Lab Test

The following Excel spreadsheet has the group allocation for your lab exam. Kindly follow the instructions.

B.Tech:https://docs.google.com/spreadsheets/d/1N0VNgayCkGJlbFQhrmgjoLs0yHjvb7KX50s74z6z508/edit?usp=sharing

M.Sc(DS):https://docs.google.com/spreadsheets/d/13zfvacuVuplKCFXf3rd1mdpyTWDEjoecz3kT5yjDo1o/edit?usp=sharing

1). The marks will be allotted to the individuals working in the group according to her/his understanding of the concept.
2). Each group has to generate a report which details their solutions for questions given in the assignment and bring it on the day of the viva.
3). Each student should know everything about the assignment, and they should know the things conceptually.
4). Any plagiarism case would lead to assignment failure.

1. Construct the training set T = { (x1,y1),(x2,y2),……,(x500,y500)} using the relation Yi = sin(5 π xi) + ϵi where ϵi ~ N(0,0.15). Similarly, construct a testing set of size 50 I,e. Test = { (x'1,y'1),(x'2,y'2),……,(x'50,y'50)}. Use RBF kernel in this question for Kernel Regression.

a) Estimate the regularized Least Squares Regression using Kernel Regression with Gradient Descent.
b) Modify the training set T by randomly picking up 5 data points from the training set T and scaling their yi values by 20. Estimate the Least Squares Estimate using Gradient Descent and state your observations.
c) Modify the optimization problem of the least squares regression model to improve the RMSE results on test data. Further, solve your modified optimization problem using Gradient Descent and calculate the percentage of improvement obtained by your model over the existing least square regression model.

2. Generate the datasets A and B in R2, each consisting of 2000 data points from a normal distribution. The dataset A and B has been drawn from the N (μ1, Σ1) and N(μ2, Σ2). Let us fix the μ1 = [-1,1] and μ2 = [2,2]. Separate the 250 data points from each class as a testing set.

Consider

$$\Sigma 1 = \Sigma 2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

a). Plot the optimal Bayesian decision boundary for this.

b). Train a logistic regression model using the Gradient descent Method and plot the separating decision line.

c). Train the SVM Model Gradient descent Method and plot the separating decision line.

d). Compare the performance of logistic Regression and SVM using Bayesian Decision Boundary.

3). Consider the two-moon dataset. Train Support Vector Machine using the Gradient Descent Method and appropriate Kernel for obtaining the separating surfaces by tuning the parameters. Plot the data points along with separating surfaces. The dataset has been attached in CSV format with this assignment.