Algorithms & DATA - COMS20017

Tutorial #2 Questions & Solutions

(Q1) For the data displayed in the following table,

x	y
-2.4	0
-1.2	0
-0.2	0
0.9	1
2.1	1

fit a Gaussian distribution to each class, and compute the posterior probability that $x = 1.3$ is in class 1, given a prior of $P(y = 1) = 0.4$.

Solution:

To fit a Gaussian distribution, estimate μ and σ^2 for each of the two classes.

Class $y = 0$:

$$\begin{aligned}
\mu_0 &= \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{-2.4 - 1.2 - 0.2}{3} = -1.267 \\
\sigma_0^2 &= \frac{1}{3}\sum_{i=1}^{3}(x_i - \mu_0)^2 = \\
&= \frac{1}{3}[(-2.4 + 1.2667)^2 + (-1.2 + 1.2667)^2 + (-0.2 + 1.2667)^2] = 0.809
\end{aligned}$$

Class $y = 1$:

$$\begin{aligned}
\mu_1 &= \frac{0.9 + 2.1}{2} = 1.5 \\
\sigma_1^2 &= \frac{[(0.9 - 1.5)^2 + (2.1 - 1.5)^2]}{2} = 0.36
\end{aligned}$$

The conditional (posterior) pdf of $y = 1$ given $x = 1.3$ is:

$$p(y = 1|x = 1.3) = \frac{p(x = 1.3|y = 1)p(y = 1)}{p(x = 1.3)}$$

The Gaussian likelihood is given by:

$$p(x|y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_y)^2}{2\sigma_y^2}\right]$$

For $x = 1.3|y = 1$ we have:

$$p(x = 1.3|y = 1) = \frac{1}{0.36\sqrt{2\pi}} \exp\left[-\frac{(1.3-1.5)^2}{2\times 0.36}\right] = 0.629$$

Similarly, in preparation for computing the evidence (i.e. $p(x = 1.3)$), for $x = 1.3|y = 0$, one can get:

$$p(x = 1.3|y = 0) = 0.00755$$

and $p(y = 0) = 1 - p(y = 1) = 0.6$

Now the evidence (the denominator of the conditional pdf) can be calculated as:

$$\begin{aligned} p(x = 1.3) &= p(x = 1.3|y = 0)p(y = 0) + p(x = 1.3|y = 1)p(y = 1) = \\ &= 0.00755 \times 0.6 + 0.629 \times 0.4 = 0.25613 \end{aligned}$$

Substituting into the expression for the conditional pdf above, we get:

$$p(y = 1|x = 1.3) = \frac{0.629 \times 0.4}{0.25613} = 0.982$$

(Q2) Consider a two (equiprobable) class, one-dimensional problem with samples distributed according to the Laplace pdf in each class, that is,

$$p(x|\omega_i) = \frac{1}{2\sigma_i} \exp -\frac{|x-\mu_i|}{\sigma_i}$$

where the first Laplace distribution has location parameter (mean) $\mu_1 = -1$ and scale parameter $\sigma_1 = 1$, while the second has parameters $\mu_2 = 1$ and $\sigma_2 = 2$. Compute the threshold value, x_0, for minimum error probability classification.

Solution:

The threshold value is found by equating the posterior probabilities of the two classes, but using Bayes theorem and given that the two classes are equiprobable ($P(\omega_1) = P(\omega_2)$), we can instead compare the likelihood functions:

$$P(\mathbf{x}|\omega_1) = P(\mathbf{x}|\omega_2)$$

Substituting, we obtain:

$$\frac{1}{2} \exp\left(-|x+1|\right) = \frac{1}{4} \exp\left(-\frac{|x-1|}{2}\right)$$

Simplifying and taking the natural logarithm of both sides:

$$\ln 2 - |x+1| = -\frac{|x-1|}{2}$$

Case 1: $x \geq 1$

$$|x+1| = x+1,$$
$$|x-1| = x-1$$

Substituting into the threshold equation:

$$2\ln 2 - 2(x+1) = -x+1$$

$$x = 2\ln 2 - 3 \approx -1.614$$

So $x_{01} = -1.614$

Case 2: $-1 \leq x \leq 1$

$$|x+1| = x+1,$$
$$|x-1| = -(x-1) = -x+1$$

Substituting these into the equation:

$$2\ln 2 - 2(x+1) = x-1$$

$$x = \frac{2\ln 2 - 1}{3} \approx 0.129$$

So $x_{02} = 0.129$

Case 3: $x < -1$

$$|x+1| = -x-1,$$
$$|x-1| = x-1$$

Substituting into the threshold equation:

