ELE888/EE8209 Intelligent Systems (2009)
Mid-term Test Paper (A) - Winter 2009

- This is a closed-book, closed-notes test. The duration of the test is 2 hours (2.45pm - 4.45pm).
- Only calculators with no text storage facilities are allowed. Internet or computers should not be accessed during the duration of the exam.
- No questions related to the content of this question paper will be answered by the instructor administering the test. Any assumptions made should be clearly stated in your answer sheet.
- The total marks for the test is worth 30% of your total course marks. There are 3 questions and all questions carry equal marks.
- Absolutely no talking during the exam period and also no peeping into others answer sheets.

Question 1 (2 + 2 + 2 + 4 Marks)

(a) What is the purpose of feature extraction in a classification system? List three desirable properties of an extracted feature. Separable

(b) Explain the issue of "generalisation" in designing a classifier.

(c) What is the primary difference between the approaches taken for Bayesian Decision Theory and Linear Discriminant Analysis in designing a classifier?

P(ω)

(d) In a two-dimensional, two-class classification problem with equal priors, the conditional densities for each class are of the form: $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \Sigma)$ Given the following information, compute the correct class label for the sample $\mathbf{x}_1$. State all assumptions made.

$$\boldsymbol{\mu}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}, \mathbf{x}_1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \qquad d=2$$

$$g_i(\mathbf{x}) = -\frac{1}{2}\underbrace{(\mathbf{x}-\boldsymbol{\mu}_i)^T}_{[1\ -2]}\Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln \cancel{P(\omega_i)}$$

Question 2 (4 + 4 + 2 Marks)

A Nintendo Wii remote controller is used to measure two separate motion gestures to be used in a computer game (ω_1 = gesture 1, ω_2 = gesture 2). Each gesture is represented by a two-dimensional feature $\mathbf{x}_i = (dx_i, dy_i)$, denoting relative displacement in the +x and +y directions respectively. A training set of sample feature vectors (measured while performing the gestures) is given in the table below.

	$x_1 = dx_i$	$x_2 = dy_i$	*gesture*
$\mathbf{x}_1$	1	3	ω_1
$\mathbf{x}_2$	1	2	ω_1
$\mathbf{x}_3$	2	1.5	ω_1
$\mathbf{x}_4$	1	-1	ω_2
$\mathbf{x}_5$	-1	1	ω_2
$\mathbf{x}_6$	-2	0	ω_2
$\mathbf{x}_7$	-2	-2	ω_2

(a) Consider a decision boundary that passes through the origin, with $g(\mathbf{x}_1) = 4.5$ and $g(\mathbf{x}_7) = -5$. Calculate the weight vector $\mathbf{a}$ and sketch the decision boundary. Comment on the effectiveness of this boundary.

(b) Using the following sequence of training samples: $[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_4]$, compute four iterations of the single-sample perceptron algorithm to find the weight vector $\mathbf{a}$. Compute the new decision boundary. Based on your result, do you expect any further iterations of the algorithm to adjust $\mathbf{a}$? Explain. Additional information: $\mathbf{a}(0)^T = [\ -0.1 \quad -0.1 \quad -0.1\ ]$, $\eta(\cdot) = 1$

(c) What are the main differences between the single sample perceptron technique and the least mean squares technique (LMS) when estimating $\mathbf{a}$ using basic gradient descent?

Question 3 (2 + 3 + 2 + 2 + 1 Marks)

For a one-dimensional, two-class classification problem, the conditional densities are of the form:

$$p(x|\omega_i) = \left\{ \begin{array}{cc} 0 & x < 0 \\ \frac{2}{b_i}\left(1 - \frac{x}{b_i}\right) & 0 \le x \le b_i \\ 0 & otherwise \end{array} \right\} \quad (1)$$

(a) Assuming $b_1 > b_2$, sketch the conditional densities and find the likelihood function.

$p(\omega_1)$

(b) If the prior probabilities are unknown, state the conditions for minimum error-rate classification, find the decision boundary and associated decision rule.

(c) Find an expression for the probability of error in terms of an arbitrary decision boundary x^* and justify that the boundary found in (b) minimises this error.

(d) Find the probability of error for (b) given that $b_1 = 10$ and $b_2 = 3$.

(e) It is decided that classifying ω_1 incorrectly is more costly that classifying ω_2 incorrectly (i.e. $\lambda_{21} > \lambda_{12}$), describe and justify the expected effect on the decision boundary, of incorporating such a penalty into the design of the above classifier.

