LECTURER: TAI LE QUY

MACHINE LEARNING

SUPERVISED LEARNING

Who am I?

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

INTRODUCTORY ROUND

Who are you?

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

TOPIC OUTLINE

UNIT 1

INTRODUCTION TO MACHINE LEARNING

STUDY GOALS

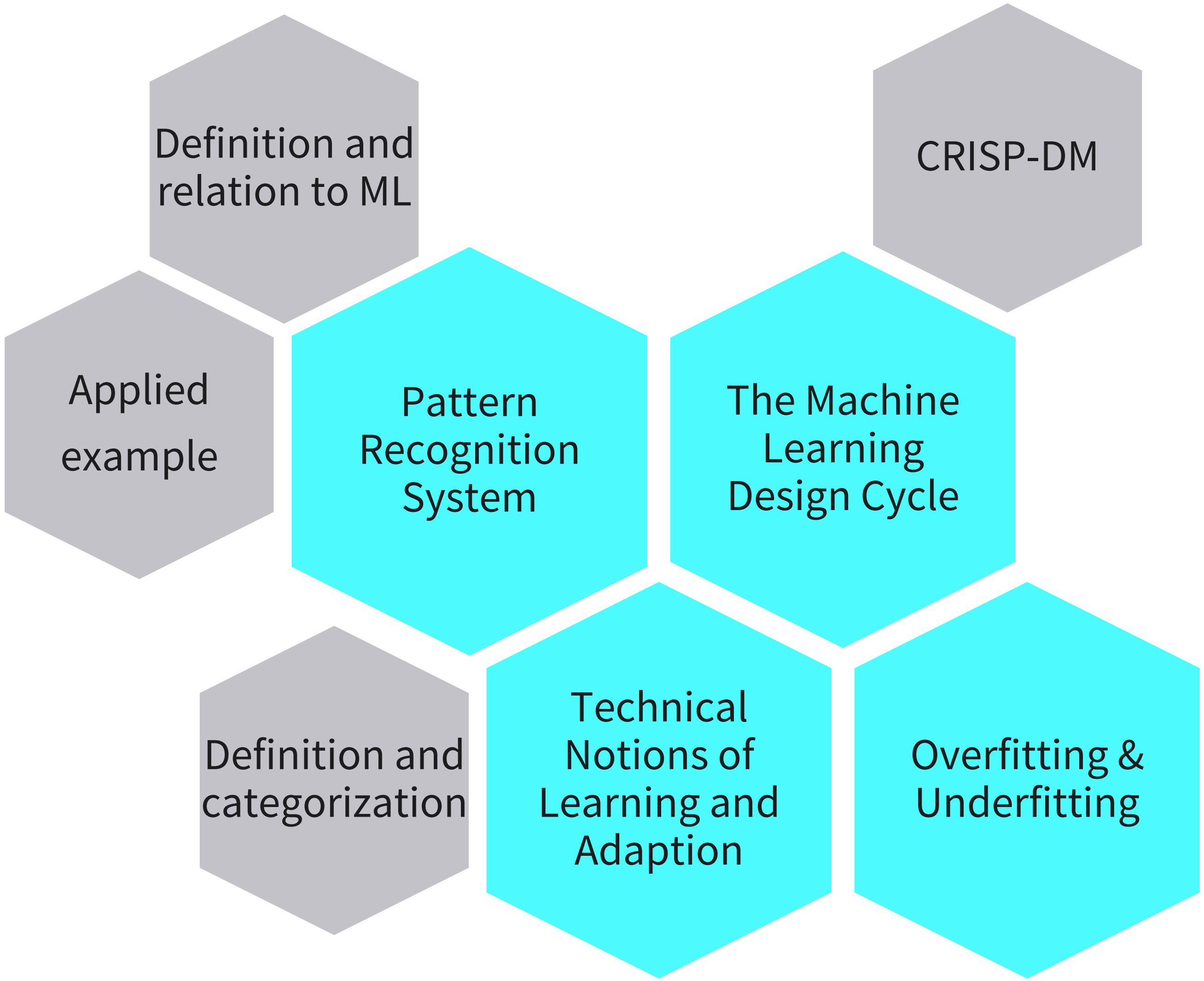
- **Define machine learning** and fully understand the concept of **supervised learning**.
- Utilize the **machine learning design cycle** as an iterative process model for building machine learning systems.
- Understand the **technical notions** of learning and adaption and how they relate to one another.
- Detect and avoid problems of **over- and underfitting**.

EXPLAIN SIMPLY

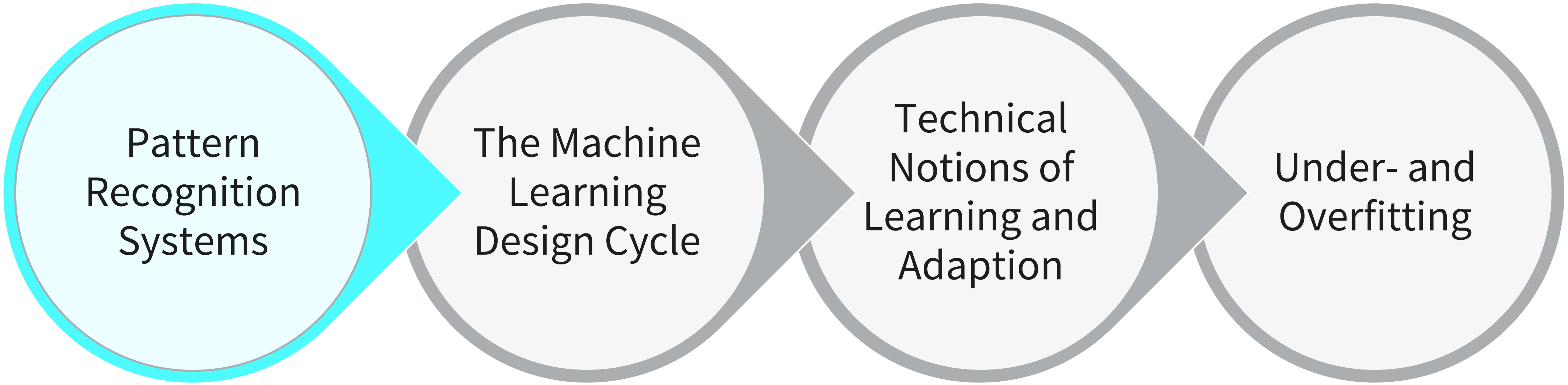
1. What is the difference between pattern recognition and machine learning?
2. What are the six phases of the CRISP-DM design cycle?
3. Why do we split the available data into training, validation, and testing data?

UNIT CONTENT

Img. 1: Introduction to Machine Learning



Source of image 1: Christian Müller-Kett, 2021.


