

## I Linear regression

During this practice, we are going to consider a dataset containing grades and chances of admission. It is available on the Madoc platform (`admissiondata.csv`).

- 1) Load the data, and, using the drop function of pandas library, remove the serial number column. Verify that there are no missing values.
- 2) Using the heatmap function of seaborn library, plot the correlation table of all the features. What can we observe?
- 3) We now consider that 'Chances of Admit ' is the target. Produce a scatter plot of all the features (individually) as a function the target.
- 4) Split your data into a training set and a testing set, and then learn a linear model.
- 5) Using the metrics of regression (see <https://scikit-learn.org/stable/modules/classes.html#regression-metrics>), evaluate your model. What are the important features? Is there any link with the correlation table?

## II Going further ...

The linear regression model is quite model, but already gives satisfying results, for this data. Now consider a new dataset, boston houses.

```
from sklearn.datasets import load_boston
dataset = load_boston()
print(dataset.DESCR)
```

For this dataset, process to same learning workflow, and compare the results with another regression method : Gradientboostedregression : <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html> This method has a number of parameters, that you may change in order to maximize the quality of the metrics.