Assignment 4

MSCI 641

27th June 2019

Question 1:

Report your classification accuracy results in a table with three different activation functions in the hidden layer (ReLU, sigmoid and tanh) What effect do activation functions have on your results?

Answer:

Activation Function	Accuracy on Test Set
tanh	**68.4%**
sigmoid	**66.28%**
Relu	**61.4%**

Looking at the results we see that tanh performs the best and both tanh and sigmoid perform better than relu because they both classify data between [0,1] & [-1,1] which means that they are good at classification problems. In our assignment we are predicting positive and negatives due to which we are getting better result by both tanh and sigmoid. Tanh is better than sigmoid because it has stronger n-order gradients.

Question 2:

What effect does addition of L2-norm regularization have on the results?

Answer:

We fit our models to search the solution space for the most fitting solution. When we are using any optimizer (sgd, adam, etc) we fit our network's parameters to the learning problem at hand, and at each iteration we take a step in the solution space towards the gradient of the loss function(optimizer). The method of learning could over fit our training data. This overfitting could result in significant generalization error and bad performance on unseen data.

The best definition is given by Ian Goodfellow

"Regularization is any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error." -Ian Goodfellow

So to avoid Generalization of the model we used l2 regularization so that training is not over fit and at training time the model accurately does punishment/penalization (L2-factor) to stop it from over fit

Question 3:

What effect does dropout have on the results?

Answer:

Drop out (with best activation tanh)	Accuracy on Test Set
0.2	**69.20%**
0.25	**70.83%**
0.3	**68.65%**
0.4	**68.54%**
0.5	**64.54%**

As you increase the dropout the accuracy at first increases but after the threshold of 0.3 the model is not being able to fit properly. Intuitively a higher drop our results in higher variance on some of the layers which degrades the training. Dropout just like regularization is that it reduces the models capacity, if you reduce the capacity too much you will get bad results as can be seen from 0.3 onwards. If you increase the drop out you increase overfitting. Another reason in our case it that cross-entropy would cause nan or 0 gradient are predicted y's are all nan or zero so when we have enough epocs, all weights suddenly become zero and hence why we witness drop in accuracy.