Fall 2013

Test 1, 10/16/13

ECE471/571 – Pattern Recognition

Name: ____________________

Total: 100

1. Simple questions and answers. (50 pts)

   a) (8) Give two examples of each of the following algorithms

      a) parametric learning
      b) non-parametric learning
      c) parameter estimation
      d) linear discriminant function

   b) (4) In practice, before we apply any pattern recognition algorithms, we usually need to normalize the data set by $\frac{x-\mu}{\sigma}$ where x is the data sample, μ is the mean vector and σ is the standard derivation of each feature vector. What's the purpose of normalization?

c) (8) When we use discriminant function as a classification method, we are actually treating the posterior probability as a function of the test sample x, g(x). With different levels of assumptions, we can simplify the original Baysian decision rule to a linear (in the format of ax+b) or a quadratic classifier (in the format of ax^2+bx+c). List the set of assumptions made in order to derive these two classifiers.

d) (6) (True or False). When k=1, the k-nearest-neighbor classification method is equivalent to the minimum distance method. Explain why or why not.

e) (8) Comment on the differences between Euclidean distance and Mahalanobis distance. Give an example showing under what scenario we'd use Euclidean distance and what condition for Mahalanobis distance.

f) (8) Plot the network structure for Perceptron. What are the limitations of Perceptron? And how can we overcome these limitations?

g) (8) Consider a multi-class classification problem with c classes. Suppose we are given n examples, labeled with one of the c classes. There are two schemes to apply a binary classification algorithm to a multi-class classification task. 1) One vs. All: For every label i, a classifier is learned over the following data set: the examples labeled with the label i are considered "positive", and examples labels with any other class are considered "negative". 2) All vs. All: For every pair of labels (i,j), a classifier is learned over the following data set: the examples labeled with one class i are considered "positive" and those with the other class j are considered "negative".

   a. How many classifiers do you learn in each algorithm?

   b. How many examples do you use to learn each classifier within the algorithm?

2. (30/20) Suppose we are given a training set and a test set as follows:

Training set

i	X1	X2	Label
1	2	10	1
2	4	4	1
3	5	7	1
4	8	9	-1
5	6	8	-1

Testing set

i	X1	X2	Label
1	5	6	1
2	6	18	-1

- Do kNN with k=1. What is the testing error?
- Do kNN with k=3. What is the testing error?
- If you only select x1 as feature, what is the testing error (k=1)?
- If you only use x2 as feature, what is the testing error (k=1)?
- Design a feature selection algorithm that can select the useful feature(s) for kNN without running kNN. Use the training set as a case study to justify your algorithm.

3. (+15/15) In a 2-D, 2-class problem, the classes may be represented by mean vectors $\mu_1 = [1 \quad 2]^T, \mu_2 = [3 \quad 4]^T$ and scatter matrices $S_1 = \frac{1}{10}\begin{bmatrix} 19 & -3 \\ -3 & 11 \end{bmatrix}, S_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$.

   - For class 2, sketch one of the isodensity contours.

     (Hint: the eigenvalues are 1, 3; and eigenvectors are [-0.7071 -0.7071] and [-0.7071 0.7071].)

   - Use Fisher's linear discriminant (FLD) to compute a projection vector. Sketch the vector.

   - Use FLD theory to classify point [3, 3]. That is, project the point onto the projection vector, and illustrate the region to which the point belongs.

4. (20/15 pts.) You are doing a test on disease diagnosis. If the patient has the disease, then he/she is classified as positive, otherwise, negative. Let class 1 be the actual negative cases with a uniform pdf on the interval [0,1]. Let class 2 be the actual positive case with a uniform pdf on the interval [0.95, 3.95]. Assume the prior probability is 0.7 (class 1) and 0.3 (class 2).

   - Given a decision threshold of 0.97, what is the probability for false-negative?
   - Write the equation for probability of error.