SML 2024, Monsoon, Quiz 3, Dur. 1 hr 10 mins.

Q1. Consider the NN in Figure 1. It has an input layer with one node and output layer with 2 nodes, and one hidden layer with a single node. It accepts a 1d input and gives a 2d output. $\hat{y}_i$ denotes the output of a i^{th} output node. S denotes the softmax operation which takes b_1, b_2 as input and gives $\hat{y}_1, \hat{y}_2$ as output. U denotes the weight of the edge connecting node x to the hidden layer node. Activation on feature a is identity. In other words $z = a$. The loss function is multi-class cross entropy. Find the equation for updation of U using the given information. Assume all weights are randomly initialized and only one train sample is used $\{x, y\}$ where the true one-hot encoded label is $y = [1, 0]^\top$. [3]

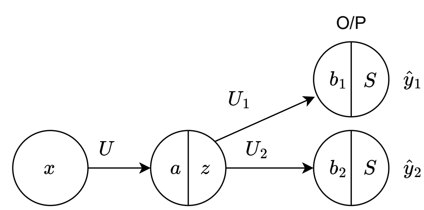

Figure 1: NN with one hidden layer of single node. S denotes softmax operation. $\hat{y}_i$ denotes the output of respective nodes.

Q2. Rosenblatt's perceptron covered in the lecture uses distance of miss-classified points to update the weight vector. It first computes the signed distance as $\beta^\top x + \beta_0$ and combines it with true label to obtain the loss function. Weights are updated via gradient descent. Finally sign of distance is used to compute the class of a test point.

a. What is the condition under which the gradient descent converges for Rosenblatt's perceptron, that is the weights will not be updated any further? [.5]

b. Give a case under which gradient descent does not converge for Rosenblatt's perceptron. [.5]

c. Suppose we modify Rosenblatt's perceptron as given in Figure 2. x is a scalar input, $z = sign(b(\beta_1 x + \beta_2 x))$, b is a known constant and is not learnable. β_1, β_2 are weights and learnable parameters. $a_i = \beta_i x$, The decision for class of x is given by z. If $z = 1$, then class is 1 else -1. How will the loss function used in Rosenblatt's perceptron modify if we use this network. [.5]

d. Find the update equation for β_1. [.5]

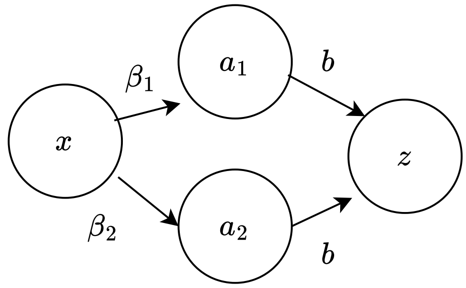

Figure 2: Modified Rosenblatt's perceptron.

Q3. Given a regression train dataset $D = \{(x, y)\} = \{(0, 2), (2, 1), (1, -1), (-1, -1), (3, -1)\}$, find the gradient boosting tree $F(x) = F_0(x) + 0.1 * h_1(x)$. Here, $F_0(x) = 0.5$. So you only need to determine $h_1(x)$. You need to use absolute loss.
a. What will be the labels that are used to determine $h_1(x)$. [1]
b. Determine $h_1(x)$. Use decision stump and choose to split at two different points, say at 0.5 and 2.5 to find the best split. [1]
c. Find the prediction of sample 3.5 using the gradient boosted tree. [.5]

Q4. Consider a fully connected NN with an input layer, single hidden layer and output layer. Input layer has 4 nodes, hidden layer has 2 nodes and output has 1 node. Assume there are no activations at any layer. All the 4 nodes are connected to both the hidden nodes. And the hidden nodes are connected to output node. Suppose a dropout is applied at both input and hidden layer. Dropout vector obtained from Bernoulli random process at input layer looks like $[1, 0, 0, 0]^\top$ (as input has 4 nodes and at hidden node it is $[0, 1]^\top$. Find the output of this network in terms of its inputs and weights. [1]