Student ID:
Name:

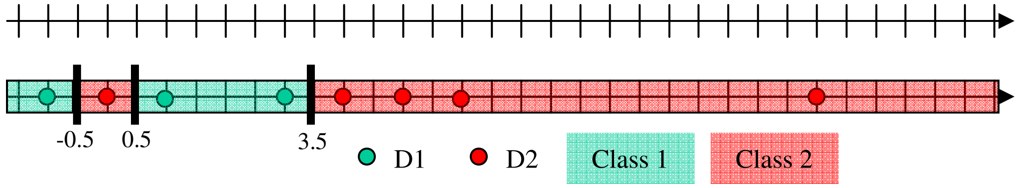
CS 442b-542b
Short Exam 1

Instructions: Show all the work you do. Use the back of the page, if necessary. Calculators are allowed, laptops are not allowed.

Problem 1 : Suppose we have a collected the following one dimensional samples from two classes: D1={-1,1,3}, D2={0,4,5,6,12}

(a) (10%) Draw the decision regions and decision boundaries for nearest neighbor approach with k=1, that is 1NN

Solution:

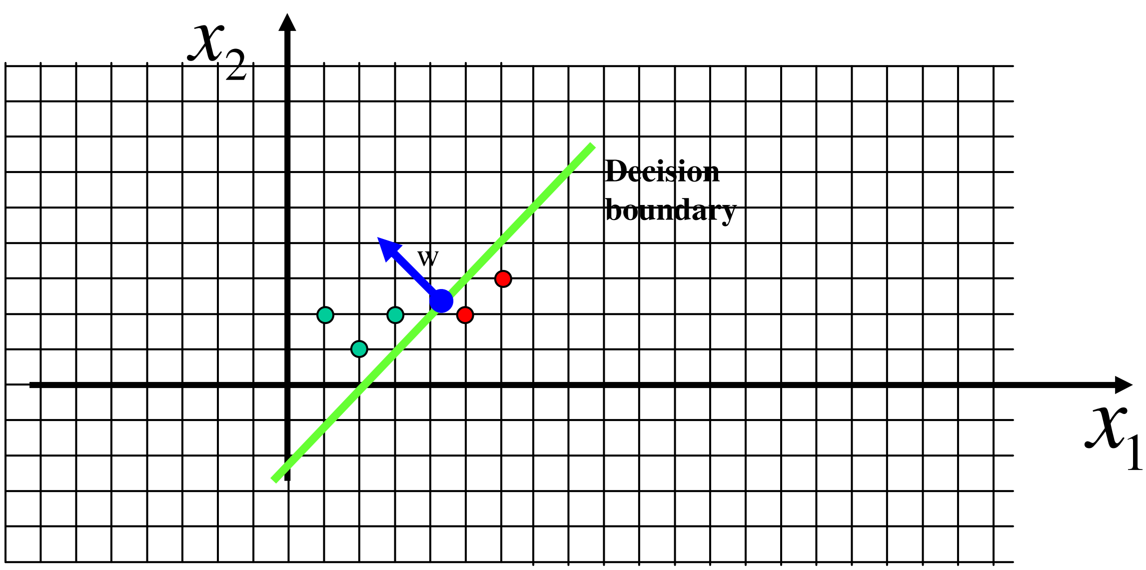


(b) (5%) classify sample ***0*** using ***knn*** algorithm with ***k*** = ***8***

Solotion: There are only 8 training samples in total, so when k=8, the class which has more samples will always win, no matter what the testing sample is. That is class 2.

Problem 2: Suppose we have 3 two dimensional training samples from class 1: $\begin{bmatrix}1\\2\end{bmatrix}$, $\begin{bmatrix}3\\2\end{bmatrix}$, $\begin{bmatrix}2\\1\end{bmatrix}$ and 2 two-dimensional training samples from class 2: $\begin{bmatrix}6\\3\end{bmatrix}$, $\begin{bmatrix}5\\2\end{bmatrix}$.

(a) (5%) In the graph ruled space provided below, sketch the samples and sketch the linear decision boundary which separates the samples. Also draw **w**, the normal vector to the decision boundary.



(b) (10%) Suppose you initialize the weight vector **a** to $a = \begin{bmatrix} 8 \\ -1 \\ -1 \end{bmatrix}$ and the step size to η=1. What is the value of the Perceptron objective function before you start optimizing? Recall that Perceptron objective function is $J_p(a) = \sum_{y \in Y_M} (-a^t y)$.

Solution: Let $y = [1 \quad x_1 \quad x_2]$

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \\ -1 & -6 & -3 \\ -1 & -5 & -2 \end{bmatrix} \times \begin{bmatrix} 8 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 5 \\ 1 \\ -1 \end{bmatrix}$$

$$Y_M = \{[-1 \quad -5 \quad -2]\}$$

$$J_p(a) = \sum_{y \in Y_M} (-a^t y) = [-1 \quad -5 \quad -2] \times \begin{bmatrix} -8 \\ 1 \\ 1 \end{bmatrix} = 1$$

(c) (10%) What is the weight vector **a** after you apply Perceptron algorithm to all samples one time, starting with weight vector **a** initialized as in (b)?

Solution:

$$a^{t+1} = a^t + \eta^t y_M = [8 \quad -1 \quad -1] + 1 \times [-1 \quad -5 \quad -2] = [7 \quad -6 \quad -3]$$

(d) (10%) Did your objective function increase or decrease with this new value of **a** that you computed in (c)? You can take an educated guess instead of actually computing the new value of the objective function.

