Machine Intelligence

TUTORIAL SHEET 3: Solutions

1. The Bayes minimum error decision rule assigns pattern vector $\mathbf{x}$ to class ω_j if

$$P(\omega_j)p(\mathbf{x}|\omega_j) = \max_{i=1}^{m} P(\omega_i)p(\mathbf{x}|\omega_i)$$

2. Bookwork

3. The discriminant function $f(\mathbf{x})$ developed in 2 is

$$f(\mathbf{x}) = (\mathbf{x}-\mu_1)^T\Sigma_1^{-1}(\mathbf{x}-\mu_1) + \log|\Sigma_1| - 2\log P(\omega_1) - (\mathbf{x}-\mu_2)^T\Sigma_2^{-1}(\mathbf{x}-\mu_2) - \log|\Sigma_2| + 2\log P(\omega_2) \leq 0$$

To check the class membership of $\mathbf{x}$ according to $f(\mathbf{x})$ we need:

$$|\Sigma_1| = 2 \qquad |\Sigma_2| = 5$$
$$\Sigma_1^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma_2^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$$
$$(\mathbf{x}-\mu_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad (\mathbf{x}-\mu_2) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Substituting into $f(\mathbf{x})$ we find

$$f(\mathbf{x}) = -0.121$$

Thus $\mathbf{x}$ should be assigned to class ω_1

4. (a) In the two class one-dimensional case the Bayes minimum error rule assigns pattern x to class ω_1 if

$$P(\omega_1)Ae^{ax} \leq P(\omega_2)Be^{b(x-1)}$$

Taking the logarithm of both sides we find

$$(a-b)x \leq \ln P(\omega_2)B - \ln P(\omega_1)A - b$$

and finally

$$x \leq \frac{\ln P(\omega_2)B - \ln P(\omega_1)A - b}{a-b}$$

(b) Constants A and B must be such that

$$\int_{\infty}^{\infty} Ae^{ax}dx = 1 \qquad \int_{\infty}^{\infty} Be^{b(x-1)}dx = 1$$

From this condition it follows

$$A = \frac{a}{e^{2a}-1} \qquad B = \frac{b}{e^{2b}-1}$$

(c)

$$A = \frac{1}{1-e^{-2}} = 1.15 \qquad B = \frac{0.5}{e-1} = 0.29$$
$$x_T = \frac{\log B - \log A - 0.5}{-1.5} = \frac{-1.23 - 0.14 - 0.5}{-1.5} = 1.24$$

(d) Before making the observation each class is equally probable, i.e. $P(\omega_i) = 0.5 \quad i = 1, 2$. After the pattern is observed we have

$$P(\omega_i|x) = \frac{P(\omega_i)p(x|\omega_i)}{p(x)}$$

where $p(x)$ is the mixture distribution given as

$$p(x) = P(\omega_1)p(x|\omega_1) + P(\omega_2)p(x|\omega_2)$$

For our equal prior case we have

$$P(\omega_i|x) = \frac{p(x|\omega_i)}{p(x|\omega_1) + p(x|\omega_2)}$$

Substituting $x = 1.5$ we have

$$P(\omega_1|x) = \frac{0.256}{0.256 + 0.328} = 0.438$$

and

$$P(\omega_2|x) = 1 - P(\omega_1|x) = 0.562$$

5. Given a set of N training patterns $X = \{\mathbf{x_1}, ..., \mathbf{x_n}\}$, the Nearest Neighbour rule assigns pattern $\mathbf{x}$ to class ω_j if the nearest point $\mathbf{x_k} \epsilon X$ to $\mathbf{x}$ belongs to class ω_j. The Nearest Neighbour rule is intuitive and can be used without having to estimate the class conditional probability density functions. This is particularly useful when the training data set is small and when no assumptions can be made about the form of the class distributions. However the Nearest Neighbour rule is computatationally intensive and its error rate is worse than the Bayes error rate.

6. Bookwork

7. (a)

$$P(\omega_i|\mathbf{x}) = \frac{P(\omega_i)p(\mathbf{x}|\omega_i)}{\sum_{j=1}^{2} P(\omega_j)p(\mathbf{x}|\omega_j)}$$

Hence

$$P(\omega_1|\mathbf{x}) = \begin{cases} \frac{2(1-x)}{2(1-x)+1} & 0 \le x \le 1 \\ 0 & elsewhere \end{cases}$$

$$P(\omega_2|\mathbf{x}) = \begin{cases} \frac{1}{2(1-x)+1} & 0 \le x \le 1 \\ 1 & 1 \le x \le 2 \\ 0 & elsewhere \end{cases}$$

(b) At $x = 0.6$ the aposteriori probability of class $P(\omega_2|\mathbf{x}) = 0.555 > P(\omega_1|\mathbf{x})$. Thus the pattern should be assigned to class ω_2.

(c) The error rate of the Nearest Neighbour rule can be determined by considering the probability of a pattern from one class finding a neighbour from another. The probability of such an event is given by

$$2P(\omega_1|\mathbf{x})P(\omega_2|\mathbf{x}) = 2\epsilon(\mathbf{x})(1 - \epsilon(\mathbf{x}))$$

where $\epsilon(\mathbf{x})$ is the local error incurred by the Bayes optimal rule. Thus the average error of the Nearest Neighbour rule is given by

$$e_N = \int 2\epsilon(\mathbf{x})(1 - \epsilon(\mathbf{x}))p(\mathbf{x})dx$$

Now the average Bayes error is given by

$$e_B = \int \epsilon(\mathbf{x})p(x)dx = \int_0^{0.5} P(\omega_2|\mathbf{x})p(\mathbf{x})d\mathbf{x} + \int_{0.5}^{1} P(\omega_1|\mathbf{x})p(\mathbf{x})d\mathbf{x} =$$

$$= \frac{1}{3}\int_0^{0.5} \frac{1}{2(1-x)+1}[2(1-x)+1]dx + \frac{1}{3}\int_{0.5}^{1} \frac{2(1-x)}{2(1-x)+1}[2(1-x)+1]dx = \frac{1}{4}$$

The second term

$$\int 2\epsilon^2(x)p(x)dx = \frac{1}{3}\int_0^{0.5} \frac{1}{2(1-x)+1}dx + \frac{1}{3}\int_{0.5}^{1}[\frac{2(1-x)}{2(1-x)+1}]^2[2(1-x)+1]dx$$

$$= 0.0329$$

Finally we have

$$e_N = 2e_B - 0.0329 = 0.4671$$

which shows that the recognition error will be almost twice the average Bayes error.