Hong Kong University of Science and Technology

COMP327/527: Pattern Recognition

Fall 1996

Midterm Examination

15 October 1996, 12:00–1:20pm

Student Name: ____________________

Student Number: ____________________

Instructions

1. This is an open-book, open-notes examination.
2. Check that you have all 8 pages (excluding this cover page).
3. Write your name and student number on this page.
4. Answer all questions in the space provided. Rough work should be done on the back pages.
5. Make your answers as concise as possible.

Question 1 (20%): __________

Question 2 (35%): __________

Question 3 (20%): __________

Question 4 (25%): __________

TOTAL (100%): __________

1. (20%)

   (a) [10%] Density estimation and discriminant functions represent two different approaches to the design of classifiers. Briefly discuss the major differences between these two approaches.

(b) [10%] Any classifier based on parametric density estimation can be converted into an equivalent classifier based on discriminant functions. On the other hand, not every classifier based on discriminant functions has an equivalent parametric classifier. Explain why.

2. (35%)

Consider the one-dimensional distributions for two classes, where class 1 is uniformly distributed in $[0, 2]$ and class 2 is uniformly distributed in $[1, 5]$. Assume that $Pr(\omega_1) = Pr(\omega_2) = 0.5$.

(a) [5%] Sketch the probability density functions for the two classes. Label the axes and mark the density function values appropriately.

(b) [5%] Using Bayesian (optimal) decision, what range of the x value will be classified to class 1 and what range will be classified to class 2?

(c) [5%] Using Bayesian (optimal) decision, what is the probability that a random example x generated from class 1 will be misclassified to class 2?

(d) [5%] Using Bayesian (optimal) decision, what is the probability that a random example x generated from class 2 will be misclassified to class 1?

(e) [5%] Compute the Bayes error (i.e. minimum average probability of error).

(f) [8%] Using the 1-NN classification rule under the large-sample assumption, compute the average probability of error.

(g) [2%] Does 1-NN achieve optimal decision for this problem?

3. (20%)

Consider the following distribution

$$p(x \mid \theta) = \frac{\theta^x}{x!} e^{-\theta}$$

where $x = 0, 1, 2, \ldots$ and $\theta > 0$ is the parameter of this distribution. Suppose we are given a sample S of n examples, $x^{(1)}, \ldots, x^{(n)}$, drawn independently from the underlying distribution. Using these n examples, find the maximum likelihood estimate for θ.

4. (25%)

Consider the following four training examples:

x_1	x_2	x_3	t
1	−1	3	1
1	0	3	−1
2	−1	2	−1
2	0	2	1

(a) [10%] Show that there does not exist a simple perceptron with three input units that can linearly separate the four examples into two classes correctly.

(b) [15%] Suppose x_3 of the second training example above is equal to 2 instead of 3. Show that the four examples become linearly separable. Give a set of perceptron weights that corresponds to one solution.

***** THE END *****