurl.com/aintpupython101

Python in Google Colab (Python101)

python101.ipynb

File Edit View Insert Runtime Tools Help Saving...

Comment Share

+ Code + Text RAM Disk Editing

Table of contents

Market Prediction Intraday

Market Prediction Intraday

```
url = 'http://hilpisch.com/aiif_eikon_id_data.csv'
data = pd.read_csv(url, index_col=0, parse_dates=True) # .dropna()
data.tail()
```

Date	AAPL.O	MSFT.O	INTC.O	AMZN.O	GS.N	SPY	.SPX	.VIX	EUR=	XAU=	GDX	GLD
2019-12-31 20:00:00	292.36	157.2845	59.575	1845.22	228.92	320.94	3219.75	14.16	1.1215	1519.6451	29.40	143.12
2019-12-31 21:00:00	293.37	157.4900	59.820	1846.95	229.89	321.89	3230.56	13.92	1.1216	1517.3600	29.29	142.93
2019-12-31 22:00:00	293.82	157.9000	59.990	1850.20	229.93	322.39	3230.78	13.78	1.1210	1517.0100	29.30	142.90
2019-12-31 23:00:00	293.75	157.8300	59.910	1851.00	NaN	322.22	NaN	NaN	1.1211	1516.8900	29.40	142.88
2020-01-01 00:00:00	293.81	157.8800	59.870	1850.10	NaN	322.32	NaN	NaN	1.1211	NaN	29.34	143.00

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 5529 entries, 2019-03-01 00:00:00 to 2020-01-01 00:00:00
Data columns (total 12 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   AAPL.O   3384 non-null   float64
 1   MSFT.O   3378 non-null   float64
```

https://tinyurl.com/aintpupython101

Summary

- Financial Econometrics
  - Financial Theories, OLS Regression
- Machine Learning
  - Learning, Evaluation, Bias and variance
  - Cross-validation
- AI-First Finance
  - Efficient Markets
  - Market Prediction Based on Returns Data
  - Market Prediction with More Features

Source: Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O'Reilly Media.

References

- Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly Media, https://github.com/yhilpisch/aiif .
- Chris Brooks (2019), Introductory Econometrics for Finance, 4th Edition, Cambridge University Press
- Oliver Linton (2019), Financial Econometrics: Models and Methods, Cambridge University Press
- Tom Mitchell (1997), Machine Learning, McGraw-Hill.
- Max Tegmark (2017), Life 3.0: Being human in the age of artificial intelligence. Vintage.
- Ajay Agrawal, Joshua Gans, and Avi Goldfarb (2018). Prediction machines: the simple economics of artificial intelligence. Harvard Business Press.
- Eugene F. Fama (1995), "Random walks in stock market prices." Financial Analysts Journal 51, no. 1, 75-80.
- Ruey S. Tsay (2005), Analysis of financial time series, Wiley.
- Min-Yuh Day (2023), Python 101, https://tinyurl.com/aintpupython101