Artificial Intelligence
Chapter : Machine Learning

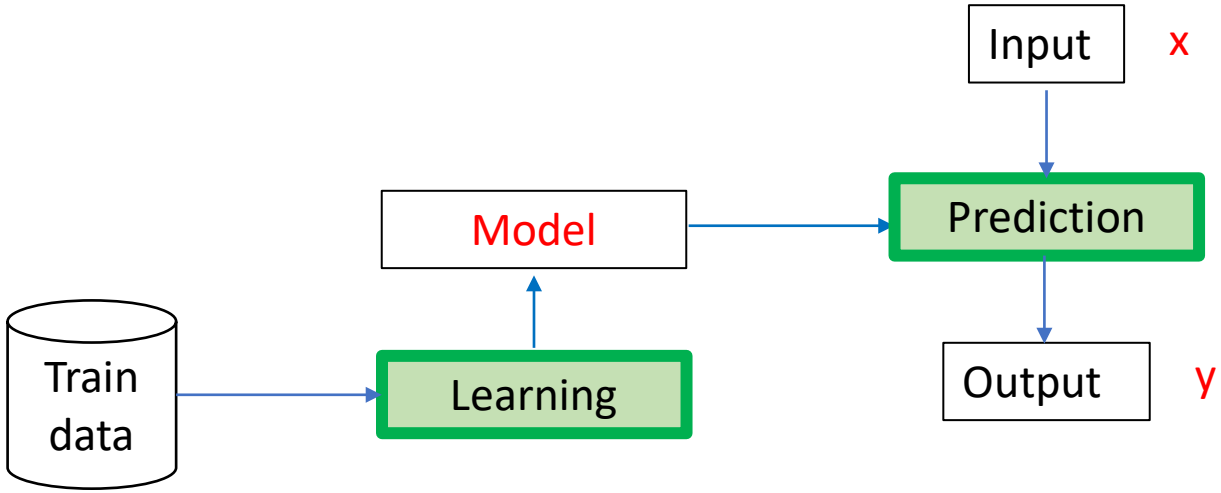
Marouane Ben Haj Ayech

Outline

- Definition of ML
- Prediction process
- Learning process
- Clustering
- Classification
- Regression
- Methodology
- Machine Learning in practice

Definition

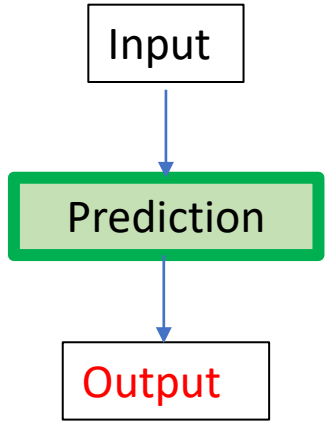
- Machine Learning (ML) is the collection of techniques that allow to learn a model using training data. This model is used later to predict an output value given an input value.



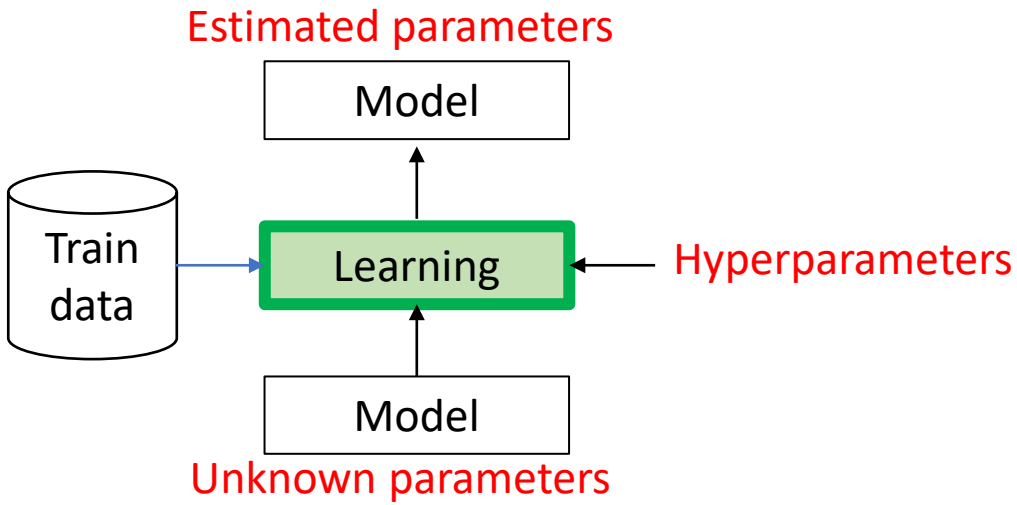
Prediction process

- Depending on the output nature, there are 3 main prediction tasks in ML :

Prediction task	Description	Output Nature	Examples
Clustering	Grouping data points into clusters based on similarity or patterns, often used for unsupervised learning.	Unlabeled classes or clusters	- Customer segmentation based on purchase behavior.
Classification	Assigning data points to predefined categories or classes based on their features, typically used for supervised learning.	Discrete categories or labels	- Email spam classification (spam or not spam). - Image classification (cat, dog, car, etc.).
Regression	Predicting a continuous numeric value or quantity based on input features, typically used for predicting numerical outcomes.	Continuous numeric values	- House price prediction based on features like size, location, and age. - Temperature value prediction based on historical data.



