Experiments performed on fully connected neural network (Rose Exeus)

Team Members: Ivan Liuliaev, Conor Finnegan, Rose Exeus

Task 2 Analysis

1) using small values near zero (-0.01, 0.01)

---

gradient Impact: small initial weights create small activations throughout the network. This also leads to small gradients during backpropagation.

Expected Behavior: It will take many epochs for us to see performance in the graph. The training loss will also decrease at a relatively slow rate.

2) Using the Xavier initialization formula

This method is most suited for this network architecture as it sets the weights based on how many neurons are connected to each layer.

Gradient Impact:The gradients of this network follow a consistent magnitude during backpropagation.

Expected Behavior: As seen in the graphs, the training loss begins to decrease at a steady rate. Accuracy improves as well. There are no dramatic fluctuations in the graphs.

3) Uniform/ Too fast method

Gradient Impact: Using large intitial gradients causes dramatic weight updates during optimization.

Expected Behavior:Early on the training is very rapid and unstable. The network will also be oscilating instead of decreasing/growing at a steady rate.

Learning Rate Experiment

Small LR = 0.0001

Gradient Impact: Each parameter receives very small updates and the dropout rate further reduces the magnitude of each gradient.

Expected behavior: During the initial training, it will be hard to tell if there is any improvement with loss or accuracy. Training time is fast, but the model will have poor accuracy.

Medium LR = 0.01

Gradient Impact: since the updates to each weight are moderate, we can see progress in training without overshooting. This rate will also compensate for reduced learning from dropout.

Expected behavior: There will be a smooth training curve with decreasing loss over time.

Large LR = 1.0

Gradient Impact: Each weight is dramatically changed during each pass. We should see the weights oscillating as well.

Expected behavior: There will be large spikes in the graphs and loss will increase over time. The model seems like it is training quickly, but there is no improvement in performance.

Task 3 Analysis

I've applied the dropout probablility (0.5) to the forward pass of the ReLU activation layer.

During testing, all neurons are acive, but during training each neuron in the hidden layer has a probability of being dropped to 0.

I experimented with a high dropout rate of o.0 and this means only 10% of the neurons stay active during the forward pass.

Expected Behavior: Learning will be very slow and there should be high variance in training between each batch since the network has a limited capacity during each pass. It will take a relatively large number of epochs to see any good performance.

In the inital experiment, I used a dropout rate of 0.5 and this was more effective because it reduces graident magnitudes, and aids in the prevention of overfitting. a dropout rate of 0.5 means that only half of the neurons might be deactivated, and this allows the model to generalize more on different data instead of relying on only specific neurons.