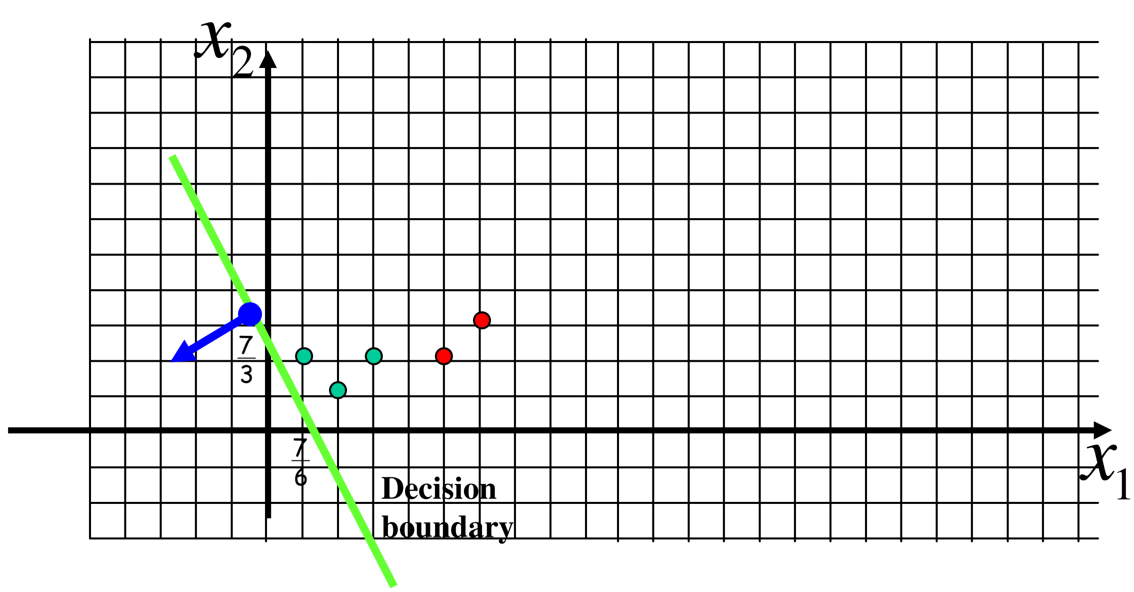
Solution: Increase

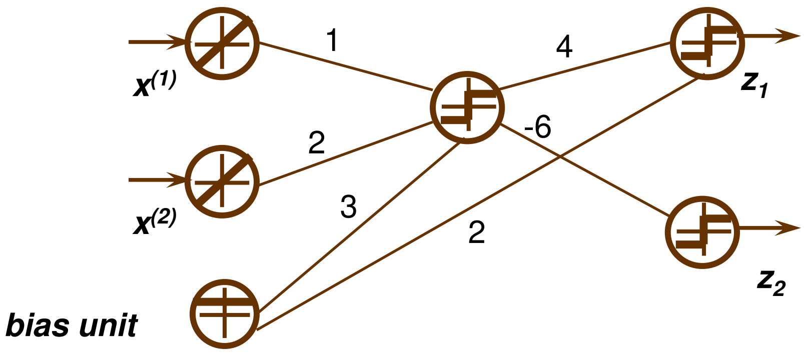
$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \\ -1 & -6 & -3 \\ -1 & -5 & -2 \end{bmatrix} \times \begin{bmatrix} 7 \\ -6 \\ -3 \end{bmatrix} = \begin{bmatrix} -5 \\ -17 \\ -8 \\ 38 \\ 29 \end{bmatrix}$$

$$Y_M = \{[1 \quad 1 \quad 2], [1 \quad 3 \quad 2], [1 \quad 2 \quad 1]\}$$

$$J_p(a) = \sum_{y \in Y_M} (-a^t y) = \text{sum}\left(\begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix} \times \begin{bmatrix} -7 \\ 6 \\ 3 \end{bmatrix} \right) = \text{sum} \begin{bmatrix} 5 \\ 17 \\ 8 \end{bmatrix} = 30$$

(e) (15 %) Plot the new decision boundary corresponding to the weight vector **a** computed in part (c), and also the normal **w**. You can plot it either below or on the previous page.





Problem 3:

(a) (15%) Write down the discriminant function corresponding to class 1 in the neural network above (the function corresponding to output unit 1). Assume that for the hidden unit and the output units we have the same function f.

Solution:

$$z_1 = f\left(4 \times f(x^1 + 2x^2 + 3) + 2\right)$$

(b) (15%) Suppose f(x) is the standard thresholding function (f(x) = 1 for positive x and f(x) = -1 otherwise). Suppose you input example $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ to the network above**.** Is it going to be classified as class 1 or class 2? Recall that you assign the class corresponding to the higher discriminant function.

Solution: Class 1

$$z_1([1 \quad -1]) = f\left(4 \times f(x^1 + 2x^2 + 3) + 2\right) = f(4 \times f(1 + 2 \times (-1) + 3) + 2) = 1$$

$$z_2([1 \quad -1]) = f\left(-6 \times f(x^1 + 2x^2 + 3)\right) = f(-6 \times f(1 + 2 \times (-1) + 3)) = -1$$

Problem 4 (5 %): Suppose you decide to do 4 fold cross validation, and the error rates for each of the 4 stages are: 10%, 5%, 20%, 5% What is the final estimate for the error rate using 4 fold cross validation?

Solution: Report the mean error: E = (10%+5%+20%+5%)/4 = 10%