Learning process



- Model : is the core component of a machine learning algorithm that captures patterns from data during the training.
- Parameters : are the internal variables of a machine learning model that are learned from the training data.
- Hyperparameters : are external configuration settings that are not learned from the data but are set prior to the training process. Tuning hyperparameters is an important part of optimizing a machine learning model's performance.
- Training dataset : Data used to train the model parameters

Learning Type	Dataset Type	Prediction Tasks	Learning models
Unsupervised	Unlabeled	Clustering	K-Means Hierarchical Clustering
Supervised	Labeled	Classification	K-Nearest Neighbors (KNN) Naïve Bayes Decision Tree Logistic Regression
		Regression	Linear Regression Polynomial Regression

Clustering

Clustering problem

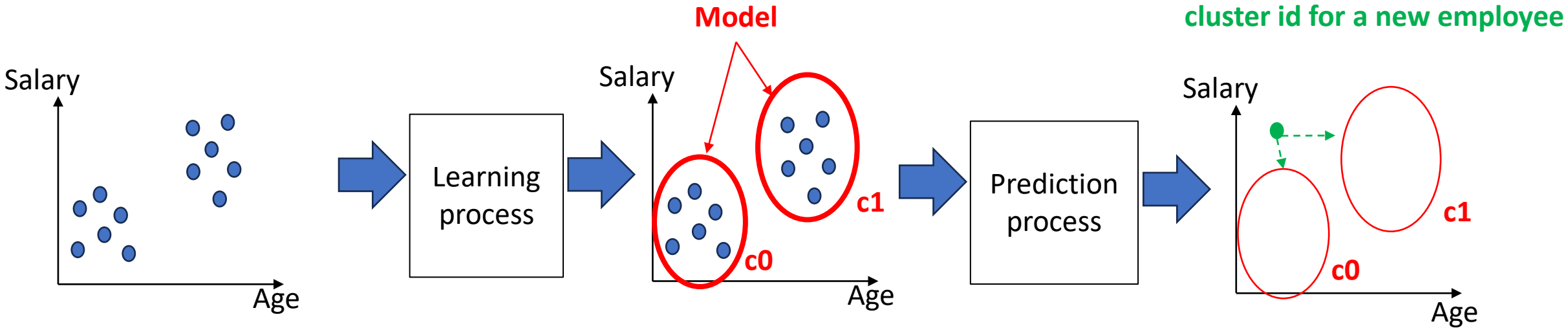
x=employee=(Age, Salary) → y=cluseter id ∈ {0,1}

