

PESIT Department of Computer Science and Engineering

Course: Data Mining
Semester: 2016 Spring (January – May)
Instructor: BNR (Dr. B. Narsing Rao)

Assignment: 09
Topic: Support Vector Machines and K-Means Clustering
Due by: Midnight on **Wednesday, March 16, 2016**
Method: Send the answers to the following questions in a single PDF file
Use standard naming conventions

Part A

Use the iris dataset and classify it using an SVM using default options (note that libsvm.jar must be in your classpath).

Question 1

- Show the confusion matrix.
- Which class is best identified using the SVM? Explain.
- Is there any effect of the choice of the type of kernel? Explain.
- Which kernel gives the lowest accuracy and is there any way to increase its accuracy?

Part B

Use the Simple K Means API to cluster the iris dataset into a number of clusters starting with 2 and going up to 5.

Question 2

- Report the results of the clustering in the following format:

No of clusters	SSW	SSB	SST	SSB/SST
2				
...				

- What are your observations based on this table?
- For each value of k show a table that shows a cross tabulation between the cluster number and class
- Based on the table in c) above, what is the optimum number of clusters? Explain.