**Name:**

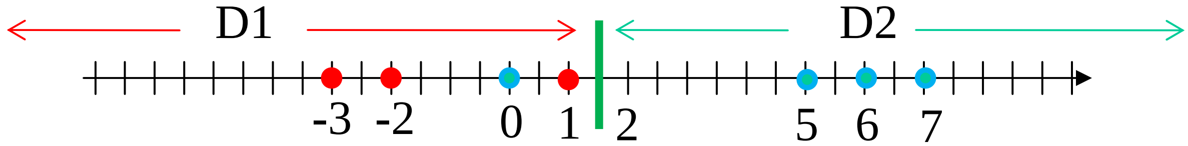
*CS 4442b-9542b, Winter 2010*

**Short Exam 1**

**Instructions:** Show all the work you do. Use the back of the page, if necessary. Calculators are allowed, laptops, cell phones, or any other communication devices are not allowed. This is an open notes/book exam. However, the sample exam 1 is not allowed.

**Problem 1 :** (20%) Suppose we have a collected the following one dimensional samples from two classes: D1={-3,-2,1}, D2={0,5,6,7}

a) (10%) Draw the decision regions and decision boundaries for nearest neighbor approach with k=3, that is 3NN




b) (10%) Can you remove any sample from your data and still have the same decision boundaries?

Answer: Remove sample -3, 6, 7 won't change the decision boundary

**Problem 2:** (10%)

Suppose we have 2 three dimensional training samples from class 1: $\begin{bmatrix} \mathbf{2} \\ -\mathbf{1} \\ -\mathbf{1} \end{bmatrix}$, $\begin{bmatrix} \mathbf{3} \\ -2 \\ -2 \end{bmatrix}$

and 2 three dimensional training samples from class 2: $\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$.

Write down a linear decision function g(x) which is positive for all samples from class 1 and negative for all samples from class 2. You do not have to "run" Perceptron algorithm for this problem, you can do it by observation. Of course, if you do wish to "run" Perceptron algorithm, you can.

Answer: Suppose the samples are x = $\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$. Then all $X_1$ in class 1 is positive and all $X_1$ in class 2 is negative. So if we choose g(x) = [1 0 0] *x+0, then it would be positive for all samples in class 1 and negative for all samples in class 2.

**Problem 3:** Suppose we have 2 two dimensional training samples from class 1: $\begin{bmatrix}2\\2\end{bmatrix}$, $\begin{bmatrix}2\\1\end{bmatrix}$

and 3 two dimensional training samples from class 2: $\begin{bmatrix}2\\-1\end{bmatrix}$, $\begin{bmatrix}0\\0\end{bmatrix}$, $\begin{bmatrix}2\\0\end{bmatrix}$

(a) (10%) What is the weight vector **a** after you apply ***batch*** Perceptron algorithm for one iteration, starting with weight vector $a = \begin{bmatrix}1\\-1\\-1\end{bmatrix}$ and using step size $\eta = 2$

Answer: 1) Augment all samples:

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

Then we have:

$$A \times a = \begin{bmatrix}1 & 2 & 2\\ 1 & 2 & 1\\ -1 & -2 & 1\\ -1 & 0 & 0\\ -1 & -2 & 0\end{bmatrix} \times \begin{bmatrix}1\\-1\\-1\end{bmatrix} = \begin{bmatrix}-3\\-2\\0\\-1\\1\end{bmatrix}$$

The misclassified samples are: $Y_m = \{[1\quad 2\quad 2],[1\quad 2\quad 1],[-1\quad 0\quad 0]\}$, therefore,

$$a_{new} = a + \eta \times \sum_{y \in Y_m} y = [1\quad -1\quad -1]' + 2 \times [1\quad 4\quad 3]' = [3\quad 7\quad 5]'$$

