Assignment 4

18-798 IVM

Due: Sept. 26, Thursday, 2013

Total Score 120/100. You may complete all or select some.

1. (10%) Both Euclidean distance and Cosine Angle coefficient can measure the distance between feature vectors. However, there are some subtle differences in the two methods. Discuss the differences with numerical examples.

2. (10%) Given a set of human subjects A, B, and C, we have binary biometric features 1 through 5. Calculate the Jaccard Index matrix among A, B, and C.

Feature	**A**	**B**	**C**
1	1	1	1
2	1	1	
3			1
4			1
5	1	1	

3. (10%) Use Chain Code to convert the shapes of a car, truck, and van to numbers manually. Then classify the feature vectors with Levenshtein distances. Analyze the classification results with a confusion matrix.

4. (10%) Based on Problem 3, improve the algorithm so that it can handle objects in difference sizes.

5. (10%) Given five test samples and a Euclidean distance matrix as below, plot a dendrogram.

	1	**2**	**3**	**4**	**5**
1	0				
2	2	0			
3	6	5	0		
4	10	9	4	0	
5	9	8	5	3	0

6. (10%) Give a one-dimensional classification problem illustrated in Figure 3.11, define a Radial Basis Function so that the projected "+" points and "-" can be as far as possible. The coordinates for the left figure are (-3,0; -2,0; -1,0; 2,0; 3,0; 5,0; 6,0)

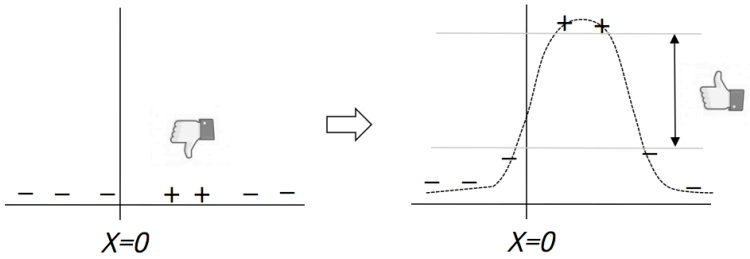

Figure 3.11 RBF Mapping

7. (10%) Given two training samples at X=0 and X=2, find the hyperplane to separate the two using the SVM method.

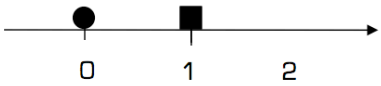

Figure 3.22 Support Vector Machine exercise problem.

8. (20%) Given a set of faces of both male and female, develop a gender recognition system so that it can tell "male" or "female" whenever a new face is entered. Use eigenspace approach to implement the algorithm. Hint: you may modify from a face recognition system, because gender recognition is a sub-problem of face recognition. Check out the MATLAB open source sample for face recognition.[1]

9. (10%) Assume you have a classification system that can tell the positive or negative results by tossing a coin. Consider two scenarios: tossing one coin and tossing two coins. Plot the ROC curves based on the sensitivity and specificity. Analyze the results in the report.

10. (20%) Use computing examples to show the trend of the classification performance (time and accuracy) for *k*-NN and SVM as the number of attributes (dimension of the feature space) increases to very large.

[1]http://www.mathworks.com/matlabcentral/fileexchange/38268-eigen-face-recognition/content/New%20folder/Mio.m