LECTURER: TAI LE QUY

MACHINE LEARNING – SUPERVISED LEARNNG

Who am I?

- Name: Tai Le Quy
- PhD candidate at L3S Research Center – Leibniz University Hannover
  - Topic: Fairness-aware machine learning in educational data mining
  - Project: LernMINT (lernmint.org)
- MSc in Information Technology at National University of Vietnam
- Profile: tailequy.github.io
- Email: tai.le-quy@iu.org
- **https://github.com/tailequy/IU-ML-Supervised**

Who are you?

- Name
- Employer
- Position/responsibilities
- Fun Fact
- Previous knowledge? Expectations?

UNIT 1

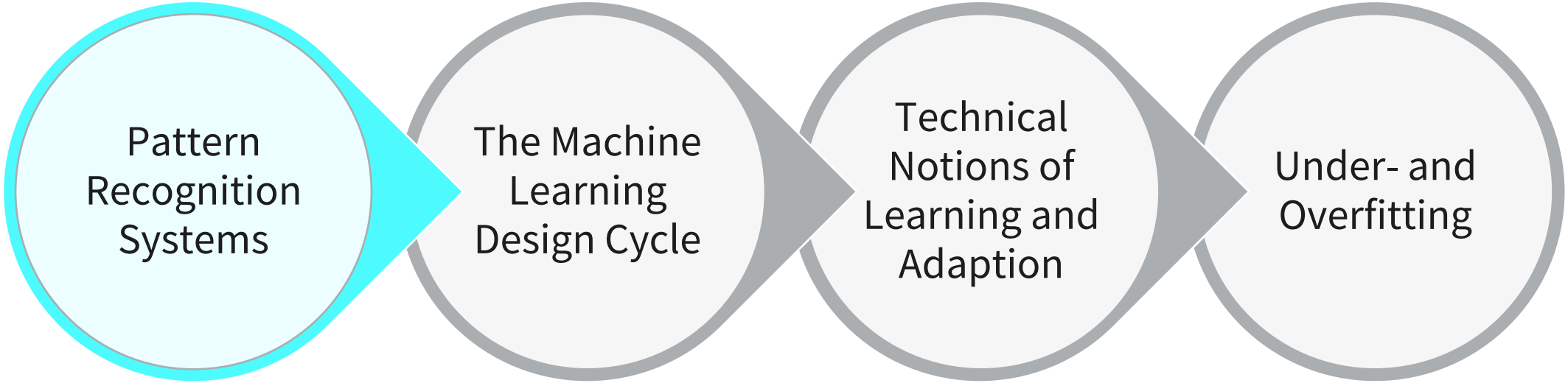
INTRODUCTION TO MACHINE LEARNING

STUDY GOALS

- define **machine learning** and fully understand the concept of **supervised learning**.
- utilize the machine learning **design cycle** as an iterative process model for building machine learning systems.
- understand the **technical notions** of learning and adaption and how they relate to one other.
- detect and avoid problems of **over-** and **underfitting**.

EXPLAIN SIMPLY

1. What is machine learning/supervised learning?
2. How is pattern recognition related to machine learning?
3. How can we design and implement a machine learning model?
4. What should we be aware of during the machine learning process


Pattern Recognition Systems
The Machine Learning Design Cycle
Technical Notions of Learning and Adaption
Under- and Overfitting

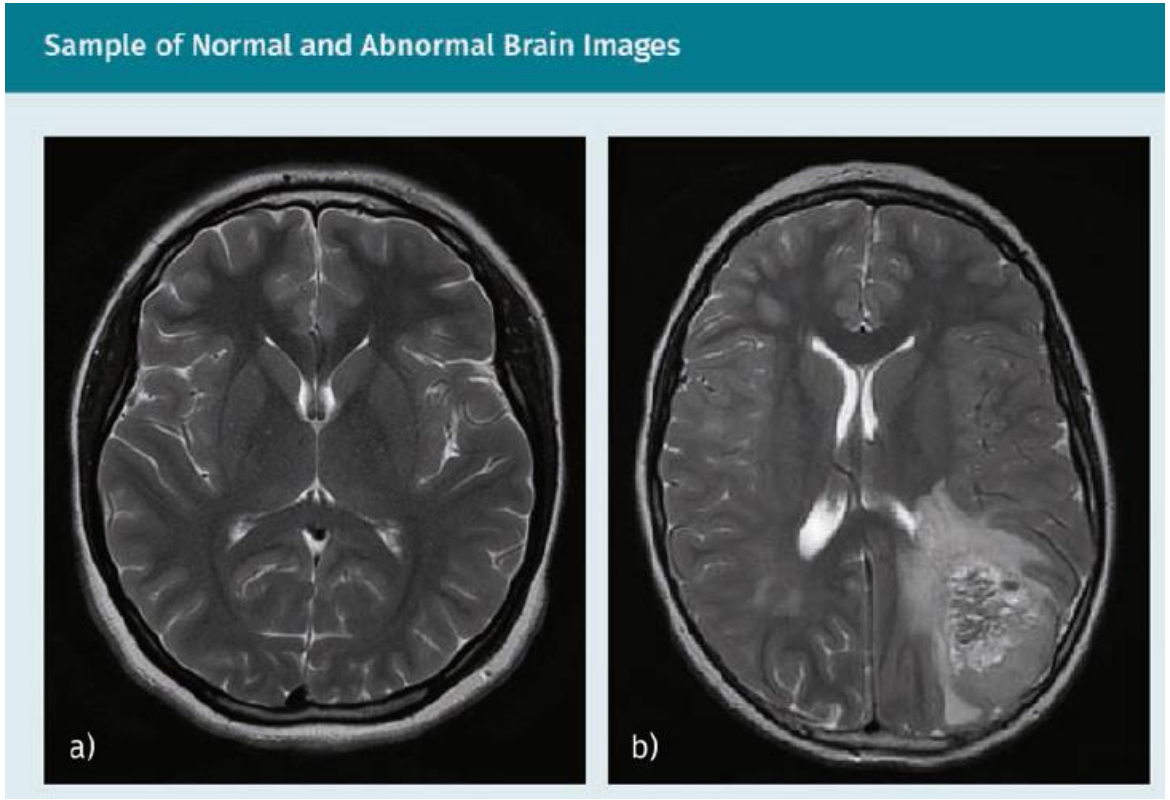
- **Pattern recognition** is the process of using **machine learning** techniques to assign a label to a given observation based on its features (i.e., the observation's attributes).
  - "to learn" is to capture the dependency structures between features and labels in order to generalize them and thereby predict labels for new, unseen observations

