School of Computer Science
Faculty of Science
National University of Engineering

Test 4

Subject: Machine Learning for Data Science **Period: 2019.2**

---

1. (5 pts.) Solve exercise 4 of chapter 4 of [2]: items a, b and c worths 1 point each one, item d worths 2 points.

2. (7 pts.) In a `1_<my_first_name>.ipynb` file:

   (a) (2 pts.) Write your answer of exercise 8 of chapter 4 of [2].

   (b) (3 pts.) Reproduce figures 12-5, 12-6,..., 12-9 of [1].

   (c) (2 pts.) Make some comments or conclusions from the graphics above.

3. (8 pts.) Using the Boston data set:

   ```
   # importing dataset
   from sklearn.datasets import load_boston
   dataset = load_boston()
   ```

   fit classification models in order to predict whether a given suburb has a crime rate above or below the median. In a `2_<my_first_name>.ipynb` file:

   (a) (2 pts.) Explore KNN models using various subsets of the predictors.

   (b) (2 pts.) Explore different Values of K in order to find the optimal one.

   (c) (2 pts.) Make some plots to back up your assertions.

   (d) (2 pts.) Make some comments or conclusions from your findings.

   N.B.: if some random result is reported the seed or state must be set in 17; otherwise, two penalty points.

Bibliography

[1] Grus, J. *Data Science par la par la partique: Fondamentaux avec Python*. Big data. Eyrolles, 2017.

[2] James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics. Springer, 2015.

October 25, 2019