Disclaimer:

1. Lab reports deadlines are strict. University late submission policy will be applied.

2. Collusion and plagiarism are absolutely forbidden (University policy will be applied).

3. Report is due 14 days from the date of running this lab

3.1 Objectives

- Implement the Least Mean Squared Error algorithm

- In this experiment, we will use the publicly dataset to verify our algorithm. Download the UCI Breast dataset: http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)

- Estimate the classification method

3.2 Estimation of Classification Methods

- (**10 marks**) Read the dataset into a list and shuffle it with the **random.shuffle** method. Hint: fix the random seed (e.g. **random.seed(17)**) before calling **random.shuffle**

- (**20 marks**) Split the dataset as five parts to do cross-fold validation: Each of 5 subsets was used as test set and the remaining data was used for training. Five subsets were used for testing rotationally to evaluate the classification accuracy.

3.3 Least Mean Squared Error

- (**5 marks**) **Training stage:** Augment the feature vector x with an additional constant dimension,

$$\hat{X} = \begin{bmatrix} X & \mathbf{1}_{N \times 1} \end{bmatrix}$$

- (**10 marks**) Scale linearly the attribute values x_{ij} into $[-1, 1]$ for each dimensional feature as follows:

$$x_{ij} \leftarrow 2 \frac{x_{ij} - \min_i x_{ij} + 10^{-6}}{\max_i x_{ij} - \min_i x_{ij} + 10^{-6}} - 1$$

where a small constant is used to avoid that the number is divided by zero.

- (**10 marks**) Reset the example vector x according its label y

$$x \leftarrow \begin{cases} x, & y \in \omega_1 \\ -x, & y \in \omega_2 \end{cases}$$

- (**15 marks**) Find the weight vector w

$$w \leftarrow (X^T X)^{-1} X^T \mathbf{1}$$

- (**10 marks**) Predict the example x

$$c = \begin{cases} \omega_1, & w^T \begin{bmatrix} x \\ 1 \end{bmatrix} \geq 0 \\ \omega_2, & \text{otherwise} \end{cases}$$

- (**10 marks**) **Test stage** : Estimate the accuracy on the test set, that is the fraction of the correctly classified examples in the total test set.

- (**10 marks**) Repeat 5-cross-fold validation for 5 times to average all accuracy performance.

3.4   Lab Report

- Write a short report which should contain a concise explanation of your implementation, results and observations.

- **Please insert the clipped running image into your report for each step with the mark.**

- Submit the report and the python source code electronically into ICE.

- The report must be written with the latex typesetting language.

- The report in pdf format and python source code of your implementation should be zipped into a single file. The naming of report is as follows:

  e.g. StudentID_LastName_FirstName_LabNumber.zip (123456789_Einstein_Albert_1.zip)

3.5   Hints

Please refer to the lecture slides.

- Latex IDE: texstudio

- Python IDE: pycharm

- Use the python numpy library flexibly.