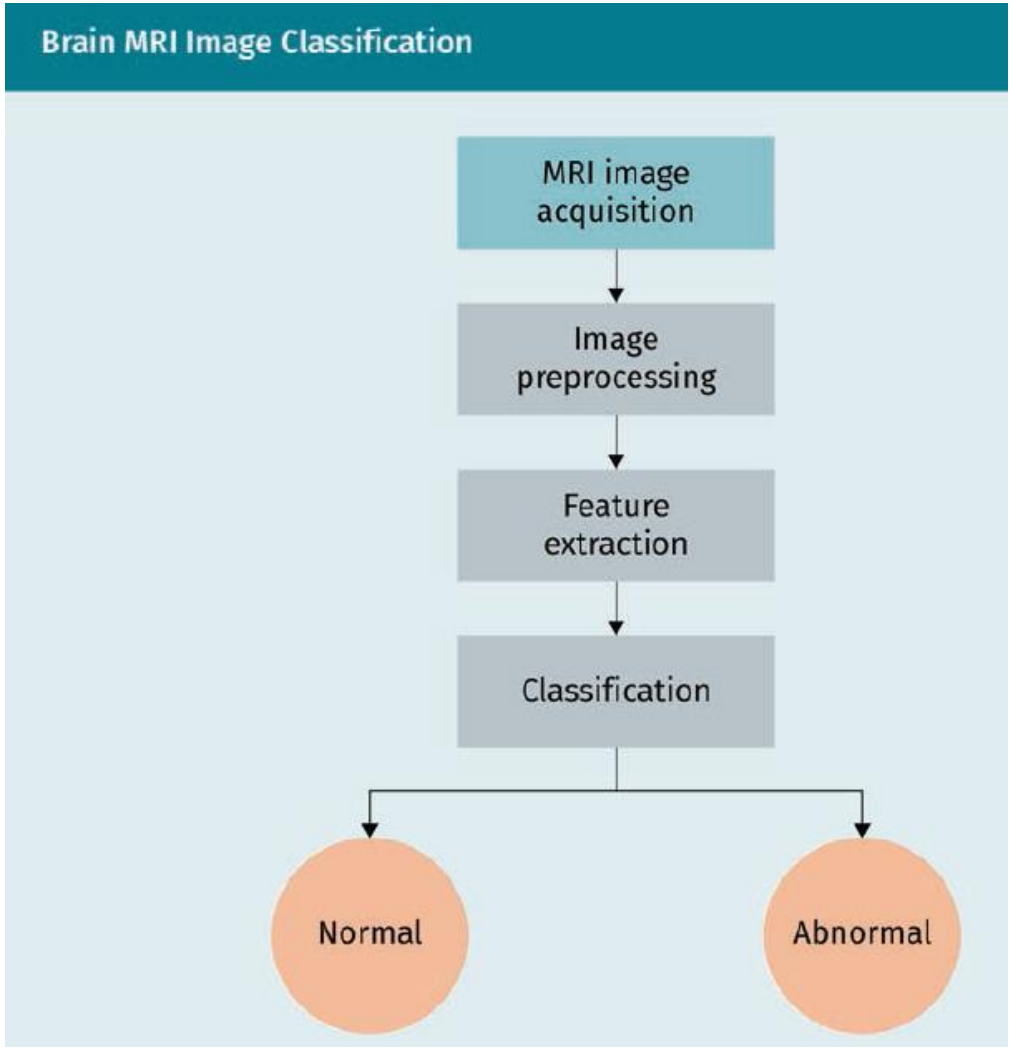
- Type of machine learning
  - Supervised learning: features and the labels are provided
  - Unsupervised learning: only features

fruit	length	width	weight	label
fruit 1	165	38	172	Banana
fruit 2	218	39	230	Banana
fruit 3	76	80	145	Orange
fruit 4	145	35	150	Banana
fruit 5	90	88	160	Orange
…				
fruit n	…	…	…	…

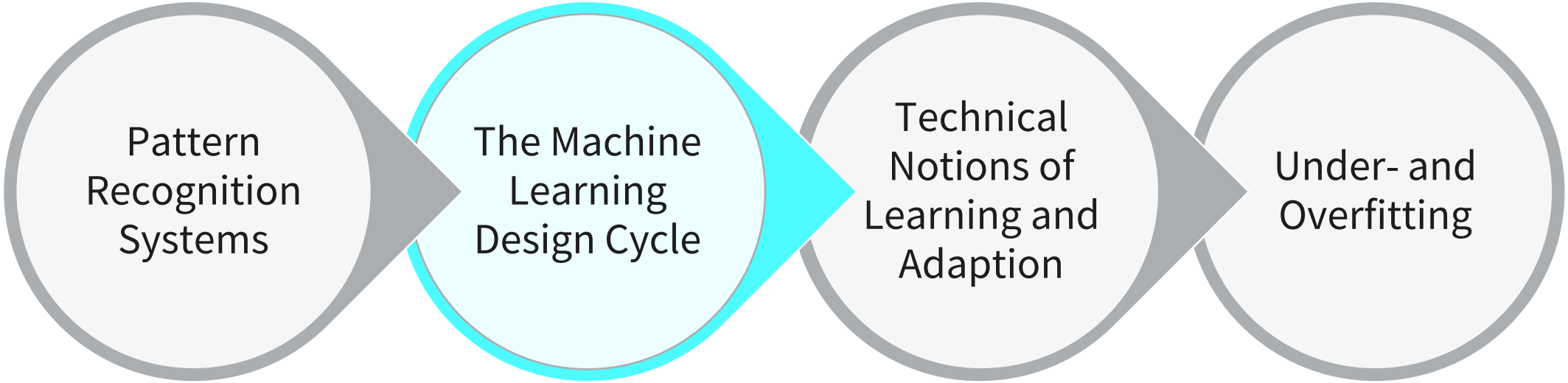
fruit	length	width	weight
fruit 1	165	38	172
fruit 2	218	39	230
fruit 3	76	80	145
fruit 4	145	35	150
fruit 5	90	88	160
…			
fruit n	…	…	…

