Indian Institute of Science

E9: 253 Neural Networks and Learning Systems - I

Instructor: Shayan Srinivasa Garani

Mid Term Exam, Spring 2019

Name and SR.No:

Instructions:

- You are allowed to refer to only 4 written sheets of A4 size during the exam and nothing else.
- There are four 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 except referring to your sheets.
- This is an in-class exam for 1.5 hrs.
- Do not panic, do not cheat, good luck!

Question No.	Points scored
1	
2	
3	
4	
Total points	

PROBLEM 1: This problem has four parts.

(1) How should the learning rate be changed when the derivative of the cost function w.r.t a synaptic weight has (a) the same algebraic sign over several iterations (b) alternating algebraic signs over consecutive iterations of the algorithm? Justify your update rules with arguments. (4 pts.)

(2) Consider a multilayer neural network where neurons operate in the linear region. This network is equivalent to a single layer feed forward network. Justify if the statement is true or false. (4 pts.)

(3) A large number of data points are collected containing a single predictor and a quantitative response. We then fit a linear regression model to the data, as well as a separate quartic regression i.e. $Y = \sigma_0 + \sigma_1 X + \sigma_2 X^2 + \sigma_3 X^3 + \sigma_4 X^4 + \epsilon$. Suppose that the true relationship between X and Y is linear i.e. $Y = \sigma_0 + \sigma_1 X + \epsilon$. Consider the training residual sum of squares (LRSS) for the linear regression, and also the training QRSS for the quartic regression. Is there any relationship between LRSS and QRSS i.e. equality or inequality? Is there not enough information to tell? Justify your answer. (4 pts.)

(4) In the class, we looked into the architecture of a convolution neural network over images using weight sharing i.e., the same filter or *kernel* is applied to various regions of an input image to extract the local features through that *kernel* before subsequent processing. Consider a 3×3 image, and a 2×2 kernel. Suppose we want a 2×1 vector after max-pooling. By rastering the image as 1D vector, illustrate the idea of weight sharing clearly through a figure by showing the connections of the input nodes (in rastered form) to the convolution layer and subsequently to the pooling layer. Indicate all the labels over the edges and input coordinates carefully. What is the advantage of weight sharing over a fully connected network? (8 pts.)

PROBLEM 2: Design a neural network to solve the XOR problem with three Boolean variables x_1, x_2 and x_3 using hidden layer(s) with *minimal* number of neurons. Show all your steps carefully including your choice of the activation function. (10 pts.)

Extra credit: Can you generalize your design to solve the XOR problem with N variables? (5 pts.)

PROBLEM 3: Prove the convergence of the batch perceptron algorithm for linearly separable patterns. You may assume that the learning rate is the same over all the epochs. Define all your variables carefully as you derive the analysis. (10 pts.)

PROBLEM 4: By expanding the error surface $\mathcal{E}_{av}(\mathbf{w}(n) + \Delta\mathbf{w}(n))$ at iteration n up to a second order approximation, how should the optimal adjustment to the weight vector $\Delta\mathbf{w}(n)$ be done? Clearly indicate every scalar element within a vector or a matrix as applicable within this set up. What are the conditions for this solution to exist? What benefits does one get using second order approximations over the standard gradient descent rule? (10 pts.)