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 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

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

Thus, for a <u>non-bias</u> term with $E(W, v|X) = \frac{1}{2}\sum_t (r^t - y^t)^2$

Choosing w_{1j} as the non-bias term.

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

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

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

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

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

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

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

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. Zhao 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.