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

- Pattern recognition

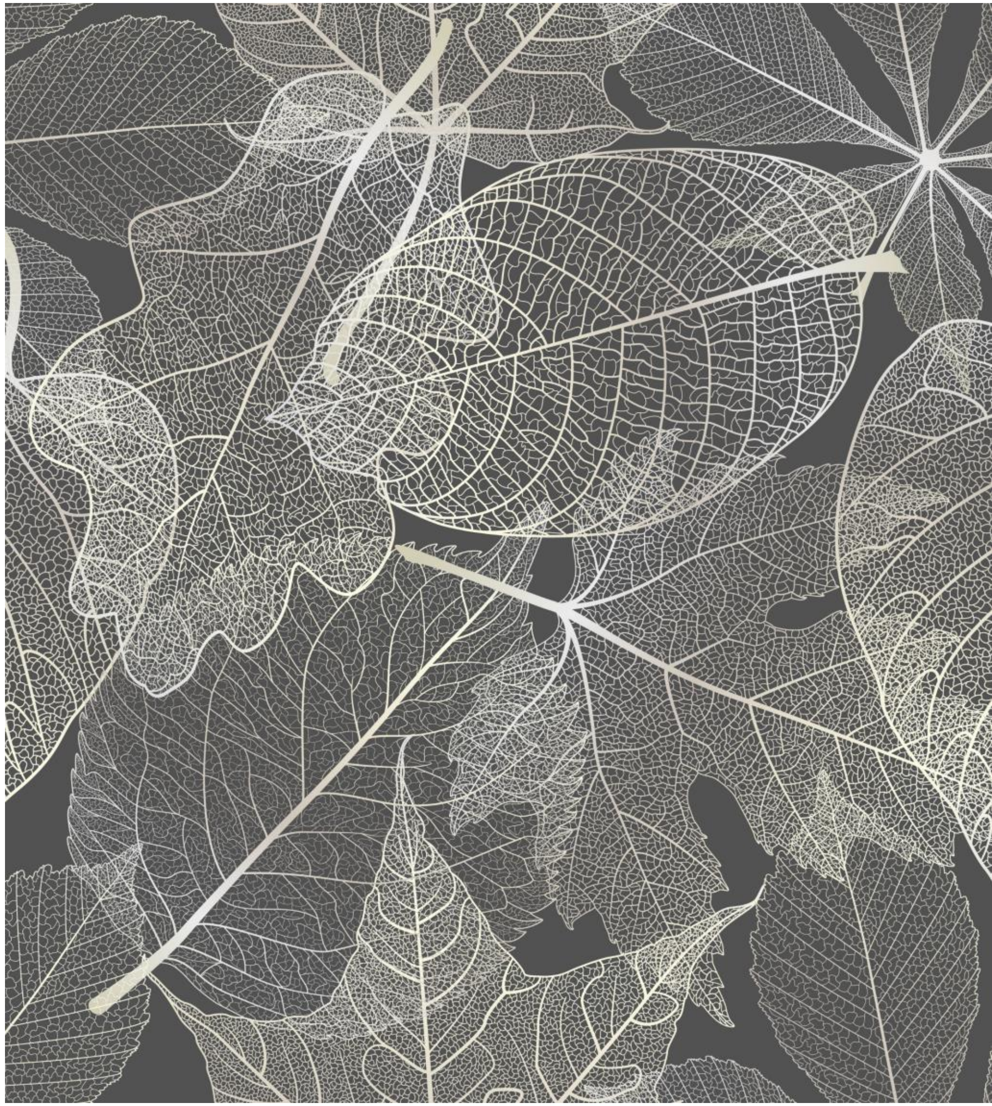
originated from the **engineering field**

- Machine learning

originated from **computer science**

- Both aim to **detect patterns** hidden in data

Img. 2: Pattern



Source of the text: Bishop, 2006, Webb, 2022
Source of image 2: Microsoft Archive

- **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

Tab. 1: Supervised and unsupervised machine learning

	Feature 1	Feature 2	Label
Sampe 1	$x_{1,1}$	$x_{2,1}$	y_1
Sample 2	$x_{1,2}$	$x_{2,2}$	y_2
Sample 3	$x_{1,3}$	$x_{2,3}$	y_3

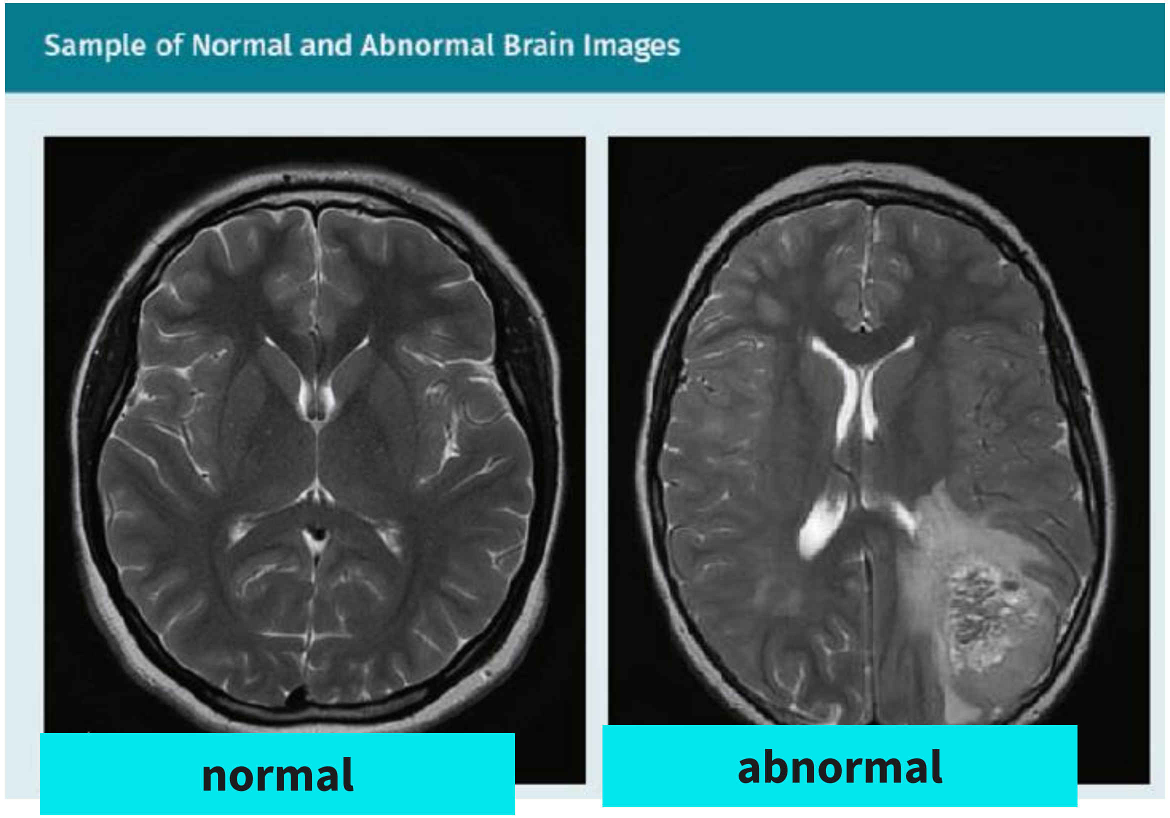
Source of table 1: Christian Müller-Kett, 2021