PATTERN RECOGNITION IN THE REAL WORLD





Source of image: Course book


Pattern Recognition Systems
The Machine Learning Design Cycle
Technical Notions of Learning and Adaption
Under- and Overfitting

- **Cross-Industry Standard Process for Data Mining (CRISP-DM)**
  - Business understanding
  - Data understanding
  - Data preparation
  - Modeling
  - Evaluation
  - Deployment

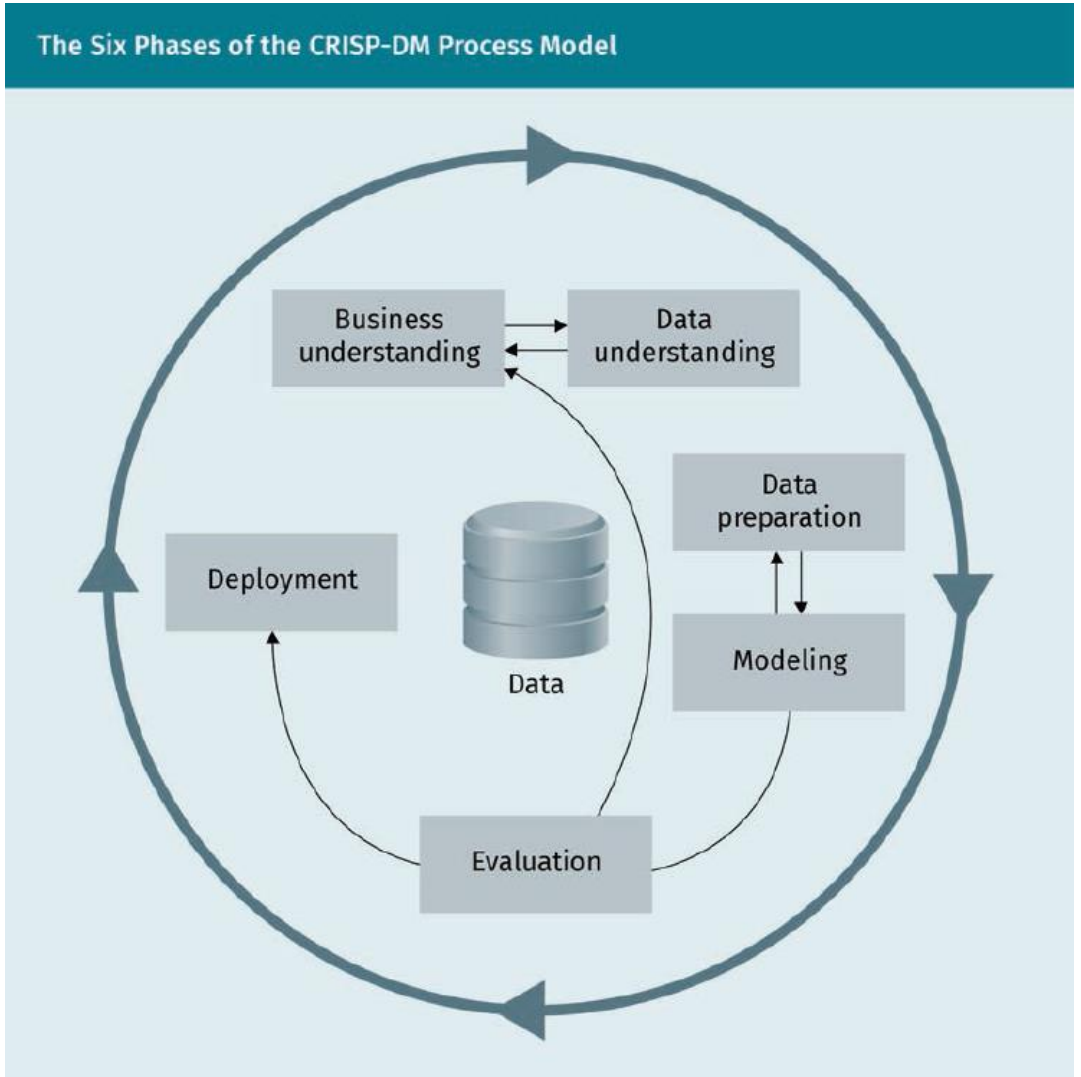
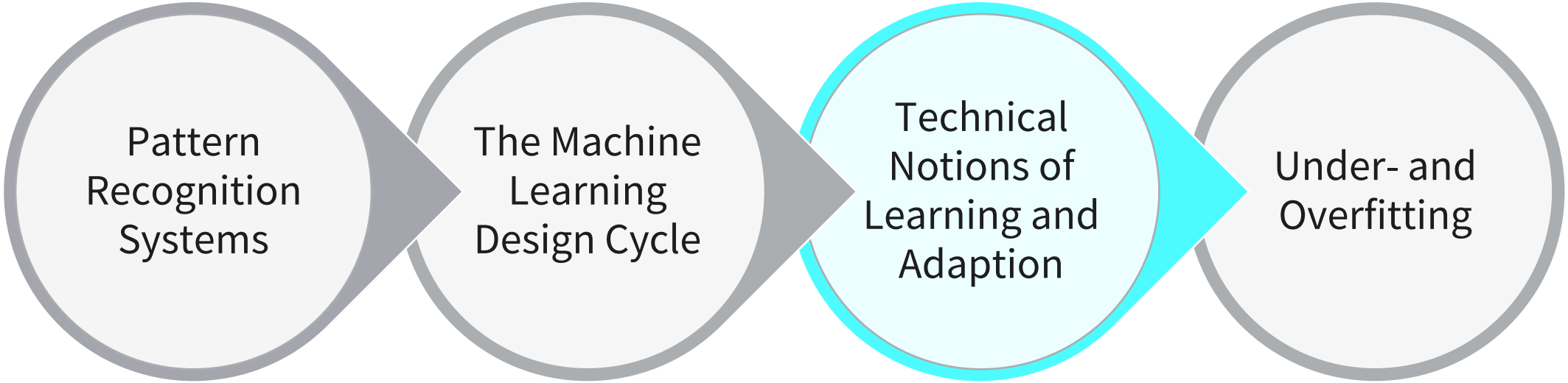


