Lab Logistic Regression

Data exploration

1. Download the data from this adress :
   https://github.com/AttilaDSA/IntilaqDSAcademy/blob/master/Deep%20Learning%20Labs/Lab%20Logistic%20Regression/email.csv

2. How many spam email there are in the dataset?

3. Explore the data in order to analyse which column(s) affect the most the spam variable.

Logistic Regression

1. Divide your dataset into two parts: train set and test set.

2. Using only one column, the one that you think the most influent on the spam variable, fit a logistic regression model to the train set (you can use the **LogisticRegression** module from **sklearn.linear_model**). Predict using this model the results (spam or not spam) for the test set. How is the accuracy?

3. Do the same operation but, this time, use all the columns of the dataset. Compare this model's accuracy with the previous one.

4. According to the results of (3), build your own fitted model using only the columns that you choose. What was the criteria you used to choose between the columns? What's the accuracy of your model?