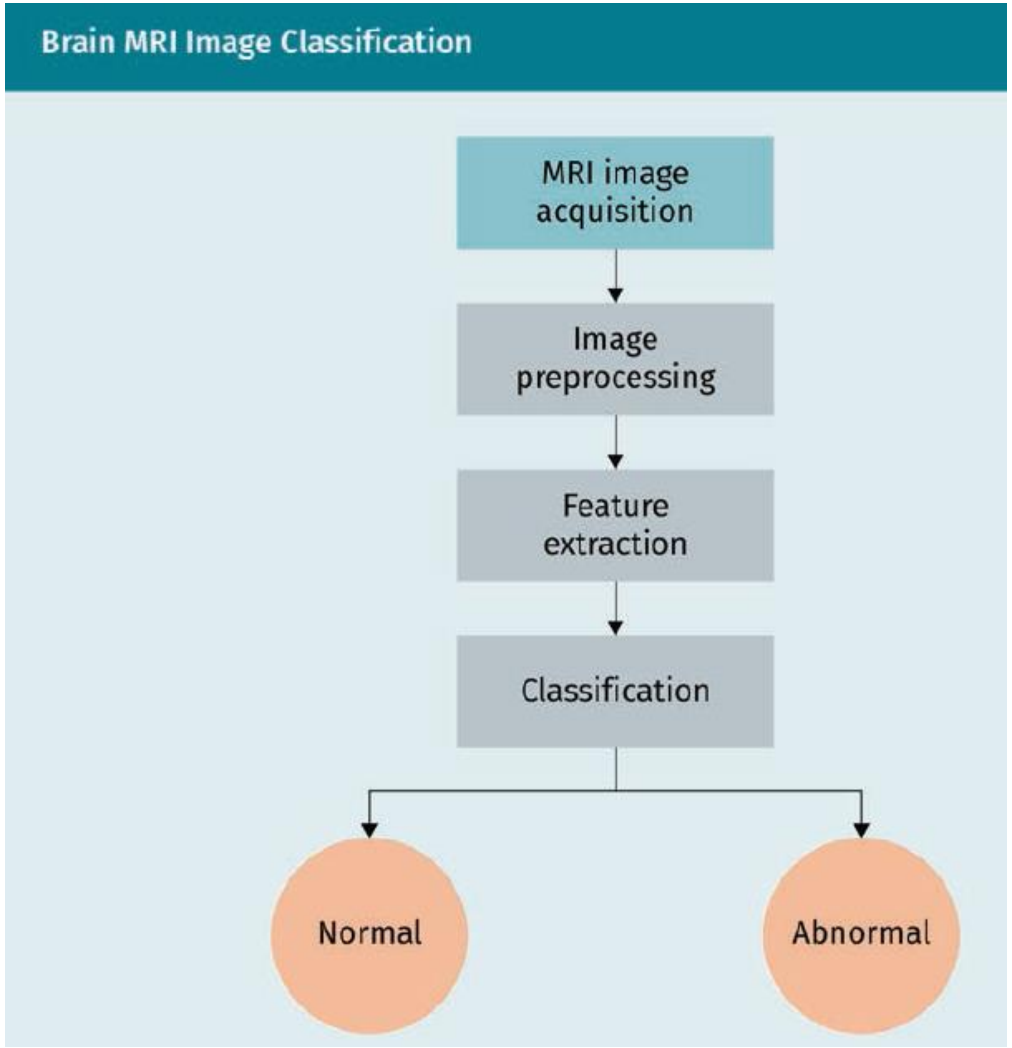
- Type of machine learning
  - Supervised learning: features and the labels are provided
  - Unsupervised learning: only features

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	...	...	...	...

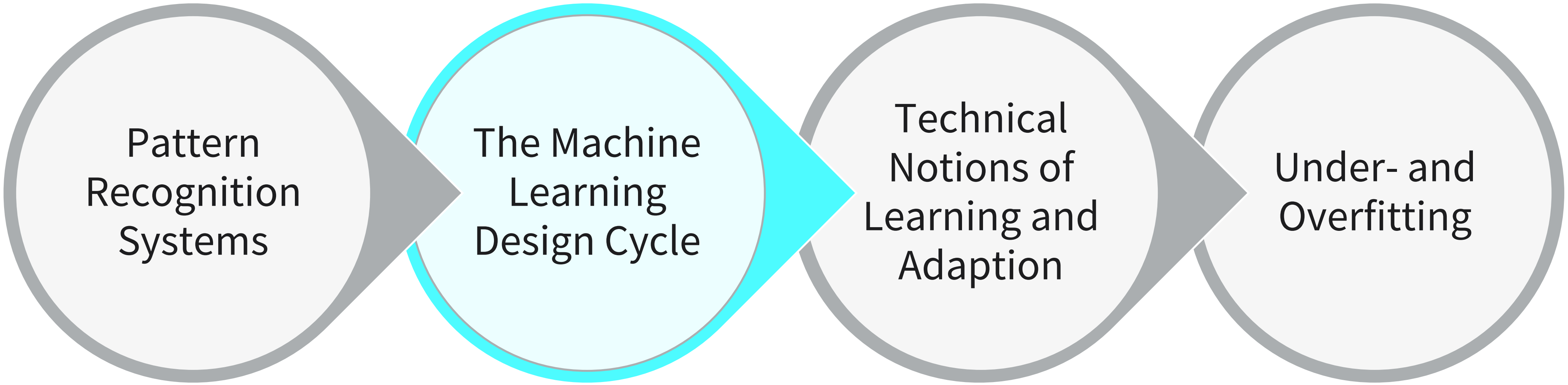
fruit	length	width	weight
fruit 1	165	38	172
fruit 2	218	39	230
fruit 3	76	80	145
fruit 4	145	35	150
fruit 5	90	88	160
...			
fruit n	...	...	...

