Machine Learning

Assignment - 1

Due Date: 13th September, 2015

Programming Questions

All programming questions have to be performed on the Image Segmentation Dataset (http://archive.ics.uci.edu/ml/datasets/Image+Segmentation). The dataset is divided into training and testing sets. You are required to use the larger set for training (2100 samples) and the smaller set for testing (210 samples). That is, use interchanged sets as compared to the ones given on the website.

Q1. Train a Random Decision Forest on the training set and report classification accuracy on the test set for the given seven-class problem. Report the mean class-wise accuracy, as well as the absolute classification accuracy of the model (correct test classifications/total test samples). Write a loop to vary the number of trees from 5 to 105, by a step size of 10. Plot two separate graphs to show the accuracy variations with the changes in number of trees (one for mean class accuracy and one for absolute class accuracy). Report any patterns observed.

Q2. Neural Networks

A. Pick data pertaining to the class labels of “BRICKFACE” and “SKY” for the training and testing sets. Code a perceptron to perform this two class classification with the given features. Report results in the form of a confusion matrix.
B. Using the perceptron trained above, create a double hidden layer network with four nodes each. Use this architecture for the two-class problem of Q2. Does your result change? Report confusion matrix.
C. Using the entire dataset to train a two hidden layer neural network of your chosen size for the complete seven-class problem. Report the mean-class accuracy of the trained network (you may use an existing implementation of Neural Network for this and next question).
D. Increase the hidden layers to three and change the activation function on all layers. Does this affect the performance of the network? Report mean-class accuracy of the same.

Q3. Implement ID3 algorithm. Using the dataset provided, train the algorithm and use it for classification on the test set. Report the mean-class accuracy and absolute classification accuracy for the test set.