Quiz 3
Foundation of Machine Learning (IT 582)
Marks 10 (4 +2+2+2)

Given the training set $\{(x_i, y_i) : x_i \in R^n, y_i \in R \text{ for } i = 1, 2, ..l, \}$, we can estimate the relationship bewteen x and y using

$$y_i = w^T \phi(x_i) + \epsilon_i \qquad (1)$$

where $\phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_1(x) \\ .... \\ ..... \\ \phi_m(x) \end{bmatrix}$ and $\phi_j(x)$ for $j = 1, 2, ..., m$, are m different basis functions. In veiw of this, answer the following questions.

1. Write down the optimization problem for regularized Least Squares regression model and obtain its solution in brief.

2. List the two important significance/advantages of the Least Squares regression model.

3. Write the main drawback of the Least Squares regression model. What are the advantages of the LASSO over the Least Squares model?

4. Why do we need the L_1- norm loss regression models? List their advantages and disadvantages.

Quiz 4
Foundation of Machine Learning (IT 582)
Marks 10 (2+2+2+4)

Multiple Choice Question

Instructions: Each question may have any number of correct choices. Clearly mark (tick) all choices you believe to be correct along with your explaination (Marks will be given only when you choose all correct answers along with explaination only).

1. Four different people are doing bias-variance estimates on regularised linear regression models. They come to you and make the following claims about certain experiments they've done. Which of these claims are definitely incorrect? (Here λ refers to the regularisation parameter as usual.)

   (a) I increased λ and the model started underfitting the data, whilst the variance went down.

   (b) I decreased λ and the model started overfitting the data, whilst the bias went up.

   (c) I decreased λ and the model started overfitting the data, whilst the variance went up.

   (d) I increased λ and the model started underfitting the data, whilst the bias went down.

2. Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to my earlier training set and re-train via gradient descent, initializing the parameters to those of the original model, in which cases will the learnt decision boundary remain exactly the same ?

   (a) When my model is Gaussian bayesian calssifier.

   (b) When my model is Logistic regression .

   (c) When my model is standard SVM.

   (d) When my model is KNN model.

3. Suppose your model is demonstrating high variance across different training sets. Which of the following is NOT a valid way to try and reduce the variance ?

   (a) Increase the amount of training data in each training set.

(b) Improve the optimization algorithm being used for error minimization.

(c) Decrease the model complexity.

(d) Reduce the noise in the training data.

Descriptive Question

4 Write the optimization problem for Logistic regression. Also, write the gradient descent algorithm for the logistic regression problem with a brief explanation.

Quiz 5

Dhirubhai Ambani Institute of Information and Communication Technology
Foundation of Machine Learning (IT-582)
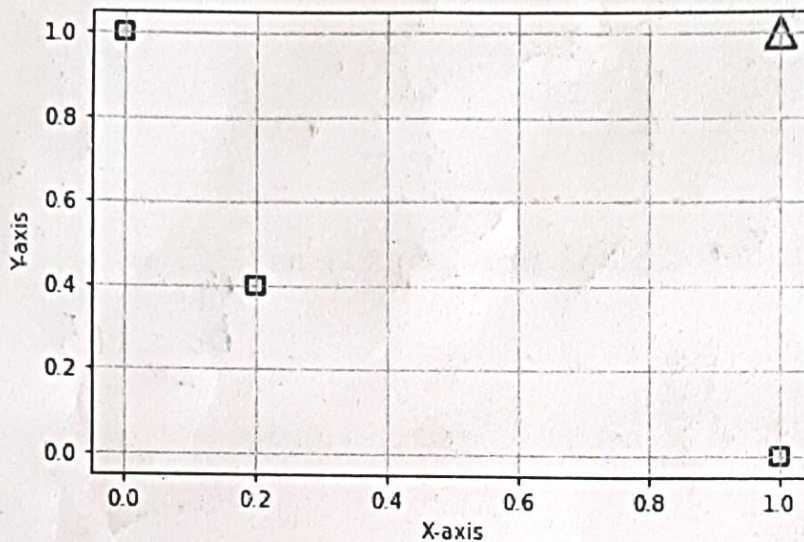Final Quiz, Date: Nov 25, 2022

Timing: 45 minutes   I Semester   Max mark: 10

Attempt all questions

1. **Principal Component Analysis:**
   Consider the following dataset, in which $X_i \in R^2$ and $y_i \in \{-1, 1\}$. Answer the following questions:

   (a) Find and draw all the principal components. (3)

   (b) Can we correctly classify this dataset by using a threshold function after projecting onto one of the principal components? If so, which principal component should we project onto? If not, explain in 1-2 sentences why it is not possible. (2)

2. **SVM**

   Let us consider the data set:

X1	X2	Y
0	1	-1
1	0	-1
2	1	1
0	0	-1

The plot of the dataset is as below:
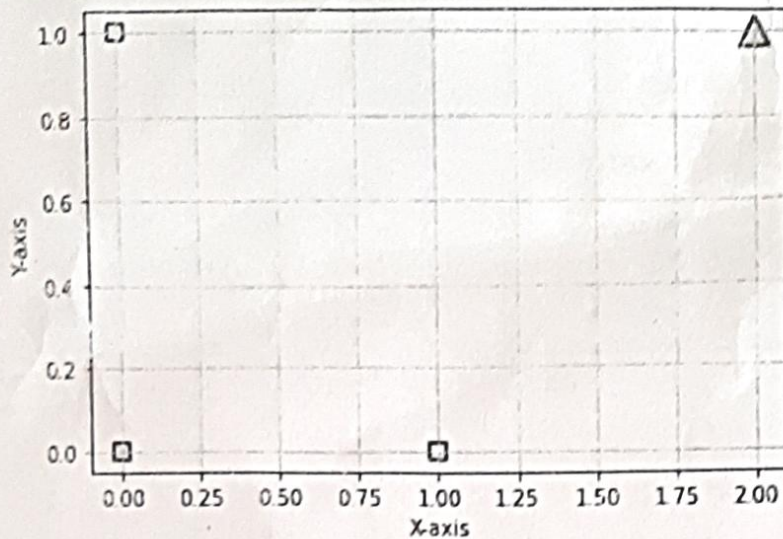
Let us consider the initial solution of SVM problem (1) as $\begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
with $\lambda = 1$.

$b_0 = 1.5$

$$\min_{w,b} \tau(w,b) = \min_{w,b} \frac{\lambda}{2}||w||^2 + \sum_{i=1}^{l} \max(1 - y_i(w^T x_i + b), 0) \tag{i}$$

(a) (i) Find the value of $\tau(w_0, b_0)$.
   (ii) Draw the classifies $w_0^T + b_0 = 0$ along with supportive hyperplanes.
   (iii) Find the width of margin. ✓   (2)

(b) (i) Let us perform two steps of gradient descent algorithm with $\eta = 0.2$ for obtaining $\begin{bmatrix} w_{(1)} \\ b_{(1)} \end{bmatrix}$ and $\begin{bmatrix} w_{(2)} \\ b_{(2)} \end{bmatrix}$.
   (ii) Find the width of margin classifier generated by $w_{(2)}^T x + b_{(2)} = 0$.   (3)

Best wishes