

CS4044 -Pattern Recognition Assignment

Problem I - Classification (15 Marks)

Note: For this assignment you may use any of machine learning softwares like WEKA, or Matlab toolbox

1. Choose a dataset(Preferably from UCL machine learning repository) with 3 attributes .

i) Apply Bayesian classification to the data (Divide the dataset into two, i.e. for training and testing)

ii) Apply Nearest neighbor classification with $k=7$. (Distance matrix: Euclidean)

ii) Apply k-means clustering with k as the number of required classes.

(K-means clustering is not covered in the class. The topic Clustering is given for self study and will be included for the topics in the End semester exam)

Problem II (5 Marks)

1 . Write a program to generate a one dimensional dataset with normal distribution as the probability density function with a given mean and standard deviation.

Probability Density Estimation :

2. Using Parzen window density estimation generate the PDF of the dataset generated in the previous question and plot the PDF.

Submission Details

Prepare a document with details of the dataset used, screen shots of the results,programs , screen shots of outputs and plots generated . Please don't forget to add your name and roll number in the document. Upload it in Moodle(CSED Eduserver: eduserver.cse.nitc.ac.in) on or before the date of submission. Plagiarism in any form will lead to zero marks for the assignment.

Submission deadline(On or before) : 02/12/2014 (Tuesday) 12:00 a.m