CSE 471 : Introduction to Artificial Intelligence
Project #6

Due Data : December 2 2018, 11:59 PM

Instruction

In this project, you will use TensorFlow to implement neural networks with different architectures. You will see how the performance of the network is affected by different parameters like the number of neurons in the layers, the number of layers in the network and the activation layer used in the network.

You will be using the skeleton code provided to you to work on this project. You need to install packages like numpy, tensorflow and matplotlib to successfully run the code.

Use the MNIST handwritten digit dataset for training and testing. You can download the data from http://yann.lecun.com/exdb/mnist/. The dataset contains 60,000 images of handwritten digits [0-9] for training and 10,000 images of the same for testing. Download all the four files from the link and place them in a folder named "mnist".

Note: All architectures given in the questions below specify the number of nodes in the hidden layers that you should add. For example (10,32,64) specifies 10 neurons in the first hidden layer, 32 neurons in the second hidden layer, and 64 neurons in the third hidden layer. You should use the **Relu** activation function unless specified otherwise.

Question 1 (10 Points)

For this question, you will analyze how changing the number of neurons affect the performance of the network, keeping the number of layers fixed. Try following network architectures and generate a plot for Epochs vs Training Loss and Epoch vs Testing Loss in the same plot for all the architectures. You can use solid lines for training loss and dashed lines for testing loss. You also need to submit a plot showing testing accuracy for each architecture.

1. (2048,1024,512,128,32)
2. (1024,512,128,64,32)
3. (512,256,128,32,16)
4. (256,128,32,16,8)
5. (128,64,32,16,8)

Question 2 (10 Points)

For this question, you will analyze how changing the number of layers affect the accuracy of the network. Try following architectures and submit the same two plots as the previous question.

1. (2048)
2. (1024,512)
3. (512,256,128)
4. (256,128,64,32)
5. (256,128,64,32,16)

Question 3 (10 Points)

For this section, you will see how an activation function affects the performance of the networks. You will use different activation functions for **hidden layers** with a fixed architecture and check the performance. The architecture will be (512,256,128,32,16). For this question, you will plot Training Loss and Testing Loss vs Epoch for all the activation functions.

1. Sigmoid
2. Tanh
3. Relu
4. Leaky Relu
5. ELU

Question 4 (5 Points)

for this part, you need to write a brief explanation of the findings for all the above questions.

Submission

You need submit a pdf along with the code in a zip file named with ASUid_lastname_firstname_p6.zip. The pdf should contain all the plots and the explanation. Plots should have proper labels, legend, and titles.