Quiz 3

● Graded

Student

Brian Bertness

Total Points

30 / 30 pts

Question 1

Q1

9 / 9 pts

✔ **+ 9 pts** Everything correct

- **+ 3 pts** A is NOT selected (A incorrect), because more parameters are more likely to lead to overfitting
- **+ 3 pts** B is selected (B correct), because that is how weights in MLP are updated by definition
- **+ 3 pts** C is selected (C correct), because perceptrons need the threshold function (which is an activation function) to solve logical operations (e.g., AND, XOR)
- **+ 0 pts** Unattempted/Incorrect

Question 2

Q2

11 / 11 pts

✔ **+ 11 pts** Everything correct

- **+ 3 pts** Step 1: Specify the correct objective function.
- **+ 2 pts** Step 2a: Specify the bias term to be updated
- **+ 2 pts** Step 2b: Specify the non-bias term to be updated
- **+ 2 pts** Step 3a: Use the correct update formula for the specific objective and parameter (bias term)
- **+ 2 pts** Step 3b: Use the correct update formula for the specific objective and parameter (non-bias term)
- **+ 2 pts** Partial credits for step 3: The update formula does not match the selected parameter/objective (grade for bias and non-bias term separately)
- **+ 2 pts** Partial credits for step 3: Incorrect update formula due to wrong derivation (e.g., incorrect calculation of derivatives, missing the summation notation) (grade for bias and non-bias term separately)
- **− 1 pt** Arithmetic error
- **+ 0 pts** Irrelevant answer/No attempt

Question 3

Q3

10 / 10 pts

✔ **+ 10 pts** Everything is correct

+ 4 pts Correct examples of activation functions (ReLU, sigmoid, tanh, softmax, etc.)

+ 1.5 pts Linearity correct for activation function I

+ 1.5 pts Linearity correct for activation function II

+ 3 pts Correct role of activation function in MLP. Expected answer should include one of: (1) introducing non-linearity to the function being modeled by the MLP (2) whether the neuron should fire an output/not (3) transforming original output values to 0-1 range or probability (normalize output).

+ 1.5 pts Inaccurate role of activation function but put some related process related to it (e.g., (1) calculating posterior of input to belong to a certain class (this only applies to last layer's activation function (2) calculating real value of node's ouput)

+ 0 pts Incorrect/no attempt

Question assigned to the following page: 1

Name+Email ID

CSCI 5521: Machine Learning Fundamentals (Spring 2024)

Quiz 3 (Mar. 28, 2024)

Due on Gradescope by 2:00 pm, Mar. 29

Instructions:

- This quiz has 3 questions, 30 points, on 1 page.
- Please **write your name & ID on this cover page.**

Brian Bertness
1478201

1. (**9 points**) Select all the correct statement(s).

   F (a) Adding more neurons to a single hidden layer of a Multilayer Perceptron always results in higher test accuracy. Should be False as this makes the model more complex and could very well cause overfitting.

   T (b) The weights in a Multilayer Perceptron are updated during the backpropagation phase based on an error function.

   T (c) A Multi-layer Perceptron with no activation functions $\boxed{\text{cannot}}$ solve the XOR problem if provided with the right weights and biases.

2. (**11 points**) Pick **one bias term** and **one non-bias term** in a (Multilayer) Perceptron **without** any activation functions, and write their update equations. Start with writing the objective function.

   See Sheet

3. (**10 points**) Give two examples of activation functions in Multilayer Perceptron (You can write either the name or the mathematical representation of the function). Answer whether the activation functions that you wrote down are linear or non-linear. And briefly describe the roles of activation functions in a Multilayer Perceptron.

   See Sheet

Question assigned to the following page: 2

2. Ok, with no activation function such as sigmoid or tanh we would simply have $z_h = w_h^T x$.

Thus, for a non-bias term with $E(W,v|X) = \frac{1}{2} \sum_t (r^t - y^t)^2$
Choosing w_{1j} as the non-bias term.

$$\Delta w_{1j} = -\eta \frac{\partial E}{\partial w_{1j}}$$

$$= -\eta \sum_t \underbrace{\frac{\partial E}{\partial y^t}}_{1} \underbrace{\frac{\partial y^t}{\partial z_1^t}}_{2} \underbrace{\frac{\partial z_1^t}{\partial w_{1j}}}_{3} \quad \star$$

Only part that is changing from in class derivations since $\frac{\partial E}{\partial y^t}$ and $\frac{\partial y^t}{\partial z_1^t}$ have not changed!

Now, $\overset{3}{\frac{\partial z_1^t}{\partial w_{1j}}} [w_{1j} x] = x_j^t$

From class we derived: $\overset{1}{\frac{\partial E}{\partial y^t}} = -(r^t - y^t)$ and $\overset{2}{\frac{\partial y^t}{\partial z_1^t}} = v_1$

Substituting 1, 2, 3 into equ $\star$ above we get

$$\boxed{\Delta w_{1j} = -\eta \sum_t -(r^t - y^t) v_1 x_j^t = \eta \sum_t (r^t - y^t) v_1 x_j^t}$$

For bias term Δw_{0j} we have the same except $\frac{\partial z_1^t}{\partial w_{0j}} [w_{0j}] = 1$

$$\boxed{\text{so } \Delta w_{0j} = -\eta \frac{\partial E}{\partial w_{0j}} = -\eta \sum_t \frac{\partial E}{\partial y^t} \frac{\partial y^t}{\partial z_0^t} \frac{\partial z_0^t}{\partial w_{0j}} = \eta \sum_t (r^t - y^t) v_0}$$

Question assigned to the following page: 3

3. Two examples of Activation Functions in Multi-Layer Perceptrons are the Sigmoid function and the tanh function.

Both of the functions are non-linear.

An activation function (threshold function) does, as Dr. Zhuo states "If the input is bigger than a certain threshold then it returns a 1, otherwise its a zero." The activation function determines if the neuron (processing unit) is activated.