

Answers

Question 1

- a) Show the confusion matrix.

Iris	Setosa	Versicolor	Virginica
Setosa	50	0	0
Versicolor	0	47	3
Virginica	0	2	48

- b) Which class is best identified using the SVM? Explain.

Setosa is best identified using the SVM. As shown in the confusion matrix above, all the instances belonging to class Setosa is correctly identified. Also, from the statistics shown in the table below, the row corresponding to Setosa, has ideal values.

TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area	Class
1	0	1	1	1	1	Setosa
0.94	0.02	0.959	0.94	0.949	0.96	Versicolor
0.96	0.03	0.941	0.96	0.95	0.965	Virginica
0.967	0.017	0.967	0.967	0.967	0.975	

- c) Is there any effect of the choice of the type of kernel? Explain.

Yes, there is an effect on the choice of the type of kernel. The upper bound and lower bound values is different for each function.

- d) Which kernel gives the lowest accuracy and is there any way to increase its accuracy?

The sigmoid function gives the lowest accuracy(6.6667 %).

Question 2

- a) Report –

No. of clusters	SSW	SSB	SST	SSB/SST
2	62.14	1039.36	1101.5	0.944
3	7.82	1242.4	1249.96	0.994
4	6.61	1292.32	1298.93	0.995
5	6.29	1300.56	1306.85	0.995

b) What are your observations based on this table?

As the number of clusters increase, sum of squared errors within a cluster decreases and the distance between the clusters increases.

c) Tabulation of Cluster number and class.

K=2

Instances	Class
100	Versicolor
50	Setosa

K=3

Instances	Class
50	Versicolor
50	Setosa
50	Virginica

K=4

Instances	Class
26	Versicolor
50	Setosa
50	Virginica
24	Versicolor

K=5

Instances	Class
19	Versicolor
50	Setosa
50	Virginica
12	Versicolor
19	Versicolor

d) Based on the table in c) above, what is the optimum number of clusters? Explain.

The optimum number of clusters is 3 as all the instances are equally divided in each cluster.