## CSE555: Introduction to Pattern Recognition

## Midterm Exam

(100 points, Closed book/notes)

There are 5 questions in this exam.

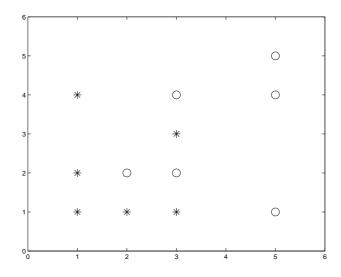
Page 3 is the Appendix that contains some useful formulas.

- 1. (15pts) Bayes Decision Theory.
  - (a) (5pts) Assume there are c classes  $w_1, \dots, w_c$ , and one feature vector  $\mathbf{x}$ , give the Bayes rule for classification in terms of a priori probabilities of the classes and class-conditional probability densities of  $\mathbf{x}$ .
  - (b) (10pts) Suppose we have a two-classes problem  $(A, \sim A)$ , with a single binary-valued feature  $(\mathbf{x}, \sim \mathbf{x})$ . Assume the prior probability P(A) = 0.33. Given the distribution of the samples as shown in the following table, use Bayes Rule to compute the values of posterior probabilities of classes.

	A	$\sim A$
x	248	167
$\sim {f x}$	82	503

- 2. (25pts) Fisher Linear Discriminant.
  - (a) (5pts) What is the Fisher linear discriminant method?
  - (b) Given the 2-d data for two classes:

$$\omega_1 = [(1,1), (1,2), (1,4), (2,1), (3,1), (3,3)]$$
 and  $\omega_2 = [(2,2), (3,2), (3,4), (5,1), (5,4), (5,5)]$  as shown in the figure:



- i. (10pts) Determine the optimal projection line in a single dimension.
- ii. (10pts) Show the mapping of the points to the line as well as the Bayes discriminant assuming a suitable distribution.

3. (20pts) Suppose  $p(x|w_1)$  and  $p(x|w_2)$  are defined as follows:

$$p(x|w_1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$
,  $\forall x$   
 $p(x|w_2) = \frac{1}{4}$ ,  $-2 < x < 2$ 

- (a) (7pts) Find the minimum error classification rule g(x) for this two-class problem, assuming  $P(w_1) = P(w_2) = 0.5$ .
- (b) (10pts) There is a prior probability of class 1, designated as  $\pi_1^*$ , so that if  $P(w_1) > \pi_1^*$ , the minimum error classification rule is to always decide  $w_1$  regardless of x. Find  $\pi_1^*$ .
- (c) (3pts) There is no  $\pi_2^*$  so that if  $P(w_2) > \pi_2^*$ , we would always decide  $w_2$ . Why not?
- 4. (20pts) Let samples be drawn by successive, independent selections of a state of nature  $w_i$  with unknown probability  $P(w_i)$ . Let  $z_{ik} = 1$  if the state of nature for the kth sample is  $w_i$  and  $z_{ik} = 0$  otherwise.
  - (a) (7pts) Show that

$$P(z_{i1}, \dots, z_{in}|P(w_i)) = \prod_{k=1}^{n} P(w_i)^{z_{ik}} (1 - P(w_i))^{1 - z_{ik}}$$

(b) (10pts) Given the equation above, show that the maximum likelihood estimate for  $P(w_i)$  is

$$\hat{P}(w_i) = \frac{1}{n} \sum_{k=1}^{n} z_{ik}$$

- (c) (3pts) Interpret the meaning of your result in words.
- 5. (20pts) Consider an HMM with an explicit absorber state  $w_0$  and unique null visible symbol  $v_0$  with the following transition probabilities  $a_{ij}$  and symbol probabilities  $b_{jk}$  (where the matrix indexes begin at 0):

$$a_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.5 & 0.1 \end{pmatrix} \qquad b_{jk} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0.4 & 0.6 \end{pmatrix}$$

- (a) (7pts) Give a graph representation of this Hidden Markov Model.
- (b) (10pts) Suppose the initial hidden state at t = 0 is  $w_1$ . Starting from t = 1, what is the probability it generates the particular sequence  $\mathbf{V}^3 = \{v_2, v_1, v_0\}$ ?
- (c) (3pts) Given the above sequence  $V^3$ , what is the most probable sequence of hidden states?

2

## Appendix: Useful formulas.

• For a  $2 \times 2$  matrix,

$$A = \left[ \begin{array}{cc} a & b \\ c & d \end{array} \right]$$

the matrix inverse is

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

 $\bullet$  The scatter matrices  $\mathbf{S}_i$  are defined as

$$\mathbf{S}_i = \sum_{\mathbf{X} \in D_i} (\mathbf{X} - \mathbf{m}_i) (\mathbf{X} - \mathbf{m}_i)^t$$

where  $\mathbf{m}_i$  is the d-dimensional sample mean.

The within-class scatter matrix is defined as

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$$

The between-class scatter matrix is defined as

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t$$

The solution for the W that optimizes  $J(\mathbf{W}) = \frac{\mathbf{W}^t \mathbf{S}_B \mathbf{W}}{\mathbf{W}^t \mathbf{S}_W \mathbf{W}}$  is

$$W = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$