ELE888/EE8209 Intelligent Systems (Formula Sheet)

$$P(\omega_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_j) \cdot P(\omega_j)}{\sum_{k=1}^{c} p(\mathbf{x} \mid \omega_k) \cdot P(\omega_k)} \tag{1}$$

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \tag{2}$$

$$\boldsymbol{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \equiv \int (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T p(\mathbf{x}) d\mathbf{x} \tag{3}$$

$$R = \int R(\alpha(\mathbf{x}) \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x} \tag{4}$$

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x}) \tag{5}$$

$$J(\mathbf{w}) \equiv \frac{1}{2} \sum_{k=1}^{c} (t_k - z_k)^2 = \frac{1}{2} \|\mathbf{t} - \mathbf{z}\|^2 \tag{6}$$

$$\triangle \mathbf{w} = -\eta \frac{\partial J}{\partial \mathbf{w}} \tag{7}$$

$$\frac{\partial J}{\partial w_{kj}} = \frac{\partial J}{\partial net_k} \frac{\partial net_k}{\partial w_{kj}} \tag{8}$$

$$\frac{\partial J}{\partial w_{ji}} = \frac{\partial J}{\partial y_j} \frac{\partial y_j}{\partial w_{ji}} \tag{9}$$

$$J_e = \sum_{j=1}^{c} \sum_{\mathbf{x} \in D_j} \|\mathbf{x} - \boldsymbol{\mu}_j\|^2 \tag{10}$$

$$J_{fuzzy} = \sum_{j=1}^{c} \sum_{i=1}^{n} [P(\omega_j | \mathbf{x}_i, \theta)]^b \|\mathbf{x}_i - \boldsymbol{\mu}_j\|^2 \tag{11}$$

$$\hat{P}(\omega_j|\mathbf{x}_i,\theta) = \frac{(1/d_{ij})^{\frac{1}{b-1}}}{\sum_{j=1}^{c}(1/d_{ij})^{\frac{1}{b-1}}} \tag{12}$$

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{i=1}^{n}[P(\omega_j|\mathbf{x}_i,\theta)]^b \mathbf{x}_i}{\sum_{i=1}^{n}[P(\omega_j|\mathbf{x}_i,\theta)]^b} \tag{13}$$

$$\mathbf{w}_{ki}(t+1) = \mathbf{w}_{ki}(t) + \eta(t)\Lambda(|\mathbf{y}-\mathbf{y}^*|,t)\left[\mathbf{x}_i - \mathbf{w}_{ki}(t)\right] \tag{14}$$

$$\begin{aligned}
d_{min}(D_i,D_j) &= min_{\mathbf{x}\in D_i,\mathbf{x}'\in D_j}\|\mathbf{x}-\mathbf{x}'\| \\
d_{max}(D_i,D_j) &= max_{\mathbf{x}\in D_i,\mathbf{x}'\in D_j}\|\mathbf{x}-\mathbf{x}'\| \\
d_{avg}(D_i,D_j) &= \tfrac{1}{n_i n_j}\textstyle\sum_{\mathbf{x}\in D_i}\sum_{\mathbf{x}'\in D_j}\|\mathbf{x}-\mathbf{x}'\| \\
d_{mean}(D_i,D_j) &= \|\mathbf{m}_i - \mathbf{m}_j\|
\end{aligned} \tag{15}$$

$$\begin{aligned}
\mathbf{S}_j &= \textstyle\sum_{\mathbf{x}\in D_j}(\mathbf{x}-\mathbf{m}_j)(\mathbf{x}-\mathbf{m}_j)^T \\
\mathbf{S}_W &= \textstyle\sum_{j=1}^{c}\mathbf{S}_j \\
\mathbf{S}_B &= \textstyle\sum_{j=1}^{c} n_j(\mathbf{m}_j-\mathbf{m})(\mathbf{m}_j-\mathbf{m})^T
\end{aligned} \tag{16}$$

$$XB(c) = \frac{1}{n}\cdot\frac{\sum_{k=1}^{c}\sum_{i=1}^{n}P(\omega_k|\mathbf{x}_i,\theta)\|\mathbf{x}_i-\boldsymbol{\mu}_k\|}{min\|\boldsymbol{\mu}_p-\boldsymbol{\mu}_q\|} \tag{17}$$

$$CH(c) = \frac{tr(\mathbf{S}_B)}{tr(\mathbf{S}_W)}\cdot\frac{(n-c)}{(c-1)} \tag{18}$$

$$Bias_{jack} = (n-1)\cdot(\hat{\theta}_{(.)}-\hat{\theta}) \tag{19}$$

$$Var_{jack} = \frac{(n-1)}{n}\cdot\sum_{i=1}^{n}\left[\hat{\theta}_{(i)}-\hat{\theta}_{(.)}\right]^2 \tag{20}$$

$$Bias_{boot} = (\hat{\theta}^{*(.)}-\hat{\theta}) \tag{21}$$

$$Var_{boot} = (1/B)\cdot\sum_{b=1}^{B}\left[\hat{\theta}^{*(b)}-\hat{\theta}^{*(.)}\right]^2 \tag{22}$$