PATTERN RECOGNITION IN THE REAL WORLD



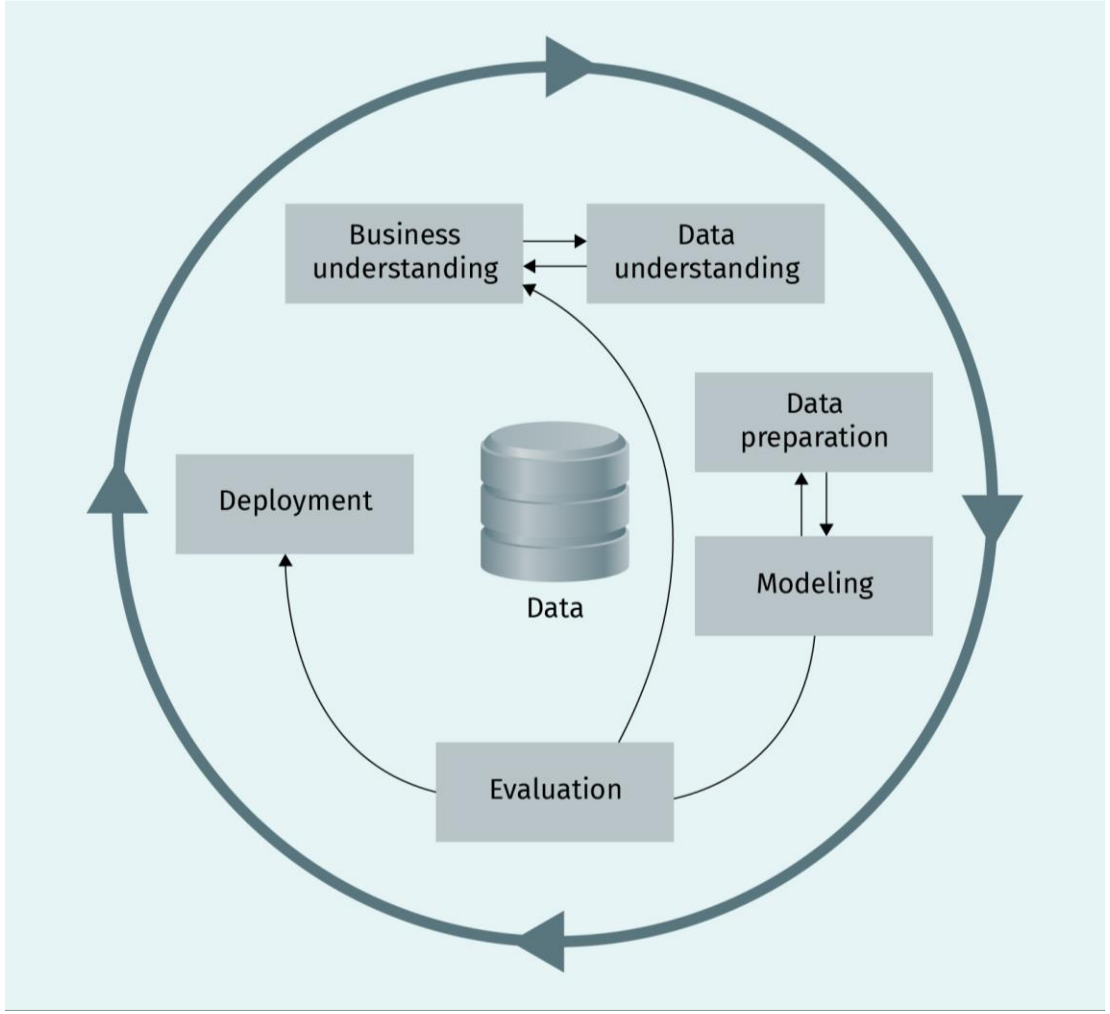


Source of image: Course book

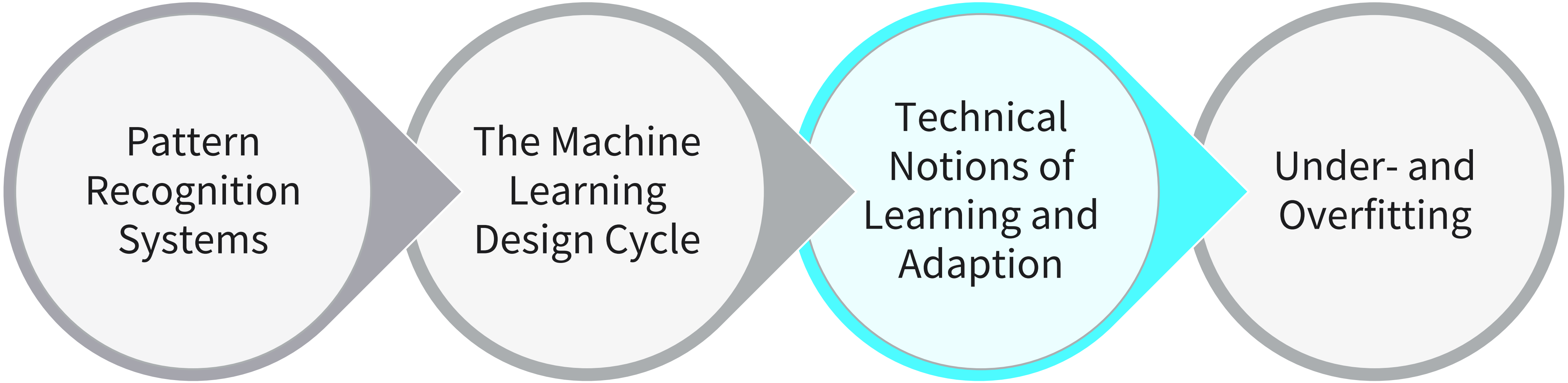

Pattern Recognition Systems
The Machine Learning Design Cycle
Technical Notions of Learning and Adaption
Under- and Overfitting

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

Img. 4: Cross-Industry Standard Process for Data Mining (CRISP-DM)



Source of image 4: Jensen, 2012 based on Wirth & Hipp, 2000.


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

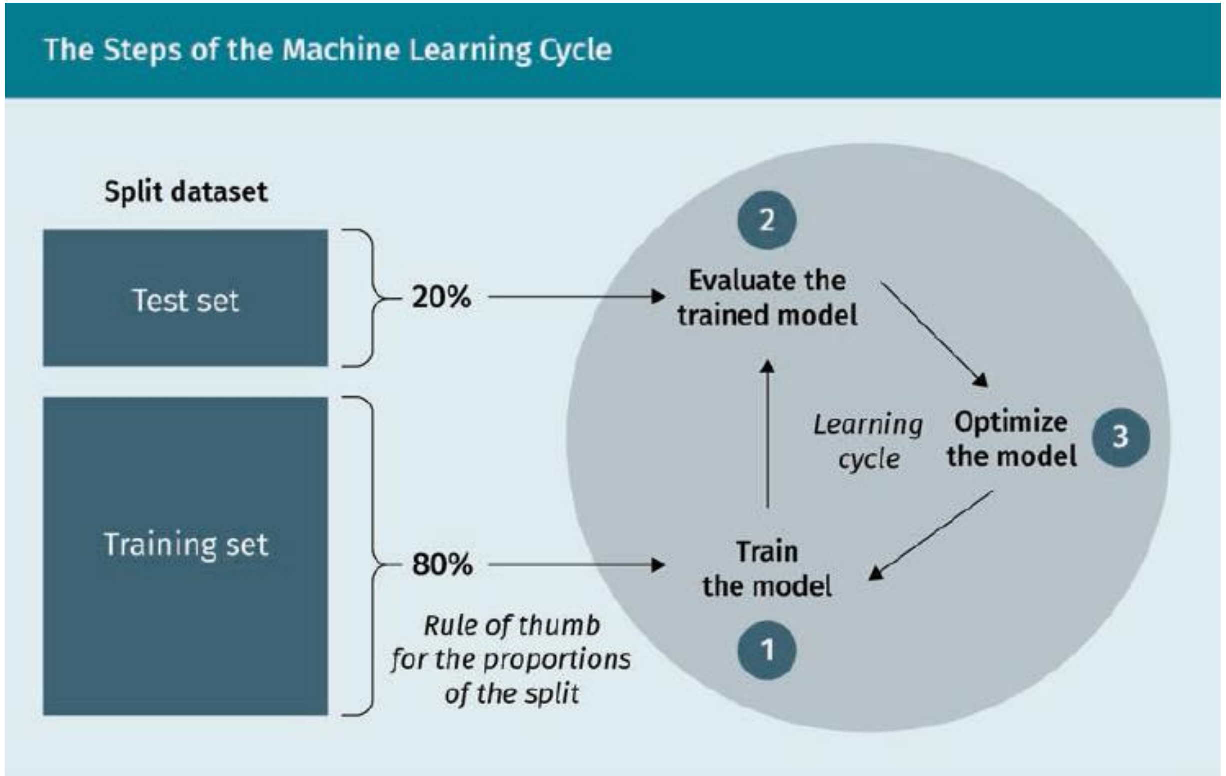
Definition of Machine Learning

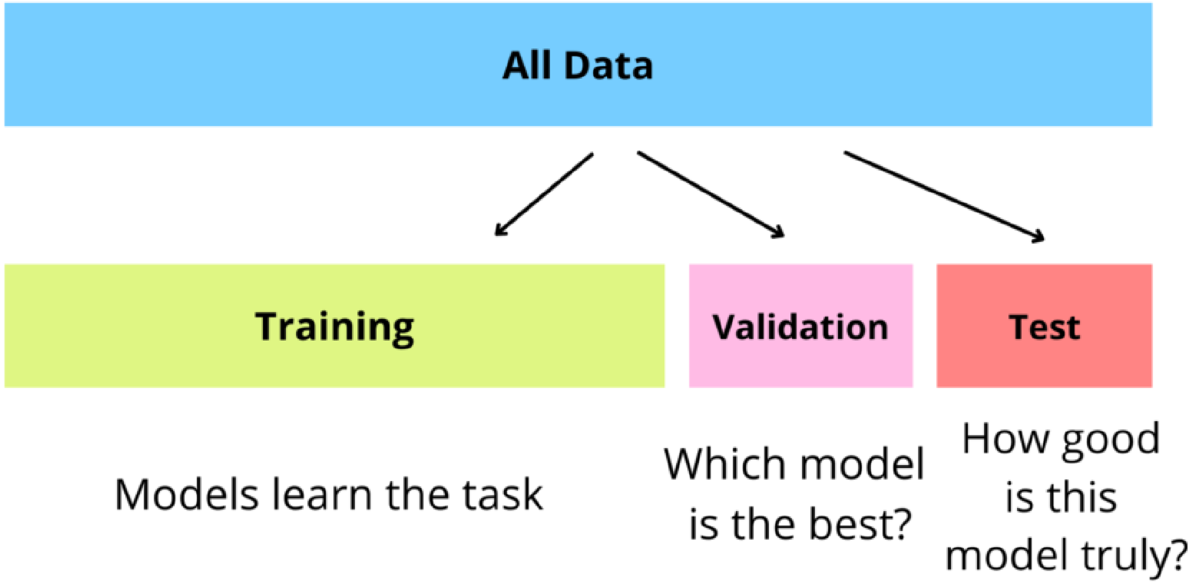
- Machine learning is a field of computer science that focuses on teaching computers to learn what humans do naturally: learn from data and past experience.
- "A computer program is said to learn from experience E with respect to some class of tasks **T** and performance measure **P**, if its performance 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

