Hong Kong University of Science and Technology

COMP527: Pattern Recognition

Fall 2001

Midterm Examination

6 November 2001, 9:00–10:20am

Student Name: ______________________

Student Number: ______________________

Instructions

1. Check that you have all 6 pages (including this cover page).
2. Write your name and student number on this page.
3. Answer all questions in the space provided. Rough work should be done on the back pages.
4. Make your answers as concise as possible.

Question 1 (10%):	____________
Question 2 (10%):	____________
Question 3 (18%):	____________
Question 4 (22%):	____________
Question 5 (15%):	____________
Question 6 (25%):	____________
TOTAL (100%):	____________

1. (10%)

   "The objective of classifier training is to minimize the classification error of the training set." Is this statement correct? Explain why or why not.

2. (10%)

   "The major difference between MAP estimation and ML estimation is that the prior parameter distribution is incorporated in the former but not the latter." Is this statement correct? Explain why or why not.

3. (18%)

Consider a classification problem with three classes where $\lambda_{11} = 0$, $\lambda_{12} = 1$, $\lambda_{13} = 1$, $\lambda_{21} = 2$, $\lambda_{22} = 0$, $\lambda_{23} = 2$, $\lambda_{31} = 4$, $\lambda_{32} = 4$, $\lambda_{33} = 0$.

(a) Compute the conditional risks $R(\alpha_i|\mathbf{x}), 1 \leq i \leq 3$, in terms of the posterior probabilities.

(b) Give the minimum-risk classification rule in terms of the posterior probabilities. Use the reject option in case of a tie in the minimum risk.

4. (22%)

   (a) Discuss the advantages and disadvantages of the two approaches to density estimation.

   (b) When only limited training data are available for building a Bayesian classifier based on the nonparametric density estimation approach, the classification error may be rather large even though the parametric approach can give much smaller error. (You might have observed this phenomenon in the first assignment.) Explain why this could happen.

5. (15%)

Consider the following five training examples:

x_1	x_2	t
2	1	0
3	2	0
4	4	0
1	1	1
3	5	1

Prove that they are linearly separable by providing a linear discriminant function for the classification task.

6. (25%)

Consider a single-hidden-layer feedforward neural network with softmax output units for representing probabilities, i.e.

$$y_k = \frac{\exp(\sum_{j=0}^{h} w_{kj} y_j)}{\sum_{k'=1}^{c} \exp(\sum_{j=0}^{h} w_{k'j} y_j)}$$

with the same notational convention as in the notes.

Using gradient-descent as in the standard back-propagation learning algorithm, show that the per-epoch updating rule for the hidden-to-output weights can be expressed as follows:

$$\Delta w_{kj} = \eta \sum_{p=1}^{n} \delta_k^{(p)} y_j^{(p)} \quad \text{where} \quad \delta_k^{(p)} = (t_k^{(p)} - y_k^{(p)})(1 - y_k^{(p)}) y_k^{(p)}$$

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