EEE 485-585 FALL 2018 PROBLEM SET 3

Due Date: Wednesday, 4 December 2018, 17:00 (room EE-212)

Question 1 [25 pts]

Download the dataset perceptrontest.mat from moodle. In this dataset **X** denotes the 10000 by 11 feature matrix, where the feature in the first column is all ones (this corresponds to the bias term). The response vector **Y** is binary.

(a) [5 pts] Code perceptron (take the learning rate equal to 1).

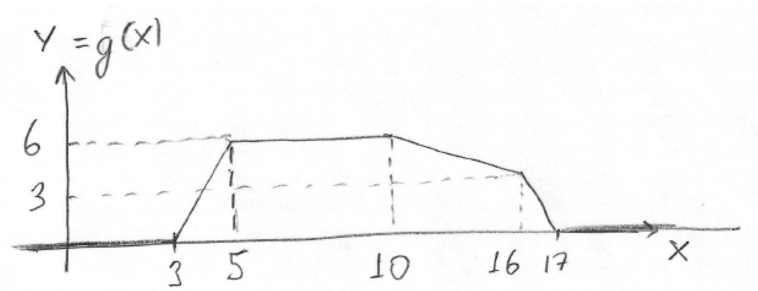
(b) [10 pts] Run perceptron by deterministically iterating over the dataset. Start with the all zero initial weight vector. In each iteration sweep through all the samples in the dataset starting from the first data instance in **X** until the last data instance in **X**. Perform 5000 iterations this way. Plot the total number of mistakes that perceptron makes on this dataset as a function of the number of iterations.

(c) [5 pts] Is the dataset linearly separable? If your answer is yes, give the equation of the separating hyperplane found by perceptron and explain if there are any other separating hyperplanes.

(d) [5 pts] Assume that some learning algorithm provides 10 different separating hyperplanes for a given binary classification dataset. Which one of these would you use as your classifier? Explain clearly.

Question 2 [25 pts]

Design a neural network that implements the function given below. The neural network must have one hidden layer and one output layer. The hidden layer uses rectified linear unit (ReLU) as activation function, i.e., $\phi(v) = \max(0, v)$. The output layer uses linear activation function, i.e., $f(v) = v$. Specify all the weights and draw the diagram of this neural network.



Question 3 [25pts]

Assume that samples are given as $x_1 = [1, 4, 400]$, $x_2 = [3, 5, 500]$ $x_3 = [3, 2, 450]$, $x_4 = [3, 1, 400]$, and $x_5 = [1, 3, 425]$. We will use Principal Component Analysis (PCA) for dimension reduction. (For this question, you can use MATLAB or Python)

(a) [10 pts] Only centralize the features. Find eigenvectors and eigenvalues of the sample covariance matrix and interpret the result. Why one of the eigenvalues is much larger than the others? If we use the eigenvector that corresponds to the highest eigenvalue, is this good representation of the samples?

(b) [10 pts] Scale each feature such that it becomes both zero mean and unit variance. Find eigenvectors and eigenvalues (for PCA). By using these the first two principal components, express each sample in the 2D space. Report the proportion of variance explained by these principal components. Is this a good representation of the samples?

(c) [5 pts] When we apply PCA to 3 dimensional data, what can we say about the data if only the largest two eigenvalues of Σ are non-zero?

Question 4 [25 pts]

(a) [10 pts] Consider ICA. We can only recover the unmixing matrix $\mathbf{W}$ up to a certain degree. Give example scenarios in which the following cases happen.

- The recovered signal $[\hat{s}_1(t), \ldots, \hat{s}_p(t)]$ will be a permutation of $[s_1(t), \ldots, s_p(t)]$.
- The sign of the recovered signal $\hat{s}_j(t)$ can be different from $s_j(t)$.

(b) [15 pts] Why do we need the sources to be non-Gaussian in ICA. Can we recover $\mathbf{W}$ correctly if the sources are Gaussian? Prove your claim.