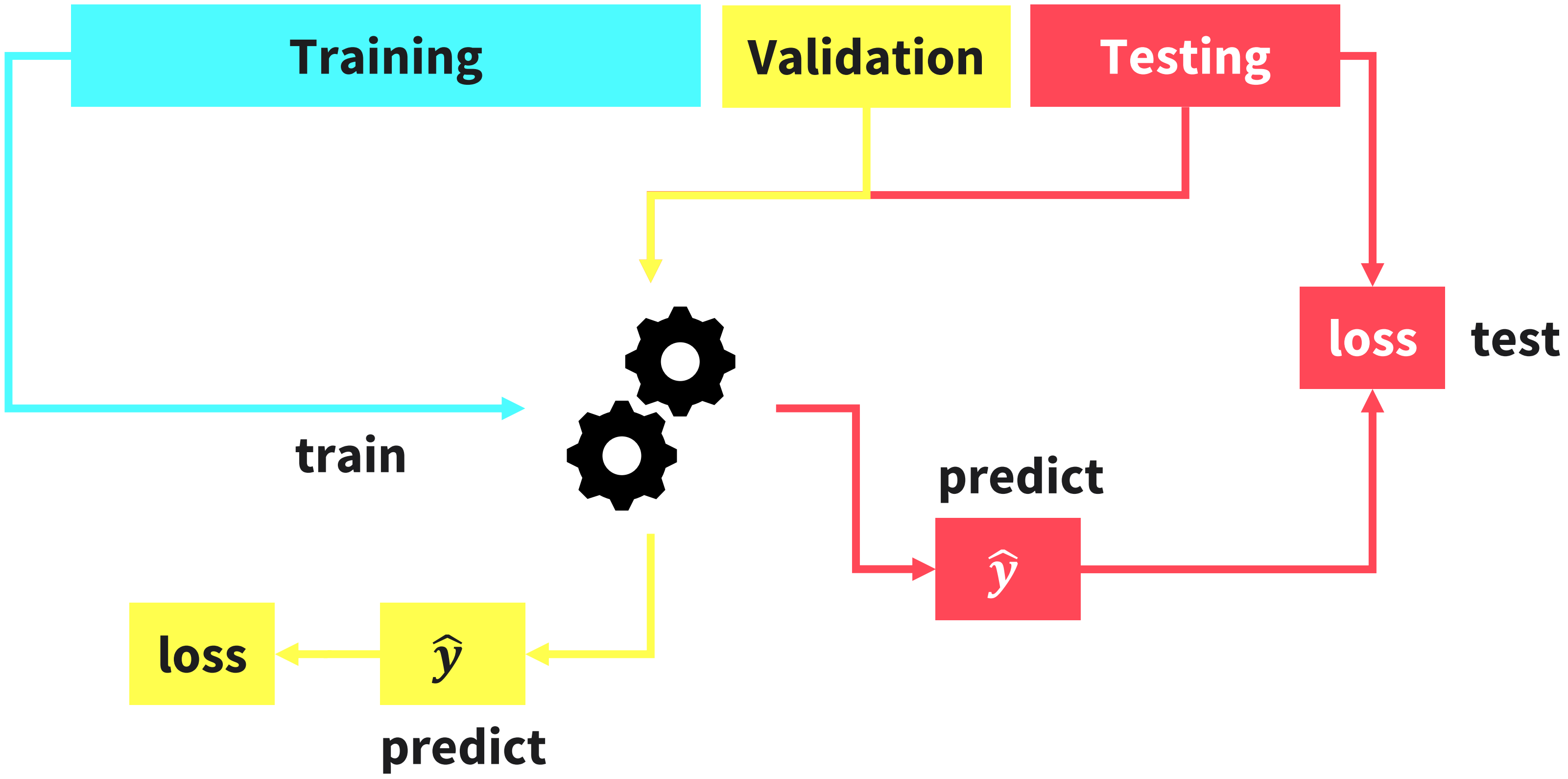




Source of image: Course book, https://galaxyinferno.com/what-is-validation-data-used-for-machine-learning-basics/

