CS5691: Pattern recognition and machine learning
Quiz - 3
Course Instructor : Prashanth L. A.
Date : Mar-22, 2019 **Duration** : 40 minutes

Name of the student :
Roll No :

INSTRUCTIONS: For short answer questions, you do not have to justify the answer. For the rest, provide proper justification for the answers. Please use rough sheets for any calculations *if necessary.* Please **DO NOT** submit the rough sheets. DO NOT use pencil for writing the answers.

I. Short answer questions

1. Consider a two-class classification problem in a three-dimensional space, where the classes are linearly separable. Suppose the optimal separating hyperplane, i.e., one with highest margin, is given by

$$2x_1 + x_2 + x_3 - 3 = 0.$$

For each of the following training examples, identify the class label, and whether it is a support vector or not: ($\frac{1}{2}$ mark for each table entry)

example	class label	support vector or not
$(1, 1, 1)$		
$(\frac{1}{2}, 1, 0)$		
$(1, 1, 2)$		

Solution:

Example	Class Label	SVM or Not
(1,1,1)	+1	yes
(0.5,1,0)	-1	yes
(1,1,2)	+1	No

2. Consider a classification dataset with points $x_1 = -1, x_2 = 1, x_3 = 100$, and corresponding class labels $y_1 = -1, y_2 = +1, y_3 = +1$. Let $w_1 = 0$, and $w_2 = 1$. Let $L_1(w)$ and $L_2(w)$ denote the square-loss, and log-loss, respectively, for any $w \in \mathbb{R}$. Recall that square-loss is employed in linear regression, and log-loss (or cross-entropy loss) in logistic regression. Which of the following statements is true? (1 mark)

(a) $L_1(w_1) \leq L_1(w_2), \quad L_2(w_1) \leq L_2(w_2)$.
(b) $L_1(w_1) \leq L_1(w_2), \quad L_2(w_1) \geq L_2(w_2)$.
(c) $L_1(w_1) \geq L_1(w_2), \quad L_2(w_1) \leq L_2(w_2)$.
(d) $L_1(w_1) \geq L_1(w_2), \quad L_2(w_1) \geq L_2(w_2)$.

Answer: $\boxed{\textbf{b}}$

3. Consider the optimization problem for finding the maximum margin separating hyperplane, assuming that the classes are linearly separable:

$$\min_{W \in \mathbb{R}^d} W^\top W \quad \text{subject to } y_i \left(W^\top X_i + b\right) \geq 1, i = 1, \ldots, n, \tag{1}$$

where $\{(X_i, y_i), i = 1, \ldots, n\}$ is the training dataset. Consider the following variant of the problem in (1):

$$\min_{W \in \mathbb{R}^d} W^\top W \quad \text{subject to } y_i \left(W^\top X_i + b\right) \geq 100, i = 1, \ldots, n. \tag{2}$$

State whether the optimal hyperplane found by solving (1) is the same as that obtained by solving (2). (1 mark)

Answer: $\boxed{\textbf{YES}}$

4. Suppose we collect data about the number of hours spent per week by students of CS5691, and whether they passed or not. Let $\{(x_i, y_i), i = 1, \ldots, 86\}$ denote the dataset, where x_i is the hours spent by student i, and y_i is a boolean indicating whether the student passed or not. We perform logistic regression on this dataset, and obtain the following output: weight $w = 1.5$, and $b = -4$.

Using the logistic regression model, what is the probability that a student who studies 2 hours per week, passes the quiz? (1 mark)

Answer:

Solution:

$$w = 1.5 \ b = -4 \ x = 2$$
$$w^T * X + b = 1.5 * 2 - 4 = 1$$
$$\sigma(w^T * x + b) = \sigma(-1) = \frac{1}{1+e^{-(-1)}} = \frac{1}{1+e}$$

II. A problem that requires a detailed solution

1. Consider a two-class one-dimensional classification dataset with points $x_1 = -2, x_2 = 2, x_3 = 3$ having class label -1, and $x_4 = 1$ with class label $+1$.

Answer the following:

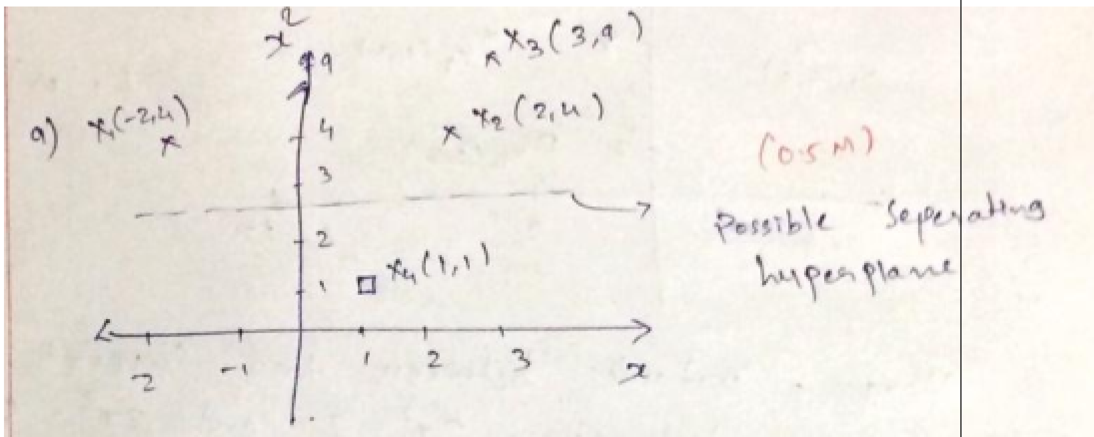
(a) Is the data linearly separable? ($\frac{1}{2}$ mark)

Answer: $\boxed{\textbf{NO}}$

(b) Consider the transformation $\phi(x) = (x, x^2)$. Form a two-dimensional dataset with inputs transformed using ϕ.

(a) Show that the transformed problem is linearly separable. ($\frac{1}{2}$ mark)

Solution:



(b) Find the maximum margin separating hyperplane in this transformed problem, either by solving the SVM optimization problem using KKT conditions, or a geometric argument. ($1\frac{1}{2}$ marks)

Solution: from geometry we can see $y = 2.5$ is the separating Hyper plane
Hyperplane is $y = w^{*T}x + b$
Let

$$
\begin{aligned}
w^* &= [\alpha, \beta], then \\
-2\alpha + 4\beta + b^* &= -1 \\
2\alpha + 4\beta + b^* &= -1 \\
\alpha + \beta + b^* &= 1
\end{aligned}
$$

Solving above three equation we get

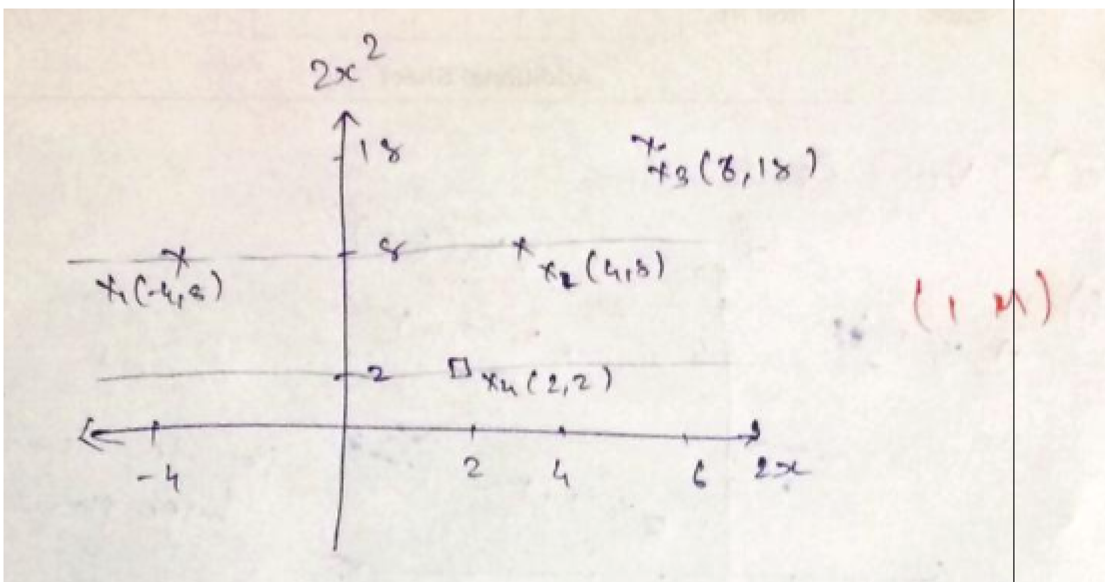
$$
\alpha = 0, \beta = \frac{-2}{3}, b^* = \frac{5}{3}
$$
$$
w^* = [0, \frac{-2}{3}], b^* = \frac{5}{3}
$$
$$
y = \frac{-2}{3}x^2 + \frac{5}{3}
$$

(c) What is the margin of the optimal hyperplane obtained above? ($\frac{1}{2}$ mark)

> **Solution:** margin $= \frac{2}{||w^*||} = \frac{2}{\frac{2}{3}} = 3$

(c) In the part above, suppose the transformation is $\widehat{\phi}(x) = (2x, 2x^2)$. Compare the margin resulting from $\widehat{\phi}$ to that obtained using ϕ. (1 mark)

Solution:



Margin= Distance between two support vectors
Which are $2x^2 = 2$ and $2x^2 = 8$
so , margin$= 8 - 2 = 6$