

Name: _____

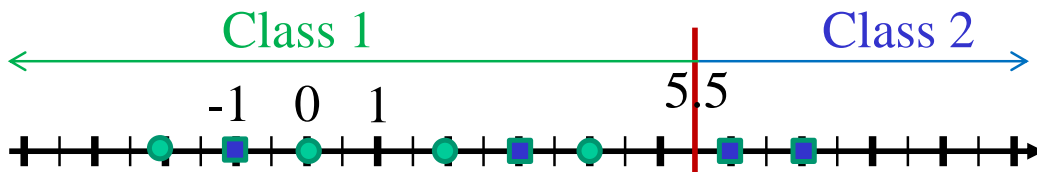
CS4442/8952: Artificial Intelligence II
Winter 2017: Quiz 1
SOLUTION

Instructions:

Show all the work you do. Use the back of the page, if necessary. Basic calculators (no graphing functions) are allowed. Laptops, cell phones, or any other communication devices are not allowed. This is an open notes exam. However, the sample exam 1 solution is not allowed.

1. (15%) Suppose we have the following one dimensional samples from two classes: $D_1 = \{-2, 0, 2, 4\}$ and $D_2 = \{-1, 3, 6, 7\}$. We use kNN classifier with $k = 2$, and whenever there is a tie (that is one neighbor is from class 1 and another from class 2), we always decide class 1. Draw the decision regions and decision boundaries for this case.

SOLUTION:



2. (10 %) Suppose the training examples from class 1 and 2 are, respectively:

$$C_1 = \left\{ \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ -3 \\ 3 \end{bmatrix} \right\}, C_2 = \left\{ \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ -1 \end{bmatrix} \right\}$$

Write down a linear discriminant function $g(x)$ which is positive for all samples from class 1 and negative for all samples from class 2. In other words, $sign(g(x))$ classifies all samples correctly. Hint: it is easiest to do by observation, not gradient descent.

SOLUTION: Many possible solutions, one of the simplest is: $g(x) = -x_1$

3. (15 %) Let $W = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & -1 \\ 3 & 1 & -1 \end{bmatrix}$ be multiclass classification matrix.

(a) (5%) Using W above, how does sample $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ get classified?

SOLUTION:

First augment sample to get

$z = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$. Then matrix multiply W and z :

$$Wz = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & -1 \\ 3 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}$$

Largest entry is in the first row of Wz , so sample x is classified as class 1.

(b) (5%) What is the loss for sample x above under quadratic loss function? Assume the true class of x is 2.

SOLUTION: Target representation for a sample of class 2 is $y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

Quadratic loss is $\frac{1}{2} \|y - Wz\|^2 = \frac{1}{2}(25 + 4) = \frac{29}{2}$.

(c) (5%) What is the loss for sample x above under Perceptron loss function? Assume the true class of x is 2.

SOLUTION: Perceptron loss is the largest row of Wz minus for of Wz corresponding to the correct class. It is $5 - 1 = 4$ in this case.

4. (35 %) Suppose the training examples from class 1 and 2 are, respectively:

$$C_1 = \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}, C_2 = \left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right\}$$

- (a) (10 %) Suppose we start Perceptron algorithm with $a = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$ and the learning rate of $\alpha = 1$. What is the weight vector a after you apply Perceptron Single Sample algorithm for one iteration?

SOLUTION:

First sample gives $a^t \cdot [1, 3, 1] = 1$, classified correctly, so no change to a . Second sample gives $a^t \cdot [1, 2, -1] = 2$, classified correctly, so no change to a . Third sample gives $a^t \cdot [1, 1, -3] = 3$, misclassified (true class is 2, classified as class 1). Change a according to Perceptron rule:

$$a = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} + (-1) \cdot \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

Forth sample with the new a gives $a^t \cdot [1, 3, 1] = -2$, classified correctly. We are now done with one iteration of the single sample rule.

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

SOLUTION: To see which samples are misclassified, matrix multiply matrix C below that contains all augmented samples and vector a :

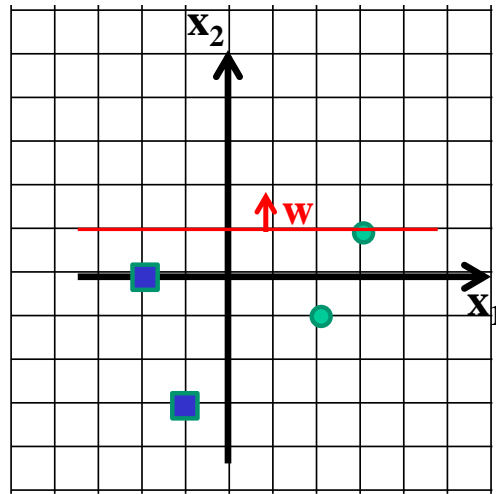
$$C \cdot a = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 2 & -1 \\ 1 & 1 & -3 \\ 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \\ -8 \\ -2 \end{bmatrix}$$

First two samples are class 1, next two are negative class. Therefore a misclassified the second sample. Loss is $-a \cdot [1, 2, -1] = 4$.

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

SOLUTION:

$$a = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix} \text{ corresponds to line } -2 + 2x_2 = 0, \text{ or, equivalently, line } x_2 = 1.$$

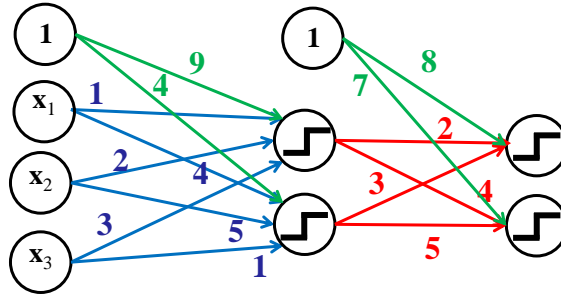


5. (25 %)

(a) (15%) Draw the neural network corresponding to the weights stored in the matrices and bias vectors below.

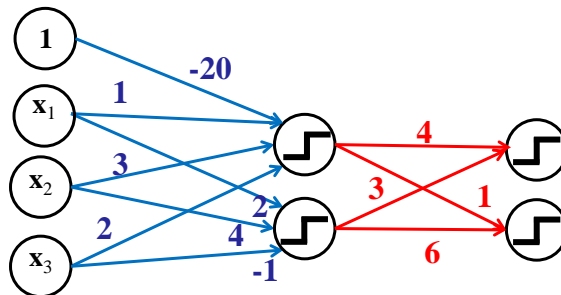
$$W^1 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & -1 \end{bmatrix} \quad b^1 = \begin{bmatrix} 9 \\ 4 \end{bmatrix} \quad W^2 = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \quad b^2 = \begin{bmatrix} 8 \\ 7 \end{bmatrix}$$

SOLUTION:



(b) (10 %) How is example $\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$ classified by the network below? Show all the work.

Assume the $sign()$ activation function for all the hidden and output units.



SOLUTION:

Hidden layer:

- First unit outputs value $sign(-20 - 1(-1) + 2(3) + 1(2)) = -1$.
- Second unit outputs value $sign(-1(2) + 2(4) + 1(-1)) = 1$.

Output layer:

- First unit outputs value $sign(-1(4) + 1(3)) = -1$.
- Second unit outputs value $sign(-1(1) + 1(6)) = 1$.

Since the second output is higher, sample classified as class 2.