Assignment 2

Homework assignments will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited. Electronic submission on Canvas is mandatory.

1. **Gradient descent algorithms** (50 points) Please download the breast cancer data set from UCI Machine Learning repository. Implement your Logistic regression classifier with ML estimator using Stochastic gradient descent and Mini-Batch gradient descent algorithms. **Do not use any package/tool**. Use cross-validation for evaluation and report the recall, precision, and accuracy on malignant class prediction (class label malignant is positive).

2. **Naive Bayes** (20 points) From Project Gutenberg, we downloaded two files: The Adventures of Sherlock Holmes by Arthur Conan Doyle (pg1661.txt) and The Complete Works of Jane Austen(pg31100.txt). Please develop a multinomial Naive Bayes Classifier that will learn to classify the authors from a snippet of text into: Conan Doyle or Jane Austen.

3. **Linear classification** (10 points) Please prove that 1) multinomial naive Bayes classifier essentially translate to a linear classifier. 2) Logistic regression is a linear classifier.

4. **Linear Discriminant Analysis** (20 points) Please download the Iris data set from the UCI Machine Learning repository and implement Linear Discriminant Analysis for each pair of the classes and report your results. Note that there are three (3) class labels in this data set.