input output

Unlabeled training dataset

Model

Prediction of the closest cluster id for a new employee

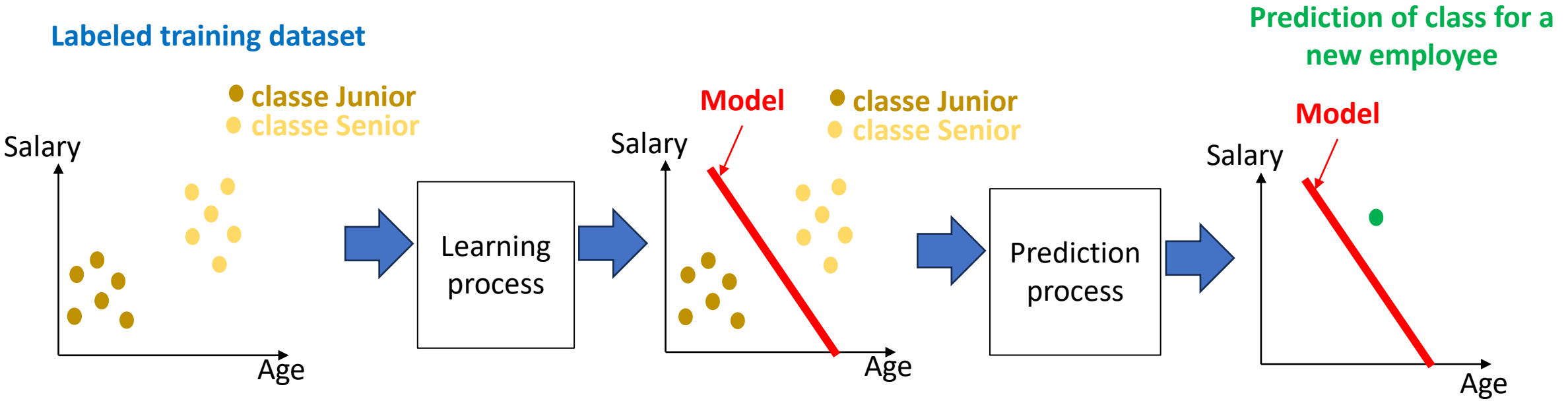


Classification

Classification problem

x=employee=(Age, Salary) → y=class label ∈ $\{0 = Junior, Senior\}$

input output



Regression

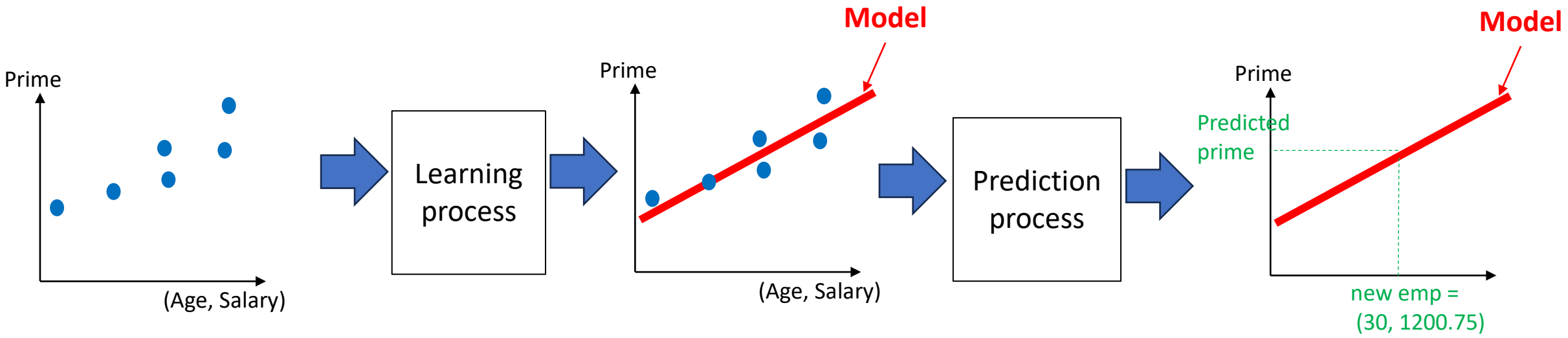
Regression problem

x=employee=(Age, Salary) → y=prime value $\in \mathbb{R}$

input output

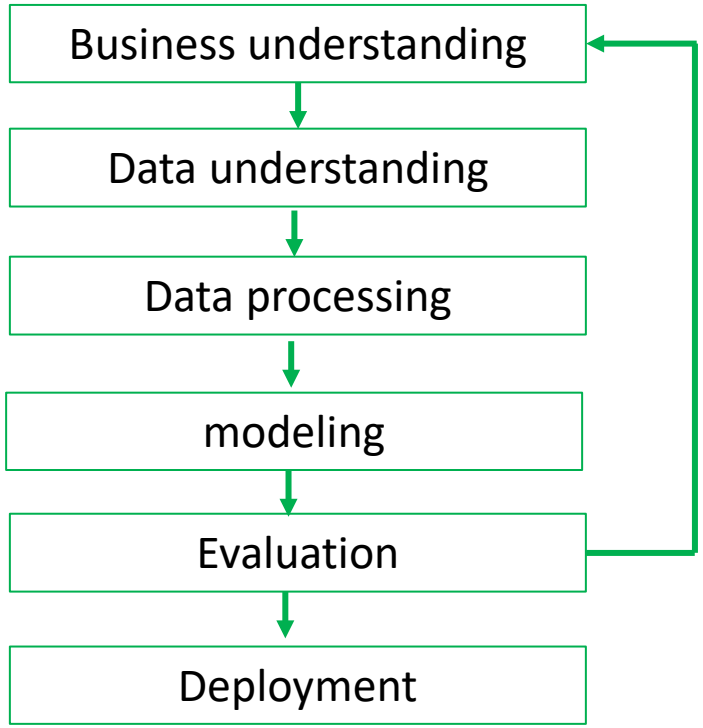
Labeled training dataset

Prediction of prime value for a new employee



Methodology

- The most adopted methodology when dealing with prediction problems is CRISP-DM



Machine Learning in practice

- Setup required tools required by working environment :
  - Python interpreter
  - Jupyter editor (integrated within visual Studio IDE):
    - Python extension
    - Jupyter extension
  - Git batch + Github account
- Tools can can be :
  - installed in local machine
  - preinstalled in google Colab service.
- Implementations are performed as Python notebooks (ipynb files)

Machine Learning in practice

- Setup working directory :
  - In visual studio, create and open a working directory called ML
  - In CMD Terminal :
    - Create a virtual environment (VE) :

      $ python –m venv venv_ml
    - Activate VE

      .\venv_ml\Scripts\activate
  - Put requirements.txt file in ML directory and run command line :
    - $ pip install –r requirements.txt

Machine Learning in practice

- Using working directory (in each session) :
  - In visual studio, open ML directory
  - Create a sub-directory called session suffixed by session number , ie. Session1, session2, …
  - Create a notebook
  - Activate virtual environment venv_ml