

1. MACHINE LEARNING REMINDER

Stephan Robert-Nicoud
HEIG-VD/HES-SO

Credit: Andres Perez-Urbe



Objectives

- ❑ Remind what is Machine Learning and what are the types of problems we can deal with it
- ❑ Remind what are the main steps of a Machine Learning approach
- ❑ Setup the tools for processing data and building Machine Learning models

References

□ Introduction à la science des données (ISD, 2021)

1. Outils pour la modélisation data-driven
2. Introduction à la science des données
3. Apprentissage automatique (Machine Learning)
4. Bibliothèques pour le calcul scientifique
5. Analyse exploratoire des données
6. Apprentissage supervisé
7. Evaluation des modèles
10. Données et caractéristiques

<http://iict-space.heig-vd.ch/ape/teaching/>



Machine learning concept

- ❑ What is Machine Learning ?
 - ❑ learning from examples
 - ❑ software 2.0
- ❑ What are the three main learning paradigms ?
 - ❑ supervised, unsupervised, reinforcement learning
- ❑ What are the two kinds of problems we try to solve with a ML algorithm using a supervised learning approach ?
 - ❑ classification & regression

Machine learning methodology (1)

- What is the first step we have to do before start building a model ?
 - Exploratory Data Analysis: filtering outliers, missing data, box plot analyses
- And then ?
 - pre-processing (e.g., normalization)
 - feature engineering

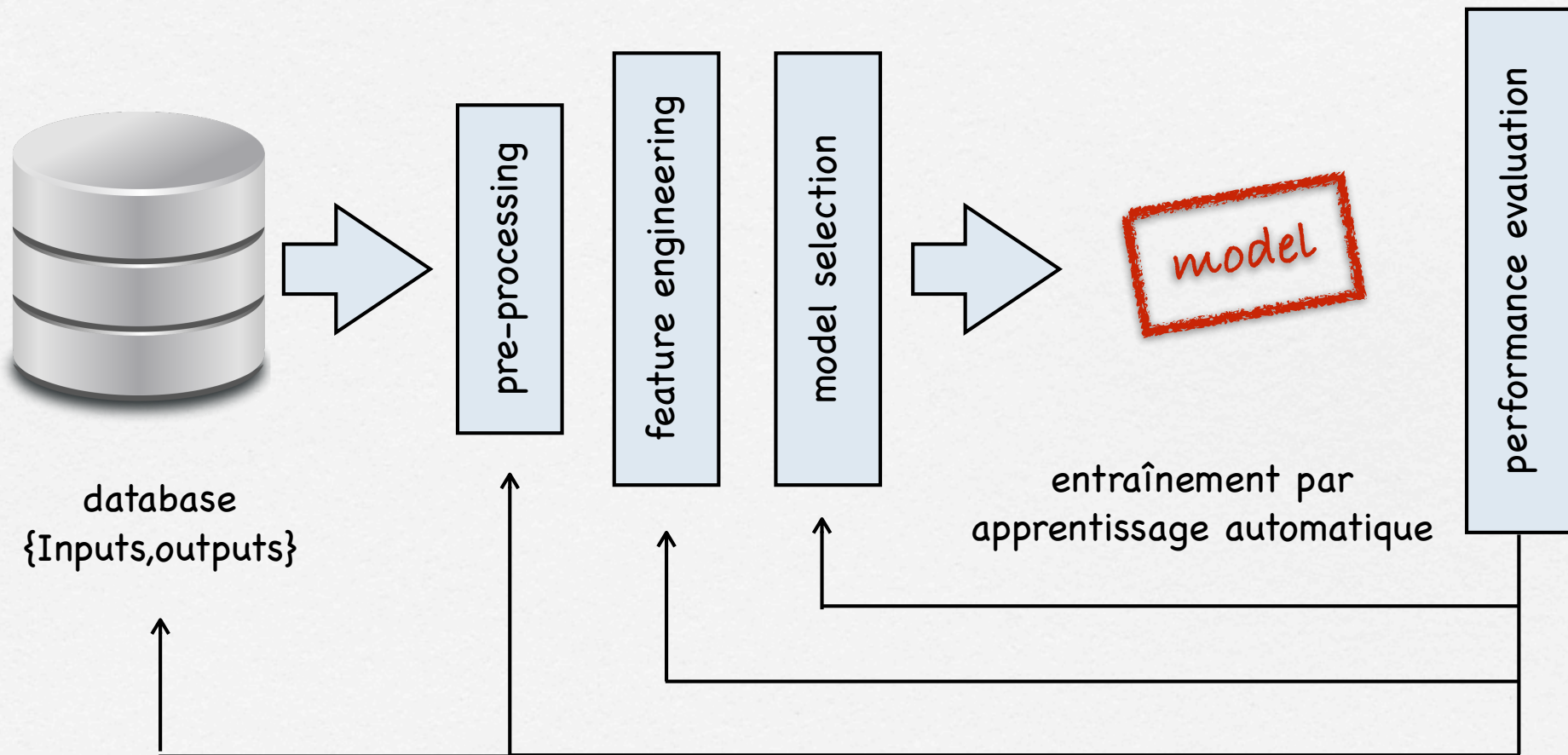
Machine learning methodology (2)

- ❑ What is feature engineering ?
 - ❑ transforming raw data to extract knowledge
- ❑ Please, give examples of features of data
 - ❑ sound -> FFT ; text -> keyword frequency; image -> color histogram; heart beat (time-series) -> heart frequency

Machine learning methodology (3)

- ❑ What data is needed for training a model using a supervised learning approach ?
 - ❑ input data and labels (desired outputs: classes or continuous values to predict)
- ❑ What are the steps in a supervised learning approach ?
 - ❑ data collection -> pre-processing -> feature extraction -> model selection -> performance evaluation

Towards Artificial Neural Networks



Machine learning methodology (4)

- ❑ Consider a supervised learning classification task. What is the main capability that a ML model should exhibit ?
Hint: It is related to model selection
- ❑ **generalization**
- ❑ each ML algorithm is characterized by a series of parameters (e.g., number of neighbors in KNN; a learning rate in linear regression & LVQ, etc.
- ❑ Model selection refers to identifying the right parameters of a model (we also speak of hyper-parameter tuning)

Machine learning methodology (5)

- What is the major risk we have when using a ML model with respect to its generalization capability ?
 - overfitting or "learning by heart" -> bad generalization

Machine learning methodology (6)

□ How do we perform model selection ?

