



Electrical and Computer Engineering Department
Machine Learning and Data Science - ENCS5341
Homework #1
Submission deadline: 11.12.2022

The `grades.csv` file contains grades of different assessments (homework, midterm, project, **final**) for a group of students in a specific course. We are interested in studying the relationship between the **final** exam grade (the target variable) and other assessments (input variables).

- 1- Some values are missing (indicated by 0 value). Address all the missing values by using the average of the available values for the corresponding variable.
- 2- Using data science techniques that we discussed in the course, examine which of the input variables would be a good predictor for the final exam.
- 3- Implement the closed form solution of linear regression and use it to learn a linear model to predict the final exam from the variable you selected in part 2.
- 4- Repeat part 3 but now by implementing the gradient descent algorithm.
- 5- Repeat part 3 but now using the linear regression implementation of [scikit-learn](#) python library.

Compare the answers from part 3, 4, and 5. In each case compute the error of the learned models.

You have to submit both the code (using python) and the results of running your code. To better structure your submission, it is recommended to use [Jupyter Notebook](#).