Photo: Markus Spiske on Unsplash


Pattern Recognition Systems
The Machine Learning Design Cycle
Technical Notions of Learning and Adaption
Under- and Overfitting

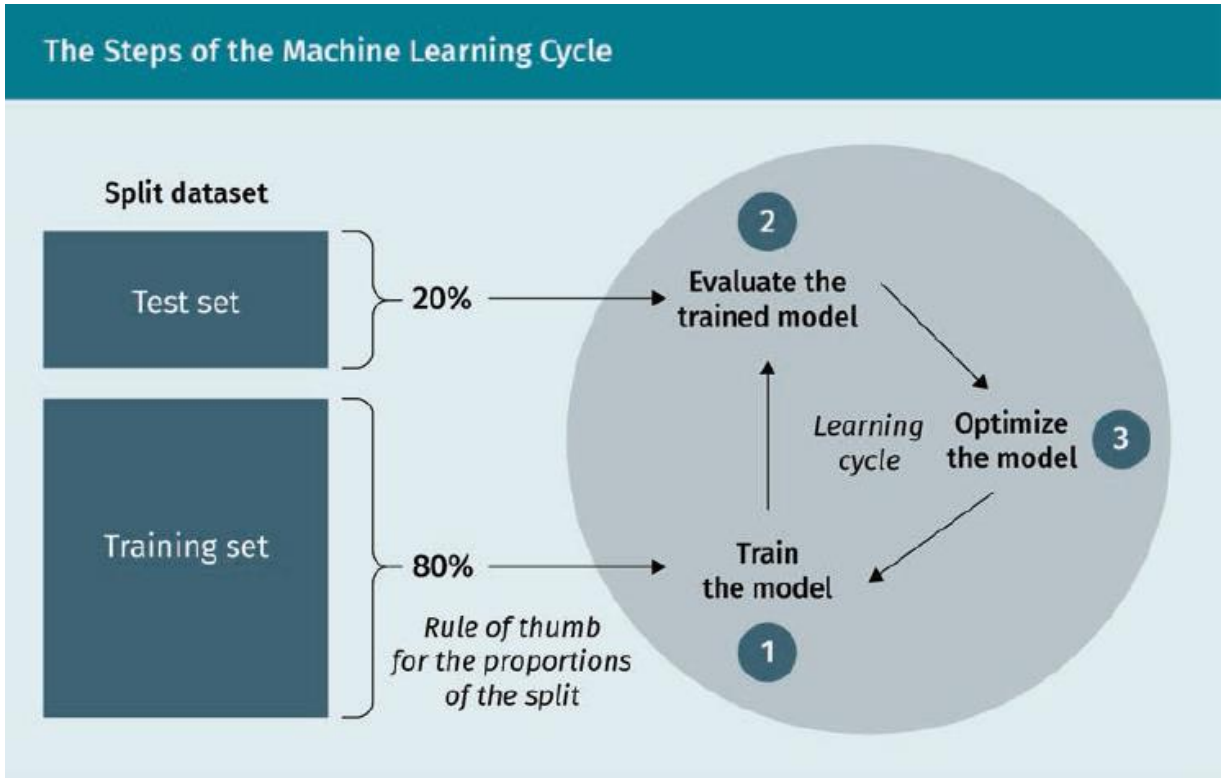
Definition of Machine Learning

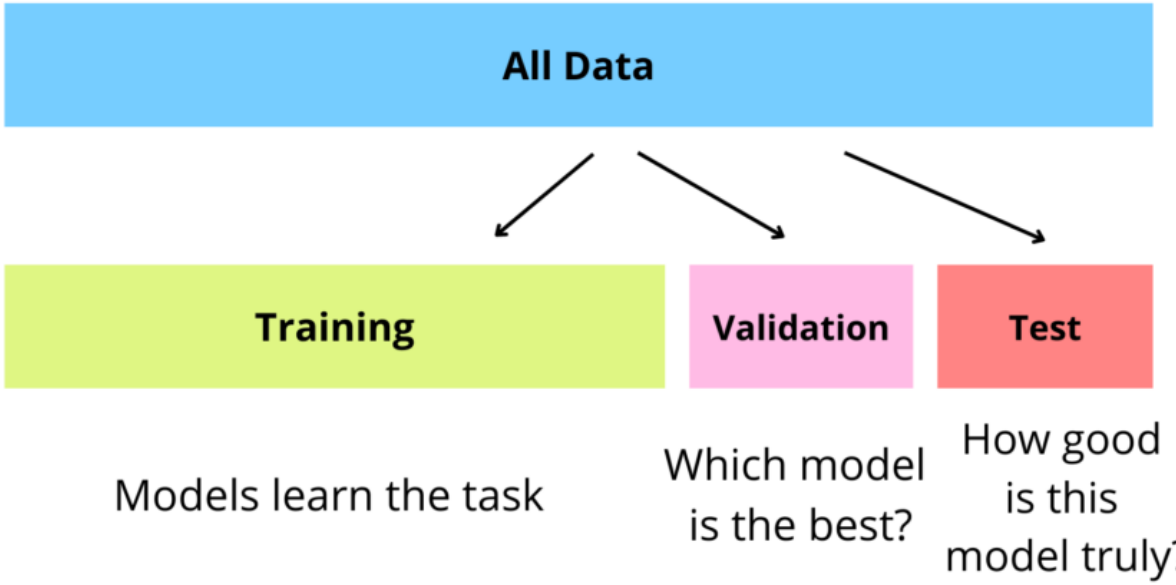
- Machine learning is a field of computer science that focuses on teaching computers to learn what humans do naturally: learn from data and past experience.
- “A computer program is said to learn from experience E with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**” (Mitchell, 1997)

Source of the text: http://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf

- Regression tasks
  - Predict continuous numeric values
- Classification tasks
  - Predict pre-known classes
- Methods
  - Linear regression
  - Logistic regression
  - k-nearest neighbors (KNNs)
  - Support vector machines (SVMs)
  - Decision trees, etc.