TRAINING, VALIDATION, TESTING

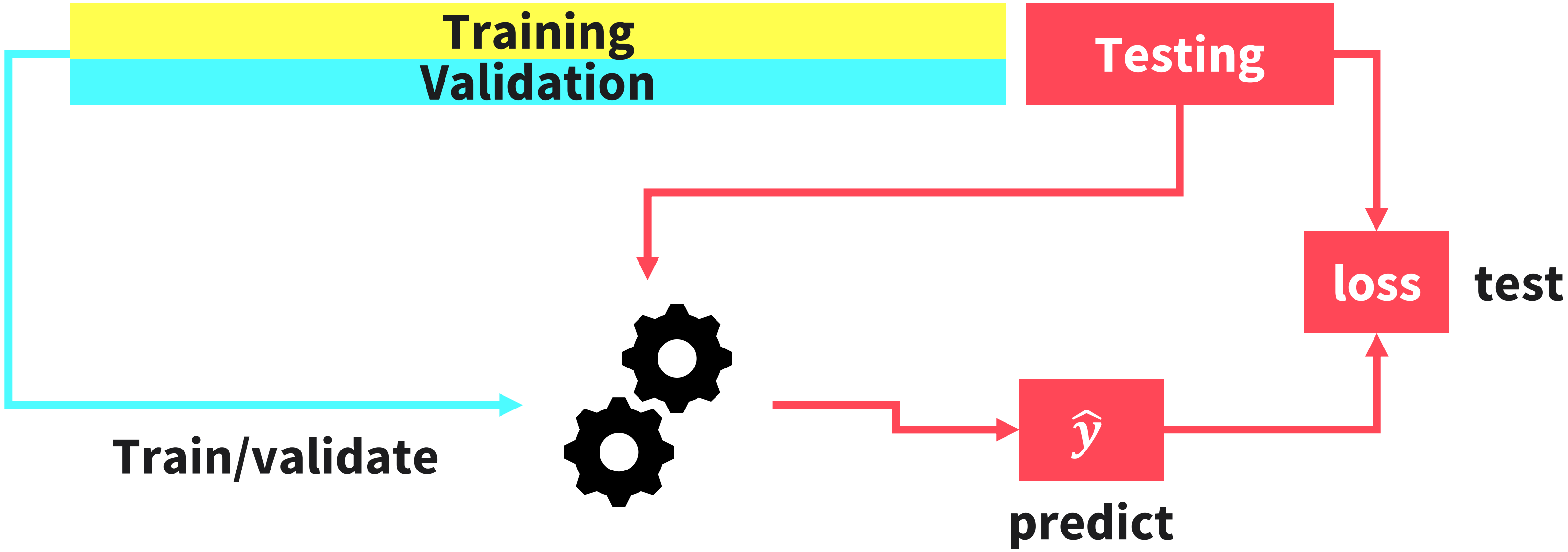
Img. 5: Training, Testing, Validation



Source of image 5: Christian Müller-Kett, 2019

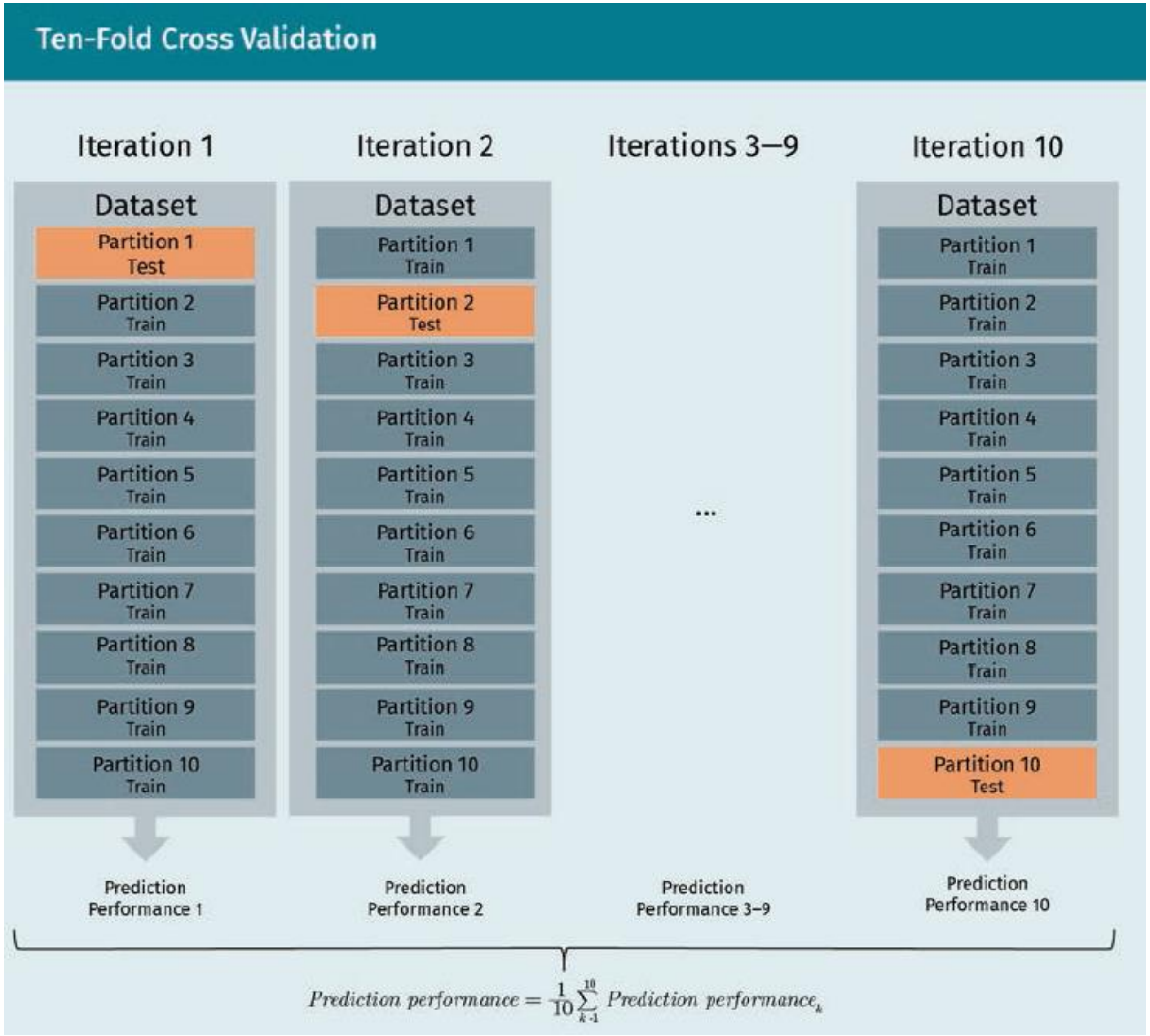
CROSS-VALIDATION

Img. 6: Cross-validation



Source of image 6: Christian Müller-Kett, 2019

- 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.

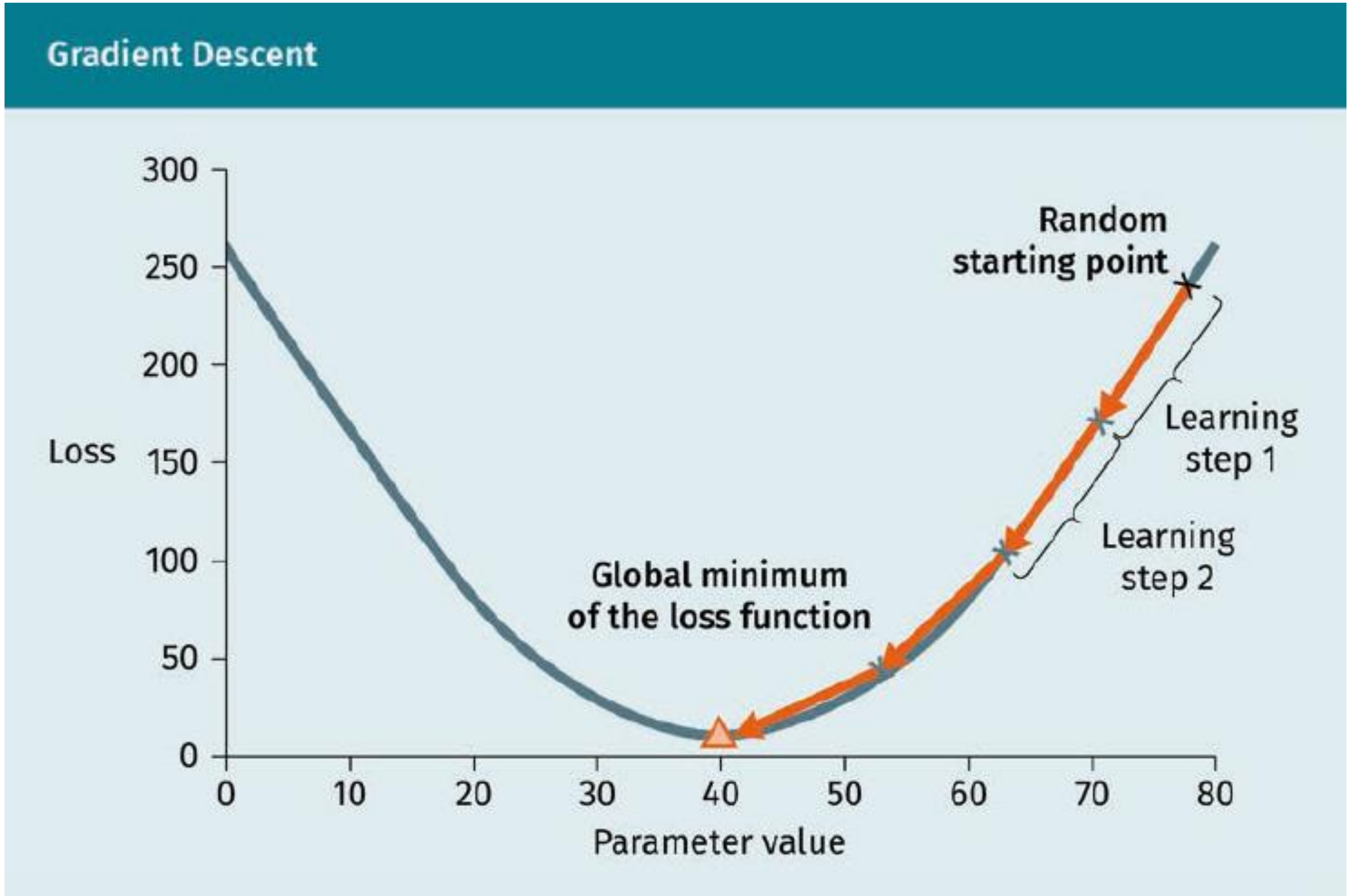
- Loss function or objective or target function:
  - Computes how much the predictions differ from the true values of the labels.
  - The model's accuracy improves based on the minimization of this loss function
  - Used to assess the quality of the model
- The optimizer
  - Iterates over the model's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

