

Assignment #1
Machine Learning
Cross Validation, KNN & Minimum Distance
Submission Deadline: 27 June 2014

You are given a dataset related to medical diagnosis. The data consists of 233 samples, 6 features and two classes. **Use 3 fold cross validation** to do the followings

1. Normalize the given data such that all features have values between 0 to 1 after normalization.
2. Find **best value of K** for K nearest neighbor by changing value of K from 3 to 9 and choose that value which gives maximum averaged accuracy
3. Use minimum distance classifier and find averaged accuracy
4. Now instead of using minimum distance classifier, do following steps
 - a. Calculate mean feature vector for **class 1** using training data
 - b. Calculate distance of each test sample from mean feature vector
 - c. Classify test sample by applying a threshold value on calculated distance and compute sensitivity and specificity value
 - d. Repeat **step c** and compute sensitivity and specificity values for different threshold (0 to 1) and draw ROC curve for this

You submission should contain title page, code, ROC curve and all tables containing values of sensitivity, specificity and accuracy for all folds and also their averaged values.

Mail softcopy of your assignment (word or pdf) at machinelearningeme@gmail.com

Note: i. Plagiarism leads to zero marks for both parties

ii. No marks for late submissions