Generate **train** - **validation** - **test** datasets (independent sets)

loop /* model exploration or hyper-parameter tuning */

for different parameters train the model (using the **train** dataset) and evaluate its generalization (using the **validation** dataset) /* cross-validation */

end loop

final performance evaluation (using the **test** set)

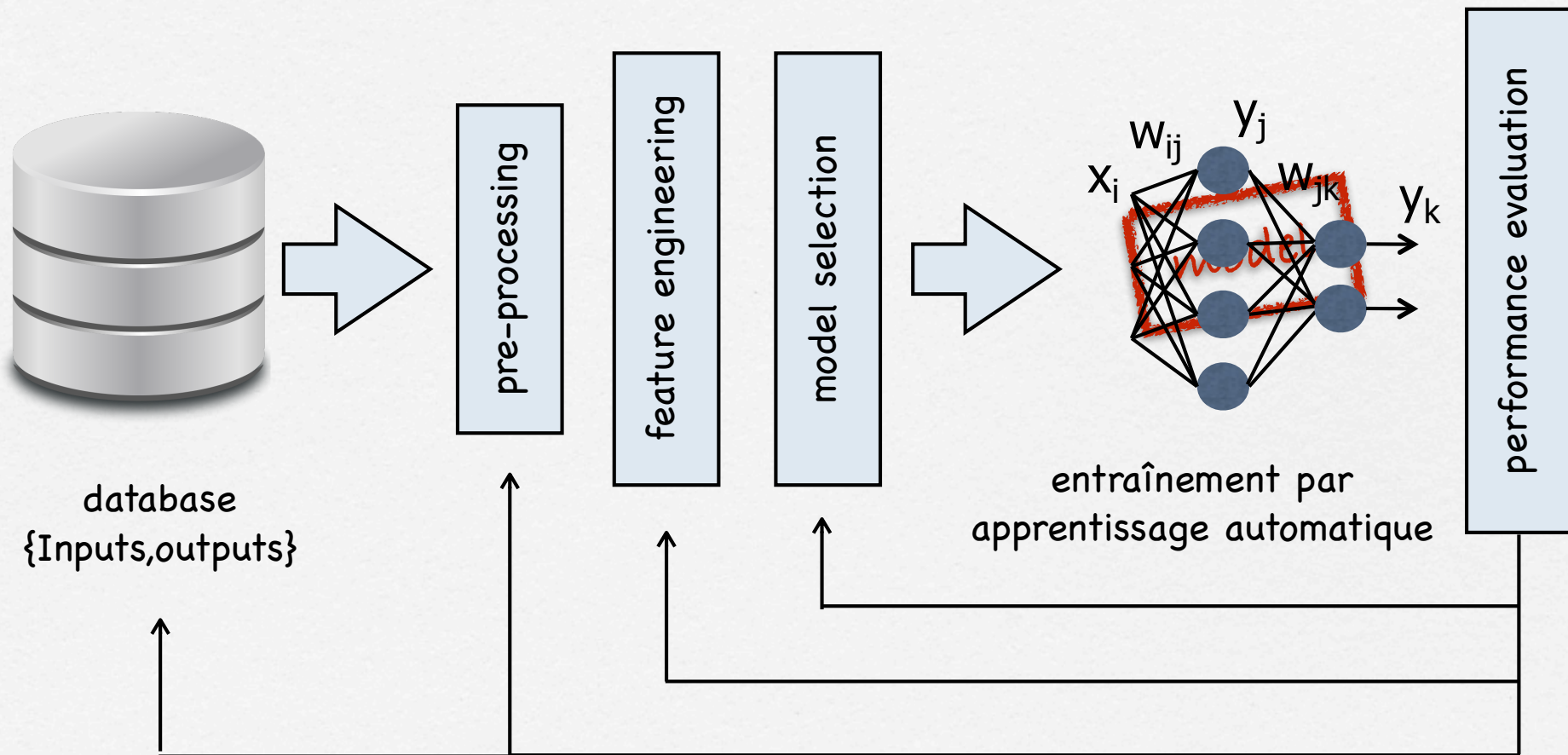
Machine learning methodology (7)

- ❑ What are the performance measures we use for supervised classification ?
 - ❑ accuracy
 - ❑ false positives; false negatives
 - ❑ positive predictive value (precision)
 - ❑ true positive rate (recall)
 - ❑ F-score
 - ❑ confusion matrix

Machine learning methodology (8)

- Why accuracy is not enough ?
 - Consider an unbalanced dataset

Towards Artificial Neural Networks



Course contents

- ☐ The Perceptron
- ☐ Multi-layer Perceptrons and Backpropagation
- ☐ Application: mice sleep phases
- ☐ Convolutional Neural Networks (Deep Learning)
- ☐ Application: object recognition app
- ☐ Survey of state-of-the-art applications