Spring 2019: CSCI 6990 Programming Assignment #3

DUE: Monday, April 29, 2019 (**Softcopy** @4 PM; **Hardcopy**@6:30 PM /in class)

Instructions

❑ All work must be your own (other than the instructor provided codes and hints to be used). You are not to work in teams on this assignment.

❑ Format: Your solution must be typed. Submit as a single compressed file (via Moodle) **containing all the related files in it**. Name it as PA3_<Your_name>.

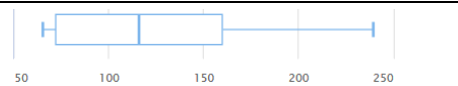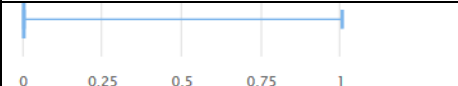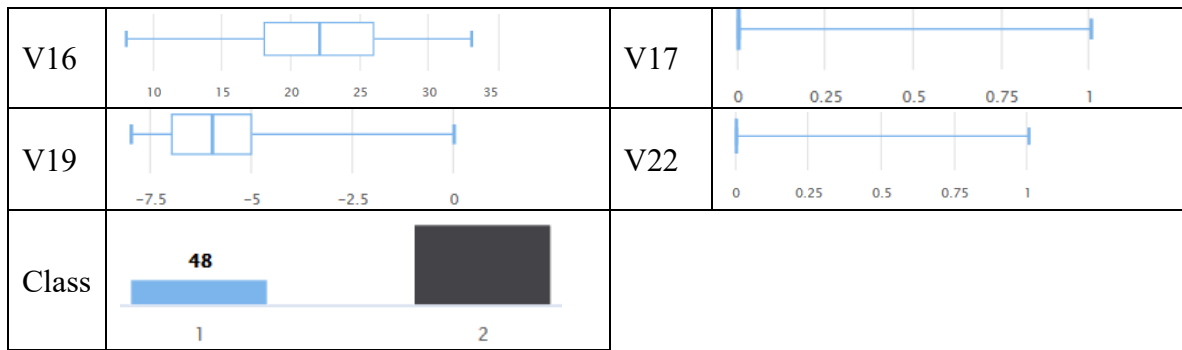❑ The top/cover page of the report should have the title, "Spring 2019: CSCI 6990 Programming Assignment #3". Then your, "Name:_____ and ID:_____

Description

- Here we will be comparing the performances of

    (i)   Linear Regression for classification (LR)
    (ii)   Linear Discriminant Analysis (LDA)
    (iii)   Quadratic Discriminant Analysis (QDA) and

    - approaches for a binary classification problem.

- We will be using 10-fold cross validation (10 FCV) to measure the average error rate.

- We will be using a dataset, called: 'Parkinson's Disease Detection from Voice'. The dataset was generated by Max Little of the University of Oxford, in partnership with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The original study published the feature extraction methods for general voice disorders (for the dataset and the related publications see /Moodle/assignment folder).

- <u>Information about the dataset</u>: This dataset consists of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals. The goal of the processed-data was to train the model(s) to discriminate healthy people from those with PD. The "class" column with value '1' indicates healthy person and '2' indicates a person with PD.

- Description of the dataset: All the input data dimensions (from V1 to V22) are numeric data and some of their properties are given by the following table:

Data	Properties	Data	Properties
V1		V2	
V3		V10	

V16	10 15 20 25 30 35	V17	0 0.25 0.5 0.75 1
V19	−7.5 −5 −2.5 0	V22	0 0.25 0.5 0.75 1
Class	48 1 2		

Reporting

1. In your own language, briefly describe, possibly why we can use voice features to detect Parkinson's disease – you may use published peer-reviewed literature using appropriate citations.   **Mark: 20%**

2. From the relevant literature (with appropriate citations), describe in your own language, how the features were extracted or, generated from the collected human voice.   **Mark: 20%**

3. Your report **must** have the following table: **Mark: 40%**

	Error Rate (%)		
Cross-Validation (CV)	LR	LDA	QDA
CV1			
CV2			
…			
CV10			
Average →			

4. Submit your program code and data (submit data if you modify the given dataset) such a way so that it can be run to verify your result-table. Thus, you will need to describe,' How to run your code', in your report.   **Mark: 10%**

5. Well commented programming code is expected.   **Mark: 10%**

--- X ---