



## Statistical Pattern Recognition (Fall 2020)

### Homework#1: Linear and Logistic Regression

---

Due date: 8<sup>th</sup> December 2020

In order to do this homework, you have to go through linear and logistic regression theories and concepts.

#### Part A: Linear Regression

---

Dataset: Data-Train, Data-Test

Linear regression tries to model the relationship between two variables by applying a linear equation to a series of data. Your task is to train linear model on the given datasets.

Implement linear regression using **closed form solution** and the Gradient Descent algorithm (**Batch** or **Stochastic**, only one) and test your implementation on the given Test Dataset.

- Use the Data-Train to train your model, and test on the Data-Test.
- Plot of the datasets and regression lines.
- Plot cost function for enough iteration for linear regression in Gradient Descent (Batch or stochastic, only one).
- Report the learned parameters ( $\theta_0, \theta_1, \dots, \theta_n$ ), and also the value of MSE error on the train and test data.
- Do not forget that you could normalize the data.

The **UCI Machine Learning Repository** is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms. The archive was created as an ftp archive in 1987 by David Aha and fellow graduate students at UC Irvine. Since that time, it has been widely used by students, educators, and researchers all over the world as a primary source of machine learning data sets.

## Part B: Logistic Regression (Binary Classification)

---

Dataset: Iris <https://archive.ics.uci.edu/ml/datasets/Iris>

- The Iris dataset consists of 4 features and 3 classes. Use only the first and second features (remove the third and fourth columns) and also delete the instances of the 'Iris-versicolor' class to reduce the data to 2 classes with 2 features.
- Consider the first 80% of the data in each class for train and the rest 20% for test
- Report the training and testing errors, and the equation of the decision boundary.
- Also, plot the decision boundary along with the samples of the two classes with different colors all in one plot.

### Notes:

- **Pay extra attention to the due date. It will not extend.**
- **Be advised that submissions after the deadline would not grade.**
- **Your implementation should be functional.**
- **Prepare your full report in PDF format and include the figures and results.**
- **Do not use sklearn or any similar library for regression and logistic regression and write your own code.**
- **The allowed programming languages are any language and feel free.**
- **Feel free for using sklearn in python for load iris or split train and test dataset.**
- **Submit your assignment using a zipped file with the name of "StdNum\_FirstName\_LastName.zip" to [compuscien@gmail.com](mailto:compuscien@gmail.com) with SPR-Fall 2020-HW#1 subject.**