PROCESS OF SUPERVISED MACHINE LEARNING

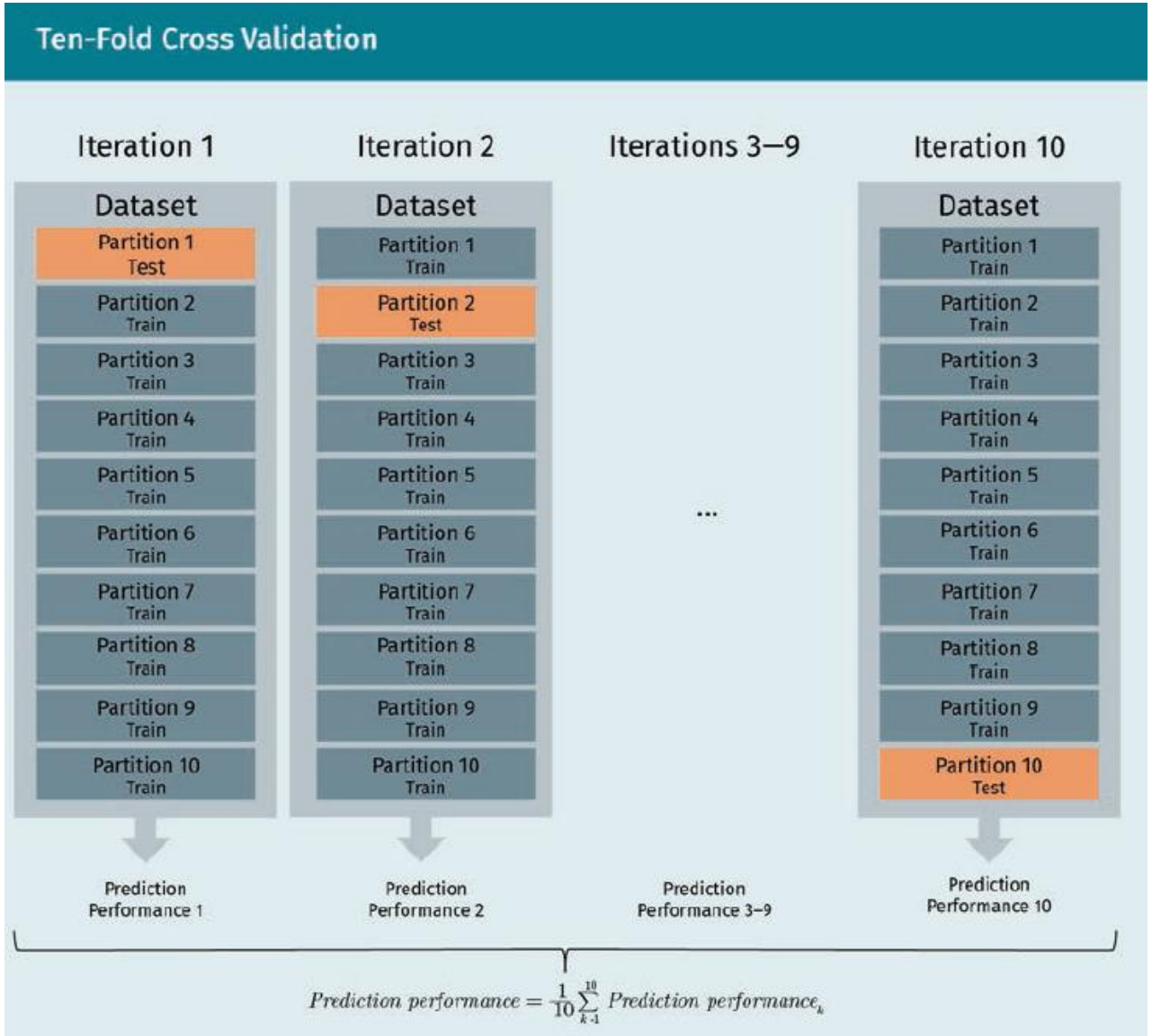




Source of image: Course book, https://galaxyinferno.com/what-is-validation-data-used-for-machine-learning-basics/

CROSS VALIDATION

- Cross-validation is a resampling method that uses **different portions** of the data to test and train a model on **different iterations** (k-fold cross validation)

