CSC4007 – Advanced Machine Learning
Vien Ngo, EEECS, QUB

Lab 3 and 4: Linear Regression

In this lab, we will use linear regression to fit a linear model to two given dataset. The implementation is with Jupyter Notebook. We will use only Numpy and Matplotlib libraries. This lab will run on week 4 (week 16) and 5 (week 17). If you could not finish them during these lab hours, you can finish at home. Through these labs, you will practice to write code to understand and re-implement algorithms learned in Lecture 03.

STEP 1: Download these files

- Files for linear regression with linearly separable data: l*inRegression.ipynb* (main code) and *dataLinReg2D.txt* (associated data file)
- Files for linear regression with non-linearly separable data: *quadRegression.ipynb* (main code) and *data_quadratic_noisy.txt* (data file)

STEP 2: Open the notebook file: "*linRegression.ipynb*" and learn how to write code in Python (using Numpy only) to implement a linear regression algorithm from scratch. You are given a dataset that is stored in the file "*dataLinReg2D.txt*". This implementation uses linear features only.

Through this task, you will learn how to load data from a file into a matrix, inspect the data, write code to compute optimum parameters for the simplest linear model, and report obtained results.

STEP 3: Open the notebook file: "*quadRegression.ipynb*" and learn how to write code in Python (using Numpy only) to implement a linear regression algorithm from scratch. You are given a dataset that is stored in the file " *data_quadratic_noisy.txt*". This dataset is not linearly separable, therefore we resort to non-linear features.

Through this task you will learn how to organize non-linear features and use linear regression to find an optimum linear model.

- Using different features: This code implemented two different features: linear and quadratic. You can choose one of them to test and observe the different results.
- The implemented quadratic features in the code is $\left[1, x_{1,} x_{2,} x_1^2, x_2^2\right]$ which is not in its complete form. Re-implement it to become $\left[1, x_{1,} x_{2,} x_1 * x_{2,} x_1^2, x_2^2\right]$.

STEP 4: Visualize the best model found in Step 3 through cross-validation: visualize the fitting function, training data points, and testing data points.