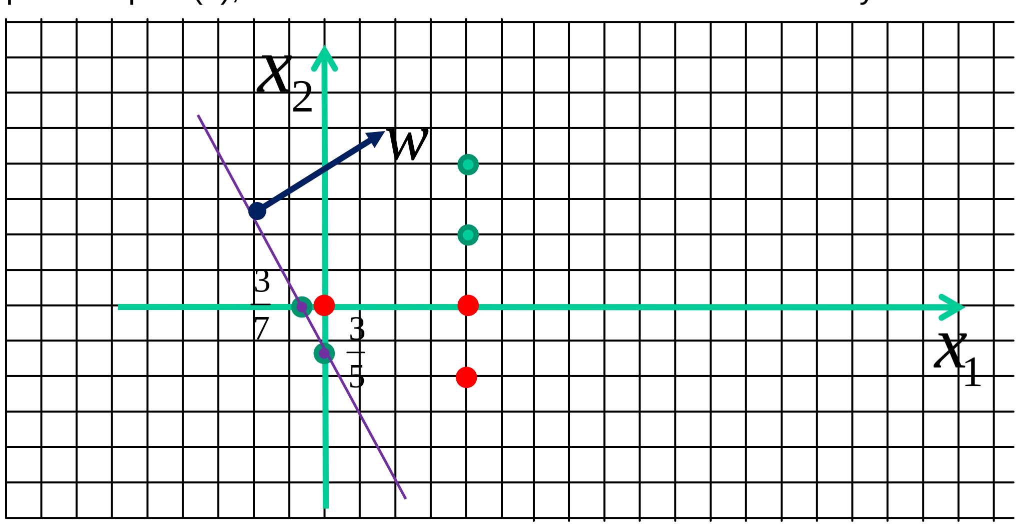
(b) (10%) What is the value of the Perceptron objective function for the weight vector ***a*** computed in part (a)?

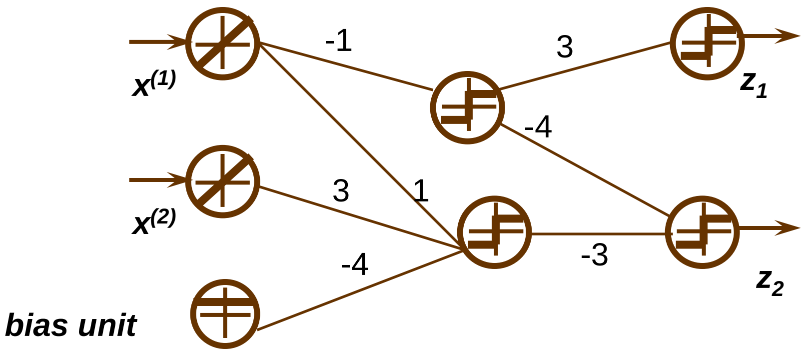
Answer:

$$A \times a_{new} = \begin{bmatrix}1 & 2 & 2\\ 1 & 2 & 1\\ -1 & -2 & 1\\ -1 & 0 & 0\\ -1 & -2 & 0\end{bmatrix} \times \begin{bmatrix}3\\7\\5\end{bmatrix} = \begin{bmatrix}27\\22\\-12\\-3\\-17\end{bmatrix}$$

Then: $J_p(a_{new}) = \sum_{y \in Y_m} (-y a_{new}) = -(-12 + (-3) + (-17)) = 32$

(c) (15 %) Plot the samples, the decision boundary corresponding to the weight vector **a** computed in part (a), and also the normal **w**. Make sure to label your axis.






**Problem 4:**

(a) (10%) Write down the discriminant function corresponding to class 2 in the neural network above ( the function corresponding to output unit 2). Assume that for the hidden unit and the output units we have the same function f.

Answer: $$z_1 = f(3\times f(-1\times x_1))$$

$$z_2 = f(-4\times f(-1\times x_1) - 3\times f(x_1 + 3\times x_2 - 4))$$

(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 $x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 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 value of discriminant function.

Answer:

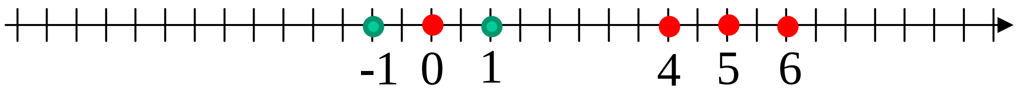
$$z_1 = f(3\times f(-1\times 2)) = -1$$

$$z_2 = f(-4\times f(-1\times 2) - 3\times f(2 + 3\times 1 - 4))$$

$$= f(-4\times f(-1\times 2)) - 3\times f(2 + 3 - 4)) = f(4-3) = 1$$

So x is classified as class 2

**Problem 5** (10 %): Suppose we have a collected the following one dimensional samples from two classes: D1={-1,1}, D2={0,4,5,6}, and you use 1NN (nearest neighbor with k = 1) classifier. What is the 3-fold cross validation error? You may wish to draw the data on the plot below, but you don't have to. **Show all the work!**




**Answer: Suppose we randomly divide the samples into three folders: {-1, 0}, {1, 5} and {4, 6}. When take {-1, 0} as the testing data, the error rate is 50%. Take {1, 5} as testing data, the error rate is 50%. Take {4, 6} as testing data, the error rate is 0%. Then the mean error rate is : (50% +50% +0%)/3 = 33.3%**

**(Or divide the samples into: {-1, 1}, {0, 4} and {5, 6}, then the error rates are: 100%, 50% and 0%. The mean error rate is (100% + 50% + 0%)/3 = 50%)**