Assignment 1 – Applied Machine Learning

Due Friday February 17th @11:55pm

This assignment must be submitted as a Jupyter Notebook, including texts, figures and codes describing the analysis. The goal of the assignment is to analyze a data set, to ask some underlying question about the data, predict features and select the best model. The dataset must be selected by you from either the UCI (https://archive.ics.uci.edu/ml/datasets.php) or Kaggle repositories (https://www.kaggle.com/datasets). You may also use Google's dataset search engine to find a dataset of your interest (https://datasetsearch.research.google.com/). The analysis should focus on the following steps:

1. Data Description and Analysis
    a. Describe the motivation to select the dataset
    b. State the objective and approach
    c. Describe the data, visualize and analyze the data
2. Prediction
    a. Choose a model
    b. Prepare data for training (test-training split, cross-validation, k-folding)
    c. Maximize performance
        i. Appropriately select the parameters (optimize, visualize)
        ii. Dimension reduction, feature selection, normalization, etc.
    d. Analyze the results (overfitting-underfitting, confusion matrixes, decision boundaries, learning curves, etc.)
3. Model Selection
    a. Apply the previous analysis for different models
    b. Compare results
4. Conclusions
    a. Discussions about parameter, structure choices and prediction results.

Submission Guidelines:

The assignment can be done individually or in groups of two. Submit your assignment as a single zip file through the moodle link provided. Use the following naming format for your zip file:

lastnameStudent-as1.zip

The zip file should contain the Jupyter notebook. Any comments, equations, images or discussion should be part of the notebook.