CS545 Machine Learning

Project Proposal for

Comparison of Classification Algorithms

By

Rohan Jhaveri

(CSUID: 830962238)

&

Ameya Sanjay Kirve

(CSUID: 831692065)

Date: November 03, 2018

Objective:

The purpose of this project is to learn and implement six different Machine Learning Classification algorithms mentioned below to categorize the samples based on certain input attributes.

Description:

As mentioned above, this project aims at applying six different classification techniques to categorize the samples which will be taken from the following datasets available at the UCI Machine Learning Repository:

- Amazon commerce reviews (https://archive.ics.uci.edu/ml/datasets/Amazon+Commerce+reviews+set)
- Letter Recognition (https://archive.ics.uci.edu/ml/datasets/Letter+Recognition)
- Australian sign language signs (https://archive.ics.uci.edu/ml/datasets/Australian+Sign+Language+signs+%28High+Quality%29)
- Burst Header Packet (BHP) flooding attack on Optical Burst Switching (OBS) Network (https://archive.ics.uci.edu/ml/datasets/Burst+Header+Packet+%28BHP%29+flooding+attack+on+Optical+Burst+Switching+%28OBS%29+Network)
- Detect Malicious Executable (https://archive.ics.uci.edu/ml/datasets/Detect+Malacious+Executable%28AntiVirus%29)

Based on different attributes and types of datasets such as number of features, sparseness of the data set etc. the classification will be demonstrated on data sets mentioned above.

The report for this project will be made taking in to consideration the following mentioned key points:

- Data Analysis, cleaning and partitioning
- Theory and Mathematics involved in each of the following classification techniques
  - *K-NN Classification,*
  - *Decision Trees,*
  - *LDA & QDA,*
  - *Linear Logistic Regression Based Classification*
  - *Non-Linear Logistic Regression (with "tanh" & "ReLU")*
- Implementation of each of the above-mentioned algorithms in python programming
- Detail analysis on data being used
- Appropriate partitioning of data sets for classification
- In-depth explanation for the observed behavior of each of the model upon applying them to the data sets
- Results of different experiments (Variation in Hidden Layer Structures, Number of training iterations etc.)
- Plots and discussions on each experiment
- Comparison of all the machine learning classification algorithms mentioned above to determine the best ones that result in maximum prediction accuracy
- Summarizing the results, plots, observations of all the experiments carried out in the course of this project
- Future scope and applications

Timeline:

The timeline below for our task is approximate, taking into account the number of data sets and their size and the time it would take to simulate them and get the results.

Task Name	Start Date	End Date
Project Proposal	10/21/2018	11/03/2018
Data Analysis, Cleaning and Partitioning	11/05/2018	11/10/2018
Study/Analysis/Implementation of Classification algorithms	11/12/2018	11/30/2018
Testing, Plots and Comparison	12/01/2018	12/03/2018
Final Report and Presentation	12/04/2018	12/10/2018