

TD1

Exercise 1: Basics

1. A programmer wants to write a program that uses the frequency of words such as "science," "public," "access," "university," "government," "finance," "education," "budget," "justice," and "law" to determine whether an article is about science policy or not. he started by annotating a thousand articles based on their subject. What kind of machine learning problem does she need to solve?
2. A person has 10000 newspaper articles that he wants to categorize by their topic. Should he use a supervised or unsupervised algorithm?
3. A person wants to examine it spam emails to determine if there are subtypes of spam. What type of machine learning algorithm should they use?

Exercise 2: Simple Linear Regression

Note: Please consider 4 digits after the decimal point.

The table shows the respective weights of fathers and their eldest sons:

Father	65	63	67	64	68	62	70	66	68	67	69	71
Son	68	66	68	65	69	66	68	65	71	67	68	70

- 1- Estimate the values of the parameters $\widehat{\beta}_0$ and $\widehat{\beta}_1$ for the weight of sons based on the weight of fathers.
- 2- Estimate the weight of a son that his father weighs 80 kg.
- 3- Compute the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the generated model.

Exercise 3: Simple Linear Regression with Gradient descent .

- 1- Initialize the parameters $\widehat{\beta}_0$ and $\widehat{\beta}_1$ with the values of the previous exercise.
- 2- Using a learning rate $\alpha = 0.1$, compute the new parameters $\widehat{\beta}_0$ et $\widehat{\beta}_1$ in the first iteration.
- 3- Compute the Mean Absolute Error (MAE). What do you observe?