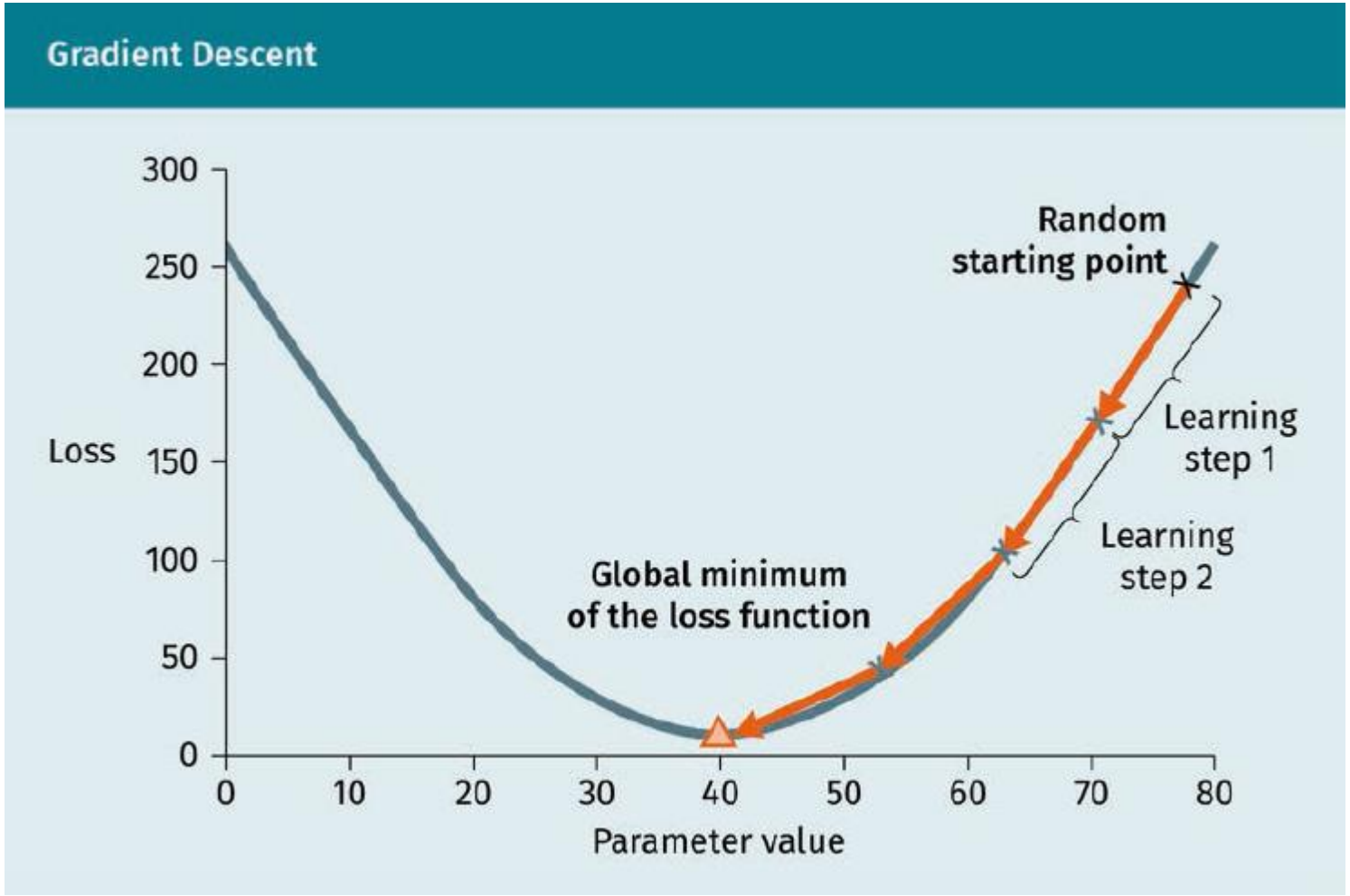


Source of iamge: Course book.

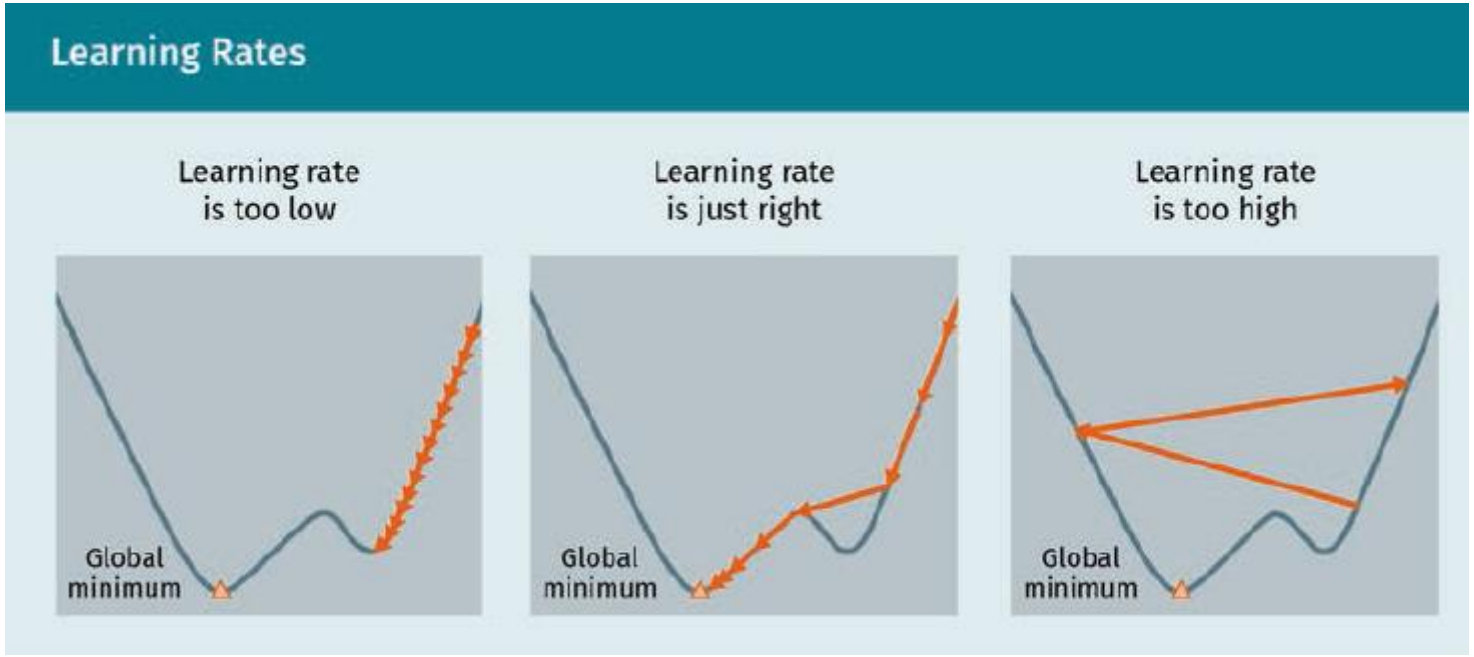
- Loss function or objective or target function:
  - Computes how much the predictions differ from the true values of the labels.
  - The model's accuracy improves based on the minimization of this loss function
  - Used to assess the quality of the model
- The optimizer
  - Iterates over the model's parameter values.
  - During each training iteration, it updates these values, thus incrementally minimizing the loss function and devising a solution

