Deep Learning : Theory and Practice

Homework # 3
Due date: April 17, 2020

1. For a neural network model with 1 hidden layer, input size of 100, output size of 10 classes and hidden layer size of 100 (which makes the weight matrix squre) and without any bias, derive the backpropagation update rule with the constraint when the weight matrix $\mathbf{W}^1$ is symmetric. (**Points** 25)

2. Show that the expression for the gradient of the last layer activations with targets $\mathbf{t}$ (δ_k^L in the notation used in the class) is same the following two cases,

   (a) A neural network with MSE loss and linear output activation ($y_k = a_K^L$)

   (b) A neural network with cross entropy loss and the output activation is softmax ($\mathbf{y} = softmax(\mathbf{a}^L)$)

   (**Points** 25)

3. Using the IRIS data of assignment 2 (150 data points of four dimensions from 3 classes), make a split (randomly of training and test) of 100 data points for training and 50 for test. Note that both the training and test data must contain all the classes. Further divide the training split into 80 data points for neural network training and 20 points for validation (again making sure that all the classes are almost equally present in both these splits). Using the 80 training data points, implement the stochastic gradient descent algorithm by hand for neural network model with 2 hidden layers of 10 dimensions and output targets which are one-hot encoded. Use a softmax nonlinearity with a cross entropy error criterion. Perform mini-batches of size 16 with a learning rate of $1e-2$ and run the SGD for 20 epochs. Compute the training accuracy, validation accuracy and test accuray for each epoch of weight update (after running through 5 mini-batches). Plot these three accuracy values as a function of epoch.
   Note : Do not make use of any standard tools and implement the back propagation by hand. (**Points** 50)