1. Under what condition(s) will the point corresponding to the weight vector of a hyperplane be on the plane's positive side? Justify your answer.
2. What is the disadvantage of batch perceptron as compared to online perceptron?
3. For the objective function shown below, give the gradient descent update equation for the weight vector $\underline{w}$.

$$J(\underline{w}) = \sum_{misclassified\ \underline{x}} (\underline{w}.\underline{x})^2$$

4. A 4-class least-squares classifier has produced the following decision functions. (a) Use them to classify the point $(1, -1, 2)$. (b) What is the decision boundary between class 3 and 4? What was the dimensionality of the training data? How many samples where used in training?

$$d_1(\underline{x}) = 2x_1 + x_2 - x_3 + 3$$
$$d_2(\underline{x}) = -x_1 - x_2 + 2x_3 + 1$$
$$d_3(\underline{x}) = x_2 + 2$$
$$d_4(\underline{x}) = -x_2 + 2x_3 + 2$$

1- When the offset is zero   or $W_{\ell+1} = 0$ 3 pts

Given $d(\underline{x}) = W_1 x_1 + W_2 x_2 + \cdots W_\ell x_\ell$ Point Corresponding to its weight vector

is $\underline{x}^* = [W_1 \ W_2 \ \cdots W_\ell]$

So $d(\underline{x}^*) = W_1 W_1 + W_2 W_2 + \cdots W_\ell W_\ell$

this quantity is always positive   3 pts

2- to do batch perceptron, All of data must be present at the   3 pts
Same time. On-line perceptron deals with the data one at a time.

3- $\nabla_{\underline{w}} J = 2 \sum_{misclassified\ \underline{x}} (\underline{W} \cdot \underline{x}) \underline{x}$ 3pts

$\underline{W}(k+1) = \underline{W}(k) - 2\int \sum_{misclassified\ \underline{x}} (\underline{W} \cdot \underline{x}) \underline{x}$ 3pts

---

4-

(a) $d_1(\underline{x}) = 2$ $d_2(\underline{x}) = 5$ $d_3(\underline{x}) = 1$ $d_4(\underline{x}) = 7$ 3pts

$\Rightarrow \underline{x} \in$ class 4

(b) $d_3(\underline{x}) - d_4(\underline{x}) = x_2 + 2 - (-x_2 + 2x_3 + 2) = 2x_2 - 2x_3$ 3 pts

(b) $d_3(x) - d_4(x) = x_2 + 2 - (-x_2 + 2x_3 + 2) = 2x_2 - 2x_3$ 3 pts

(c) $\ell = 3$ 3 pts   (d) Cannot be determined 3 pts

Max: 30 pts
Min: 3 pts