

TP2: Linear regression and logistic regression

Exercise 1: Unidimensional Linear Regression

We consider a problem of predicting the distance travelled by a car based on its speed.

For this problem, use the package of data “cars”.

- 1- Give a general algorithm of linear regression.
- 2- Implement this algorithm to solve the given task.
- 3- Give the optimal set of parameters.
- 4- Give the empirical loss of your model.
- 5- Plot in a graph the real distance and the estimated one with respect to the speed.
- 6- Comment your results.

Exercise 2: Logistic Regression

We consider a problem of predicting whether a student succeed or not based of his GPA and GRE.

For this problem, use the provided csv file named “binary”.

- 1- Give a general algorithm of logistic regression.
- 2- Implement this algorithm to solve the given task. Here, you should create four cases where the first case, the parameters are initialized by zero then the remaining ones, the parameters are initialized randomly with respect to different probability distributions.
- 3- Give the optimal set of parameters.
- 4- Give the empirical loss of your model.
- 5- Plot in a graph of three dimensions the found classifier.
- 6- Comment your results.

Questions from 3 to 6 should be answered for each case.

Repeat the same exercise using the rank as an extra feature.

Exercise 3: Multidimensional Linear Regression

Your objective is to predict the population density people per square mile based on the following variables:

- Death rate per 1000 residents.
- Doctor availability per 100,000 residents.
- Hospital availability per 100,000 residents.

- Annual per capital income in thousands of dollars.

The dataset is contained in an excel file named pop.

To model this data, you should use linear regression model.