Indian Institute of Science

E9-253: Neural Networks and Learning Systems-I

Instructor: Shayan Srinivasa Garani
Home Work #1, Spring 2019

Late submission policy: Points scored = Correct points scored $\times e^{-d}$, d = # days late

Assigned date: Jan. 29^{th} 2019 **Due date:** Feb. 12^{th} 2019 in class

---

PROBLEM 1: Write down the equation for the output y of the network as shown in Figure 1. (5 pts.)

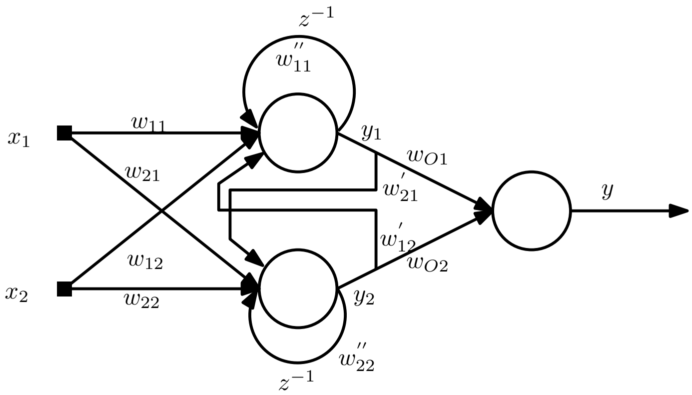

FIGURE 1. Network with lateral connections

PROBLEM 2: Let $\{\overline{x}_n\}_{n=1}^{\mathrm{N}_x} \in \mathbb{R}^{\mathrm{d}}$ be the set of data points in a d-dimensional space. Define the convex hull as follows:

$$\overline{x} = \sum_{n=1}^{\mathrm{N}_x} \alpha_n \overline{x}_n \tag{1}$$

where $\alpha_n \geq 0$ and $\sum_{n=1}^{\mathrm{N}_x} \alpha_n = 1$. Consider a second set of data points $\{\overline{z}_m\}_{m=1}^{\mathrm{N}_z} \in \mathbb{R}^{\mathrm{d}}$ and define the corresponding convex hull. The two datasets are linearly separable if there exists a vector $\overline{w}$ and a scalar b such that $\overline{w}^{\mathrm{T}}\overline{x}_n + b > 0$ for all $\overline{x}_n$ and $\overline{w}^{\mathrm{T}}\overline{z}_m + b < 0$ for all $\overline{z}_m$. Show that

(a) If two convex hulls intersect, the two datasets cannot be linearly separable.
(b) If the two datasets are linearly separable, their convex hulls do not intersect.

(10 pts.)

PROBLEM 3: Consider a N-class classification problem of the input data $\overline{x} \in \mathbb{R}^{\mathrm{d}}$. Let $\{L_{kj}\}_{k,j=1}^{N}$ be the cost associated with assigning class label k to an input which belongs to class j, $\mathrm{P}\left(\mathcal{C}_k \mid \overline{x}\right)$ be the condition probability that the input $\overline{x}$ belongs to class $\mathcal{C}_k$ and, R_k be the set of all $\overline{x}$ for which the class label is $\mathcal{C}_k$.

(a) Compute the total expected loss.
(b) Derive the optimal rule for assigning class labels.

(10 pts.)

PROBLEM 4:

(a) The perceptron may be used to perform numerous logic functions. Demonstrate the implementation of the binary logic functions **AND**, **OR**, and **COMPLEMENT**. Provide scatter plots with decision boundaries. Attach your codes in an Appendix.
(b) A basic limitation of the perceptron is that it cannot implement the **EXCLUSIVE OR** function. Explain the reason for this limitation.

(10 pts.)

PROBLEM 5:

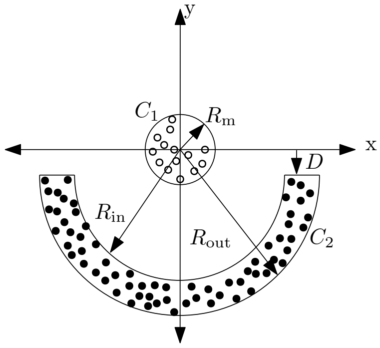

FIGURE 2. Set of data points uniformly distributed within a circle of radius $R_{\rm m}$ and crescent with inner and outer radius $R_{\rm in}$ and $R_{\rm out}$.

(a) Generate a set of $N = 2000$ data points (with 1000 data points in each class) as shown in Figure 2 with $R_{\rm m} : R_{\rm in} : R_{\rm out} = 1 : 2 : 3$, and $D = 0$. Provide scatter plots and attach your codes in an Appendix.

(b) By changing the value of D, classify the set of data points into classes $\mathcal{C}_1$ and $\mathcal{C}_2$ using the perceptron algorithm configured in online and batch modes. Provide scatter plots with decision boundaries. Attach your codes in an Appendix.

(c) Fix the D value for linear separability. Start with different initial conditions for the weight vector and the bias. Check whether you get the same decision boundary and comment upon this.

(d) With the initial weight vector fixed in online mode, randomize the sequence of inputs to the perceptron. Check whether you get the same decision boundary and comment upon this.

(e) Add Gaussian noise with 0 mean and variance ranging from 1 to $R_{\rm in}$ (in steps of 2) to the set of data points shown in Figure 2. What is your stopping criterion for learning? What can you comment upon the classification accuracy experimentally?

(f) Repeat the experiment 5(b) (fix the D value for linear separability) by varying the learning rate η from 0.1 to 1 in steps of 0.2. Report the number of steps $n_{\rm max}$ required for the convergence of perceptron algorithm for each value of η and fix the initial weight vectors in all your experiments.

(g) Justify your observations made in 5(f) theoretically.

(35 pts.)