Indian Institute of Science

E9: 253 Neural Networks and Learning Systems - I

Instructor: Shayan Srinivasa Garani

Final Exam, Fall 2020

Name and SR.No:

Instructions:

- This is a take home final exam. There are five main questions. None of them have negative marking.
- Attempt all of them with careful reasoning and justification for partial credit.
- There is absolutely no collaboration with any one or referring to any code from the web except referring to the web/class notes or text for any source of conceptual information you may need. You should not refer to any solutions if there are any. Cite all the resources used.
- This exam is assigned on 25/1/2021 at 12 midnight and to be turned in on 28/1/2021 12 noon. This is a hard deadline.
- Do not panic, do not cheat, good luck!

Question No.	Points scored
1	
2	
3	
4	
5	
Total points	

PROBLEM 1:

This problem has five parts. Consider the scatter plot of points shown in Figure 1.

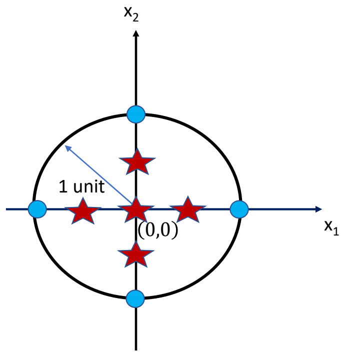

FIGURE 1. Scatter plot of points belonging to two classes ω_1 and ω_2 within a circular geometry centered at the origin and radius 1 unit. Equally likely points belonging to class ω_1 are located at coordinates $(1,0)^{\mathrm{T}}$, $(0,-1)^{\mathrm{T}}$, $(-1,0)^{\mathrm{T}}$ and $(0,1)^{\mathrm{T}}$. Equally likely points belonging to class ω_2 are located at coordinates $(0,0)^{\mathrm{T}}$, $(0.5,0)^{\mathrm{T}}$, $(0,-0.5)^{\mathrm{T}}$, $(-0.5,0)^{\mathrm{T}}$ and $(0,0.5)^{\mathrm{T}}$.

(1) Suppose we wish to non-linearly transform the points in Figure 1 via the inverse multiquadrics function given by $\phi(r) = \frac{1}{\sqrt{(r^2+1)}}$, $r > 0$. Sketch the points after the transformation. Are the points linearly separable? (2 pts.)
(2) Algorithmically, how would you approach to build a linear classifier towards solving the pattern classification problem post the non-linear transform in part 1? (3 pts.)
(3) Assuming $P(\omega_1) = p$, what are the optimal Bayesian rule and misclassification rates for pattern classification post transformation? (5 pts.)
(4) Does the transformation in part 1 belong to the class of Green's functions? Justify mathematically with details. (5 pts.)
(5) Suppose we wish to classify the points in Figure 1 using a radial basis function network, show the final neural network architecture with all the connecting weights. You need not assume any *apriori* class probability while solving this part. (5 pts.)

PROBLEM 2: Consider the architecture of a neural network taking inputs from R^4. The desired response is a vector in R^2. The data comprises $\{(\underline{x}_i, \underline{d}_i)\}_{i=1}^{N}$. The network comprises a single hidden layer with 3 neurons, an output layer with 2 neurons. This network has feedforward connections from input to the hidden and from the hidden to the output layer. However, unlike what we did in the class, there are also lateral connections among the output neurons. You can assume the usual notations as in the class for the feedforward connections. However, for the lateral connection use the notation $a_{j,k}[n]$ indicating the synaptic connection from the k^{th} output node to the j^{th} output node.

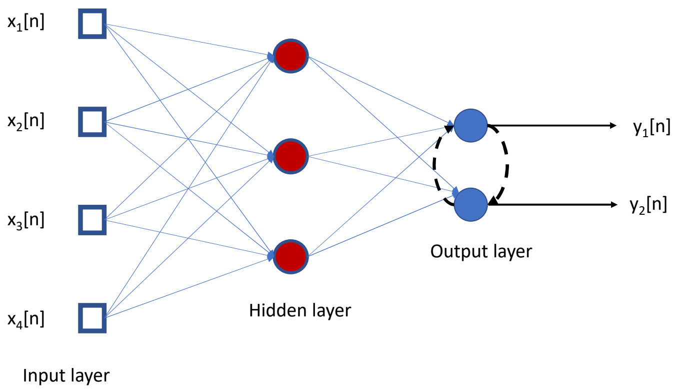

FIGURE 2. Architecture of a single hidden layer multilayer perceptron with lateral connections at the output. The inputs are from a 4-dimensional space. There are 3 hidden neurons and 2 output neurons.

For the above setup, derive the backpropagation algorithm from first principles assuming the usual quadratic cost function and a general non-linear activation function $\phi(x)$. Show all your steps carefully, distilling all the computation details in the form of a procedure towards the end. You may assume any information as needed, but state them clearly. (20 pts.)

PROBLEM 3: Points $\{\underline{x}_i\}_{i=1}^N \in R^d$ with data point density $f_{\underline{X}}(\underline{x})$ are provided to you. You need to cluster them into L clusters $L < N$. The clustering of the points is governed by a weighted cluster distortion function based on the L_2 norm-squared distance between the point $\underline{x}_i$ and the cluster representatives $\underline{\mu}_j$, $j = 1, ..., L$. The weight w_{ij} is chosen as

$$w_{ij} = \begin{cases} 1, & \text{if } \underline{x}_i \in \text{cluster } j. \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Set up the cost function and solve for obtaining optimal $\{\underline{\mu}_j\}_{j=1}^L$ analytically. How would you determine $\{\underline{\mu}_j\}_{j=1}^L$ algorithmically?

(20 pts.)

PROBLEM 4: Consider the XOR problem over 2 variables x_1 and x_2 that can take $ternary$ values $0, 1, 2$ according to the equation $y = x_1 \oplus_3 x_2$ i.e., $(x_1 + x_2) \mod (3)$. Design an SVM engine to solve this problem. Show your decision boundaries with all the relevant equations from first principles. (20 pts.)

PROBLEM 5: You are given labeled raw speech signals comprising spoken digits from various speakers. The files are essentially digital samples of length N each. You are asked to build a convolution neural network (CNN) towards training the network to identify the digit and the speaker. Show how you will build this up using a suitable mask of size L and weight sharing. Show your network architecture with all the details. You need not derive the algorithm from scratch. I need you to show how you would set up the network architecture with all the computation details as we did in the class that can be possibly machine coded using software. You can make any assumptions needed, but state them clearly. (20 pts.)