- **Set the parameter** in question to an **arbitrary value**
- Calculate the slope of the **loss function** by taking the **first derivative**
- **Multiply** this slope with the **learning rate**
- **Add the result** to the parameter in question
- **Stop when**
  - the **difference** between values is **very low** (e.g. < 0.001) from one iteration to another
  - the **derivative** of the loss function is very **close to zero**
  - a **maximum number of iterations** has been reached (e.g. 1000)

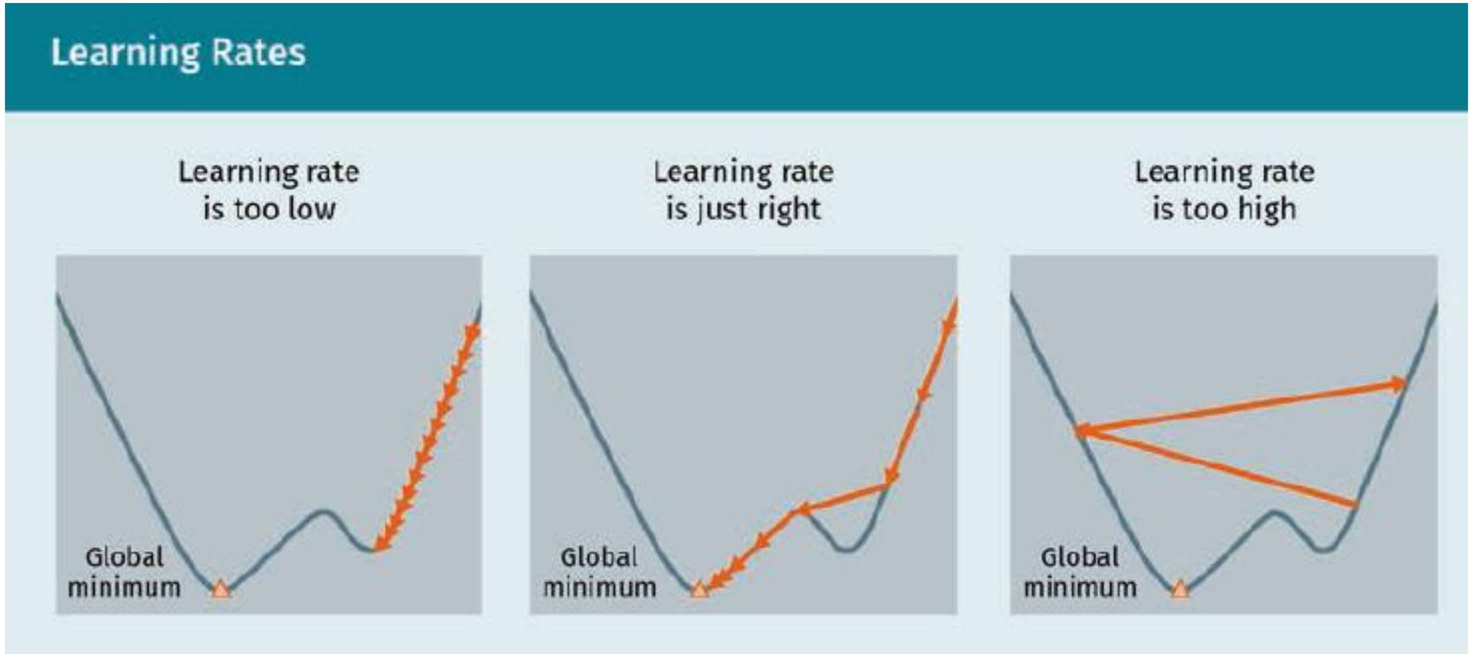
Source of the text: Last name of the author, yyyy.