- The model parameters are iteratively tweaked to minimize the loss function (and find its global minimum)
  - Initial step: model parameters are initialized by random values
  - The goal is to work in the direction of the steepest descent of the loss function gradually, i.e., to take the step offering the largest possible loss reduction.
  - The gradient descent locates the steepest descent by determining the local gradient of the loss function (i.e., the partial derivative of the loss function) and goes in the negative direction of the gradient.
  - Repeated until the function's minimum has been reached

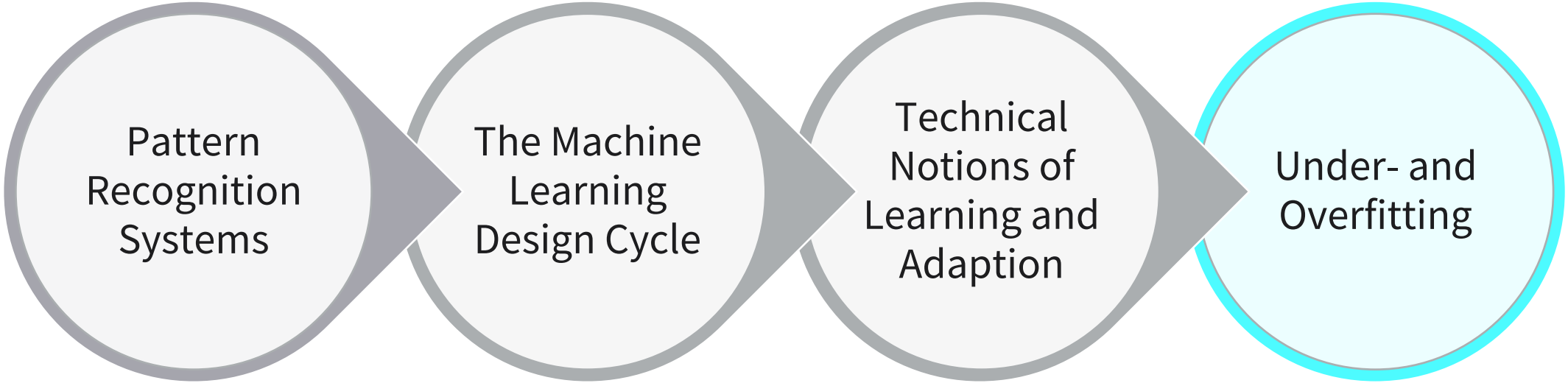
GRADIENT DESCENT



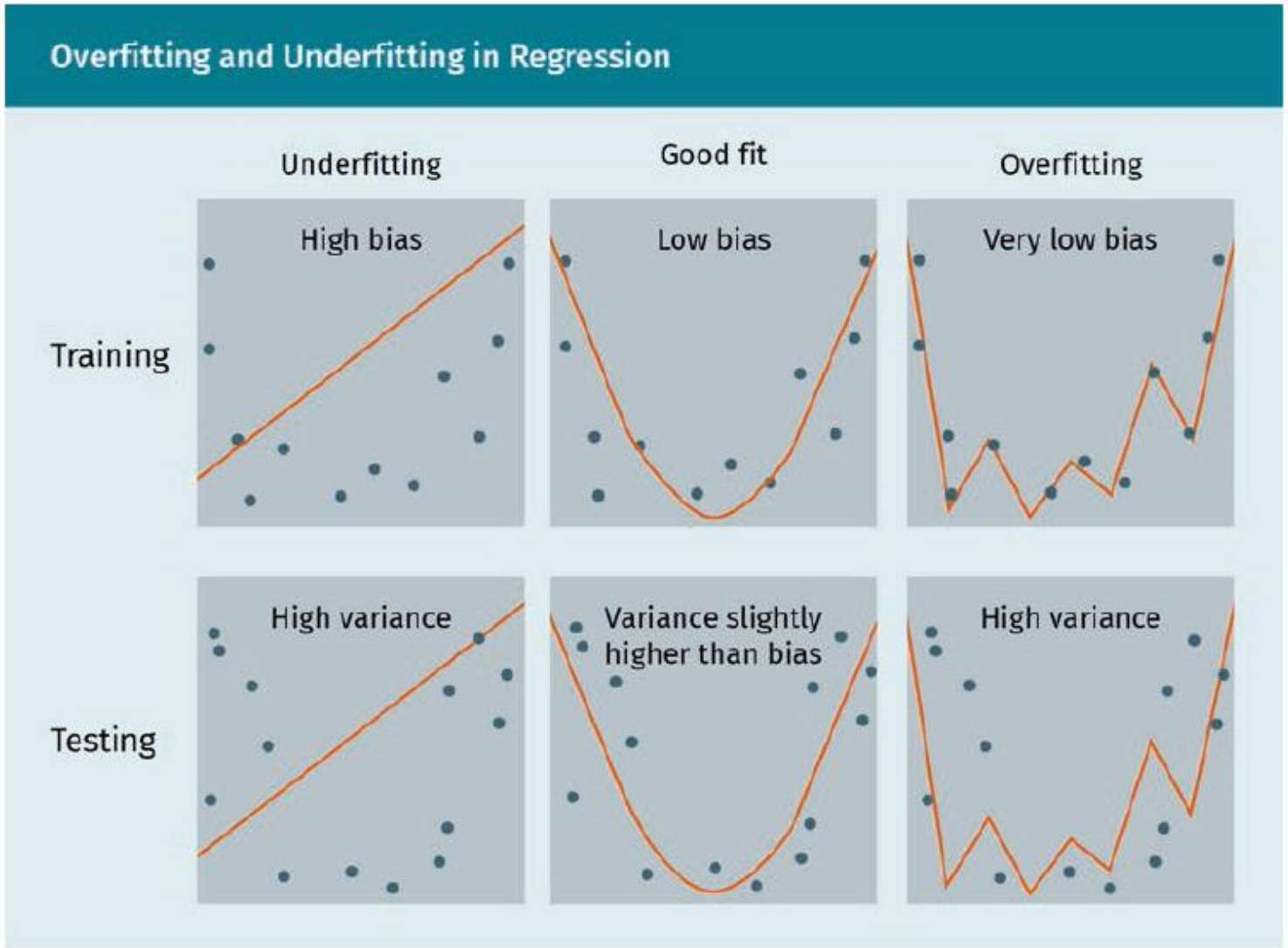
– Gradient descent is influenced by the chosen learning rate



