

Assignment 2
CS 4783/5783
Due: 10/07/2022 11:59 pm

[Question 1]

[10 points]

Derive the update rule and show how to train a 2-layer (1 hidden layer and 1 output layer) neural network with backpropagation for regression using the Mean Square Error loss. Assume that you are using the Sigmoid activation function for the hidden layer. Explain briefly how this is different from the update rule for the network trained for binary classification using log loss.

[Question 2]

[40 points]

For the given data on canvas, construct a neural network for the regression task. Your network must have 1 hidden layer and 1 output layer. Use sigmoid to be your activation function for the hidden layer(s). You can choose the number of neurons in each layer using your intuition.

The data is already split to have your input data for training (`X_train.csv`) and testing (`X_test.csv`) and their corresponding target values `Y_train.csv` and `Y_test.csv`, respectively. You can load the data as follows:

```
X_train = np.loadtxt("X_train.csv")
```

Implement the backpropagation algorithm and train your network until convergence.

Answer the following questions:

1. What is the activation function that you will choose for the output layer? Justify your answer briefly.
2. How many neurons should there be in the output layer? Why?
3. Report the average MSE loss and the accuracy.
4. Plot the loss and accuracy as a function of the number of iterations.
5. What is the effect of the learning rate on the training process? Vary the learning rate to be between 0.001 and 1.0 and plot the resulting accuracy as a function of learning rate.
6. What is the effect of the number of neurons in the hidden layer? To answer this question, you will need to consider and answer the following:
 - a. You will need to vary the number of neurons from 1 to 10. Does the update rule need to be changed/derived again? Why or why not?
 - b. Report your observations by reporting the final loss and plotting the true labels and your predicted labels, along with a brief (2-3 lines) description.
7. What is the effect of the activation functions in the network? Explore two different activation functions other than sigmoid such as tanh, linear, or ReLU.
 - a. Will you need to change the update rule?
 - b. What is the change that you need to make to achieve this experiment?
 - c. Report your observations by reporting the final loss and plotting the true labels and your predicted labels, along with a brief (2-3 lines) description.