- 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

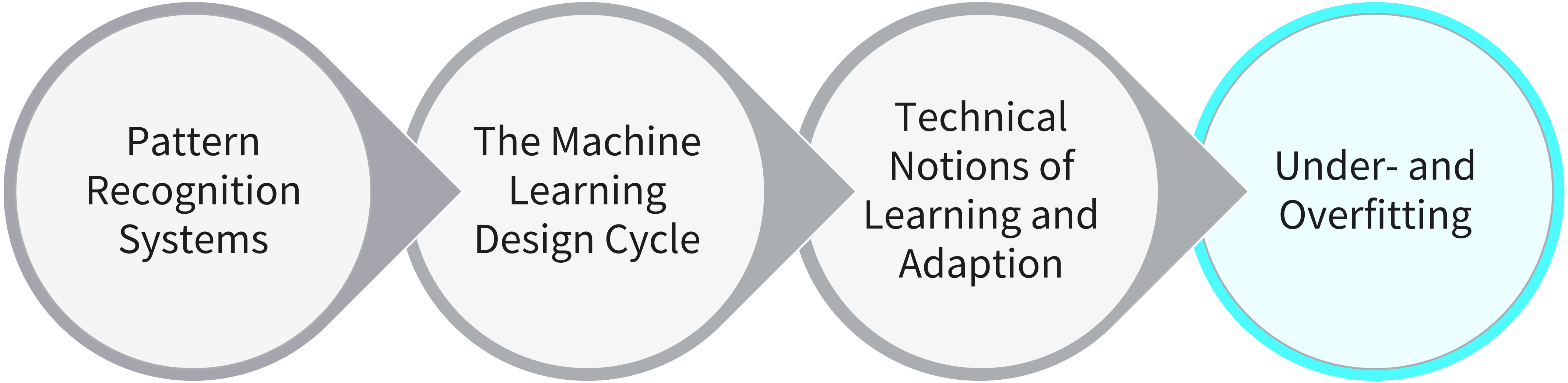
GRADIENT DESCENT



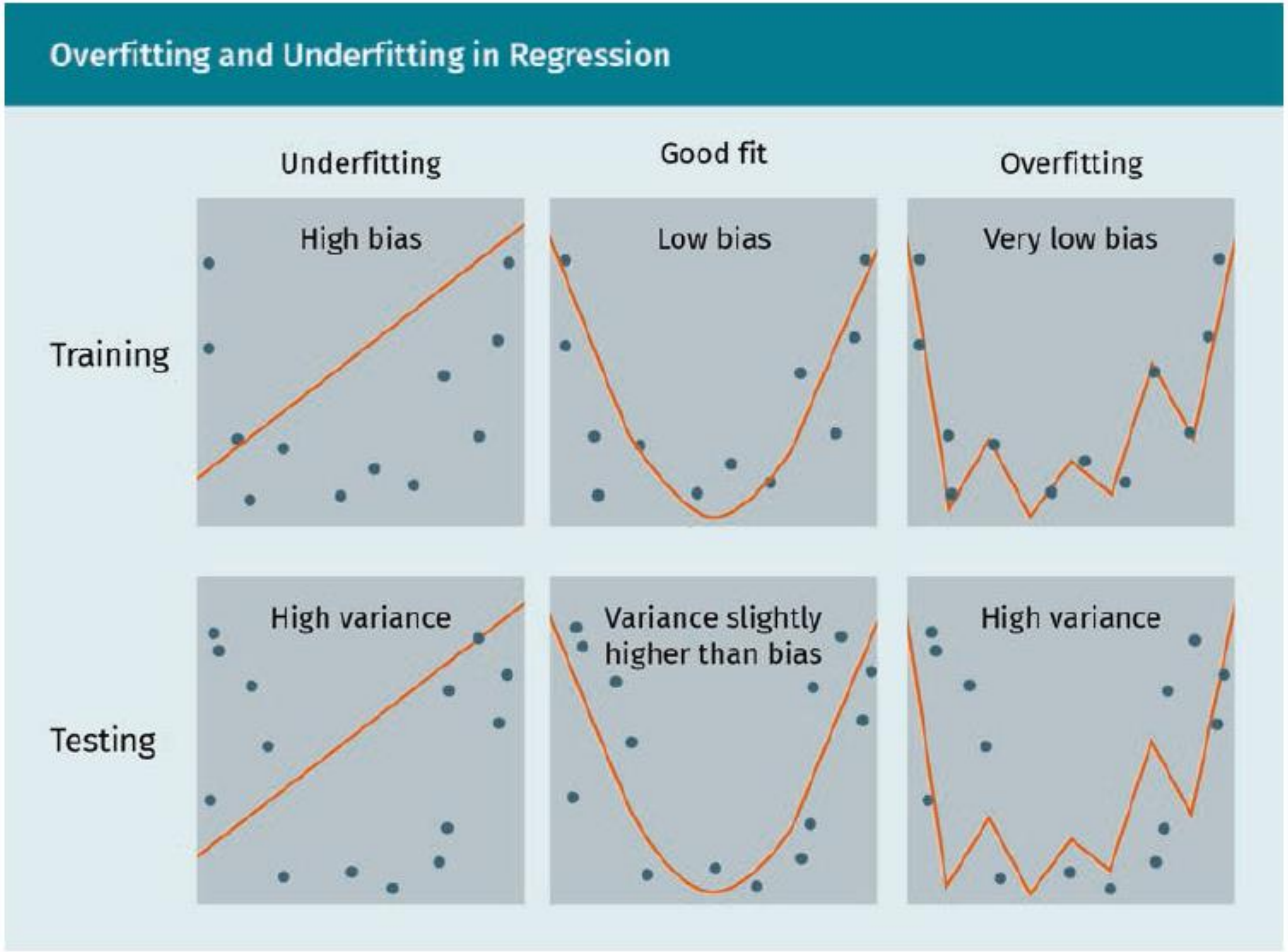
- Gradient descent is influenced by the chosen learning rate



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

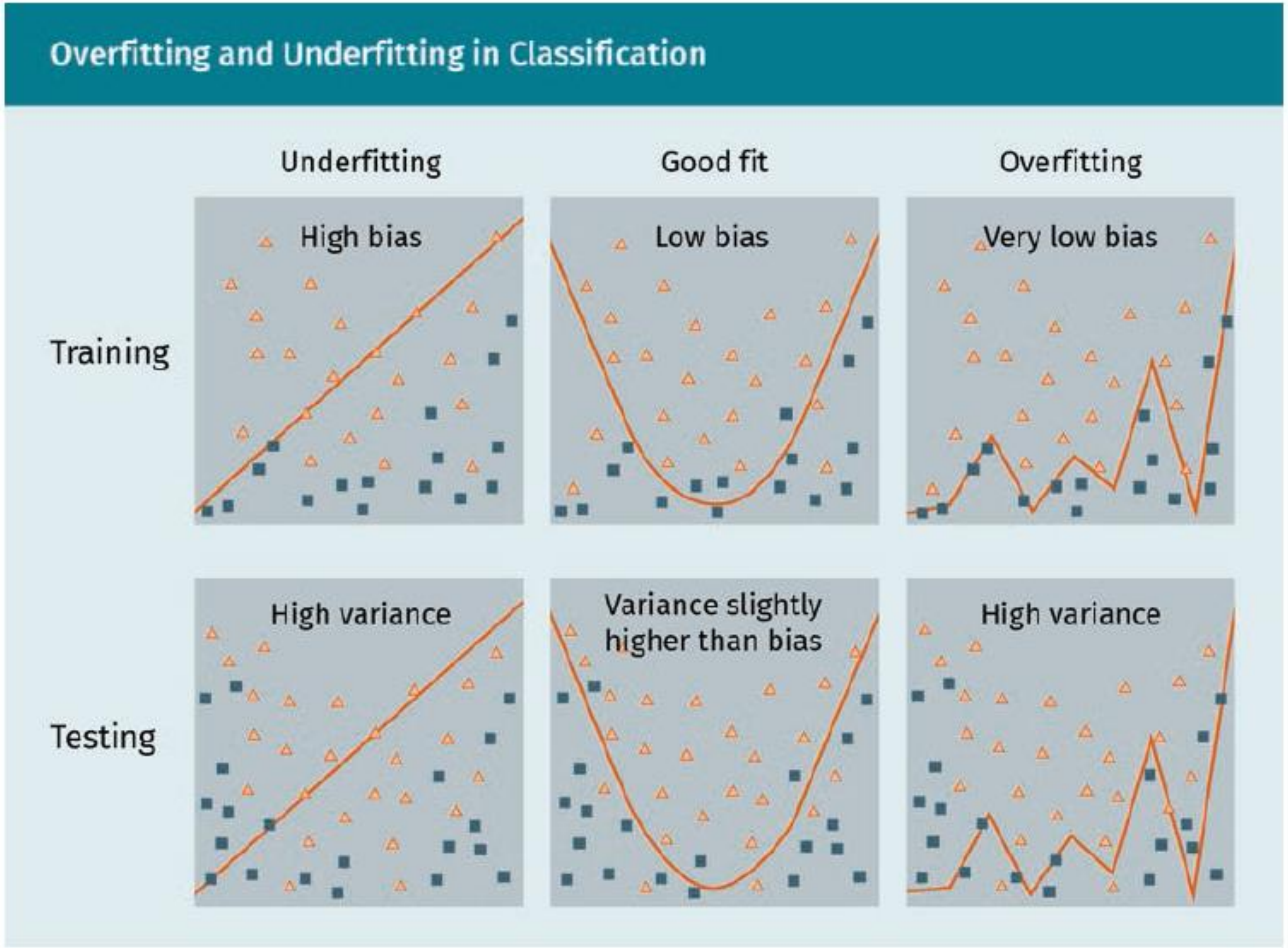

Pattern Recognition Systems
The Machine Learning Design Cycle
Technical Notions of Learning and Adaption
Under- and Overfitting

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

REVIEW 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 another.
- Detect and avoid problems of **over- and underfitting**.

SESSION 1

TRANSFER TASK

A start-up that sells **sustainable products in smaller stores** has been very successful in recent years. As a result, more stores are to be opened worldwide.

As a Data Scientist, you and your team are tasked with training a **machine learning model predicting product demand** one week ahead. Eventually, this model is supposed to be connected to the **company's ordering system**, giving well-informed advice about how many products should be ordered per store.

Create a rough **project plan** and briefly describe the **work items for each of the project's phases**.

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

Please present your results.
The results will be discussed in plenary.

1. Which one of the following is not one of the three partitions into which a machine learning dataset is often divided?
   a) Modeling set.
   b) Training set.
   c) Testing set.
   d) Validation set.

2. What does CRISP-DM stand for?
   a) cross-industry standard process for data modelling
   b) core-industry standard process for data mining
   c) cross-industry standard procedure for data mining
   d) cross-industry standard process for data mining

3. What is it called when a machine learning model memorizes the dataset and loses its ability to generalize?
   a) underfitting
   b) refitting
   c) overfitting
   d) unfitting

LIST OF SOURCES

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

Hashan, A. M. (2021). MRI based brain tumor images, version 1 (Normal (1).jpg) [Photograph]. https://www.kaggle.com/mhantor/mri-based-brain-tumor-images

Jensen, K. (2012). CRISP-DM process diagram (CRISP-DM_process_diagram.png) [Illustration]. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png

Webb, A. (2002). Statistical pattern recognition. John Wiley & Sons. https://doi.org/10.1002/0470854774

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (pp. 29—39). Springer.