– Stochastic gradient descent helps to solve the problem of local minima and plateaus

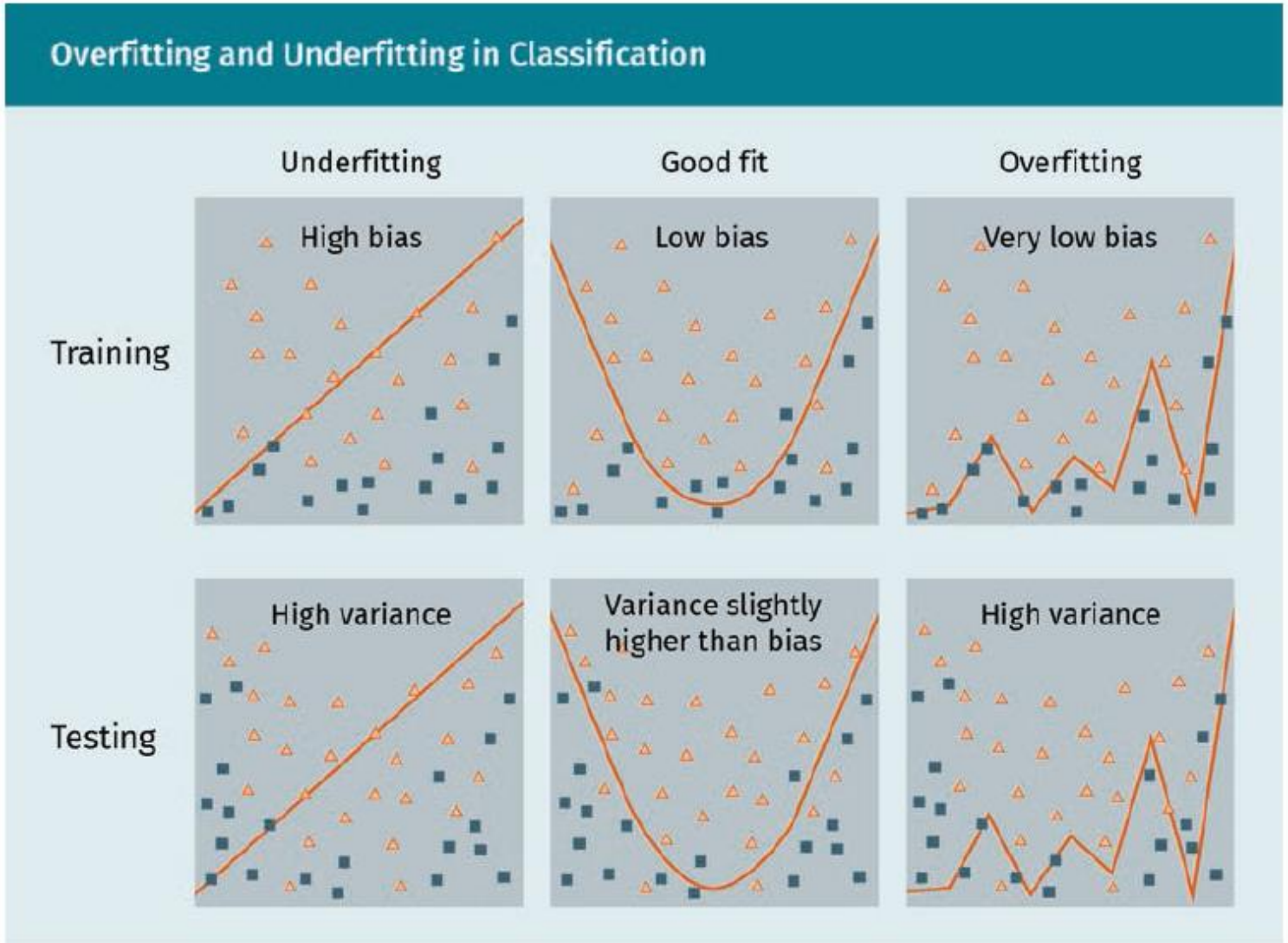

Pattern Recognition Systems
The Machine Learning Design Cycle
Technical Notions of Learning and Adaption
Under- and Overfitting

UNDER- AND OVERFITTING



Source of image: Course book.

UNDER- AND OVERFITTING



Source of image: Course book.

Avoiding Overfitting

- Choosing a less complex model
- Collecting more training data
- Reducing noise in the training data
- Using **regularization**

Avoiding Underfitting

- Choosing a more powerful, mathematically complex model
- Extracting better features
- Reducing constraints (e.g., reducing regularization) on the model

- define **machine learning** and fully understand the concept of **supervised learning**.
- utilize the machine learning **design cycle** as an iterative process model for building machine learning systems.
- understand the **technical notions** of learning and adaption and how they relate to one other.
- detect and avoid problems of **over-** and **underfitting**.

Session 1

TRANSFER TASK

Stock market prediction: Case study

– Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a **stock's future price** could yield significant profit.

– Create a rough project plan to achieve this goal. For each phase of this plan, explain which supervised machine learning techniques might be applied

Credit Score Classification: Case Study

– The **credit score** of a person determines the creditworthiness of the person. It helps financial companies determine if you can repay the loan or credit you are applying for.

– Create a rough project plan to achieve this goal. For each phase of this plan, explain which supervised machine learning techniques might be applied

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your results.

The results will be discussed in plenary.

LIST OF SOURCES

Mitchell, T. M. (1997). *Machine learning*. CMU Press