$$2\ln 2 + 2x + 2 = x-1$$

$$x = \frac{2\ln 2 + 3}{3} \approx 1.462$$

So $x_{03} = 1.462$

Of the three solutions found, only x_{02} is in the correct interval for the case analised, so the threshold for minimum error probability classification is:

$$x_0 = 0.129$$

(Q3) In a three-class, two-dimensional problem, the feature vectors in each class are normally distributed with covariance matrix:

$$\Sigma = \begin{pmatrix} 1.2 & 0.4 \\ 0.4 & 1.8 \end{pmatrix}$$

The mean vectors for each class are $\boldsymbol{\mu_1} = [0.1, 0.1]^T$, $\boldsymbol{\mu_2} = [2.1, 1.9]^T$, and $\boldsymbol{\mu_3} = [1.5, 2.0]^T$. Assuming that the three classes are equiprobable, i.e. $P(\omega_1) = P(\omega_2) = P(\omega_3)$, classify the feature vector $\mathbf{x} = [1.6, 1.5]^T$ according to the Bayes minimum error probability classifier.

Solution: The multivariate Gaussian probability density function is given by:

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

In a classification problem with M classes, $\omega_1, \omega_1, \ldots, \omega_M$, an unknown pattern, represented by the feature vector $\mathbf{x}$, is assigned to class ω_i if

$$P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}), \forall j \neq i$$

Alternatively, using Bayes theorem and the fact that the three classes are equiprobable, we can rewrite the above condition as:

$$P(\mathbf{x}|\omega_i) > P(\mathbf{x}|\omega_j), \forall j \neq i$$

In our case, $M = 3$ and we will thus evaluate $P(\mathbf{x}|\omega_i)$ for $i = 1, 2, 3$.

It is however easier to take the logarithm of the Gaussian pdf and evaluate:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu_i})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu_i}) - \frac{1}{2} \ln |\Sigma|$$

For information (but irrelevant for solving the problem) $g_i(\mathbf{x})$ is called the discriminant function.

Start by calculating Σ^{-1} and $|\Sigma|$:

$$\Sigma^{-1} = \begin{pmatrix} 0.9091 & -0.2020 \\ -0.2020 & 0.6061 \end{pmatrix}, \quad |\Sigma| = 2.08$$

Now, for class ω_1:

$$\mathbf{x} - \boldsymbol{\mu_1} = \begin{bmatrix} 1.6 \\ 1.5 \end{bmatrix} - \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1.4 \end{bmatrix}$$

$$(\mathbf{x} - \boldsymbol{\mu_1})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu_1}) = [1.5 \quad 1.4] \begin{bmatrix} 0.9091 & -0.2020 \\ -0.2020 & 0.6061 \end{bmatrix} \begin{bmatrix} 1.5 \\ 1.4 \end{bmatrix} =$$
$$= 2.7627$$

Substituting into g_1:

$$g_1(\mathbf{x}) = -\frac{1}{2}(2.7627) - \frac{1}{2}\ln(2.08) = -1.527$$

For class ω_2:

$$\mathbf{x} - \boldsymbol{\mu_2} = \begin{bmatrix}1.6\\1.5\end{bmatrix} - \begin{bmatrix}2.1\\1.9\end{bmatrix} = \begin{bmatrix}-0.5\\-0.4\end{bmatrix}$$

$$(\mathbf{x} - \boldsymbol{\mu_2})^T\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu_2}) = [-0.5 \quad -0.4]\begin{bmatrix}0.9091 & -0.2020\\-0.2020 & 0.6061\end{bmatrix}\begin{bmatrix}0.5\\-0.4\end{bmatrix} =$$
$$= 0.2945$$

Substituting into g_2:

$$g_2(\mathbf{x}) = -\frac{1}{2}(0.2945) - \frac{1}{2}\ln(2.08) = -0.467$$

For class ω_3:

$$\mathbf{x} - \boldsymbol{\mu_3} = \begin{bmatrix}1.6\\1.5\end{bmatrix} - \begin{bmatrix}1.5\\2.0\end{bmatrix} = \begin{bmatrix}0.1\\-0.5\end{bmatrix}$$

$$(\mathbf{x} - \boldsymbol{\mu_3})^T\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu_3}) = [0.1 \quad -0.5]\begin{bmatrix}0.9091 & -0.2020\\-0.2020 & 0.6061\end{bmatrix}\begin{bmatrix}0.1\\-0.5\end{bmatrix} =$$
$$= 0.2223$$

Substituting into g_3:

$$g_3(\mathbf{x}) = -\frac{1}{2}(0.2223) - \frac{1}{2}\ln(2.08) = -0.436$$

$g_3(\mathbf{x}) = -0.436$ turns out to be the largest g_i, so the feature vector $\mathbf{x}$ will be classified as class ω_3.

(Q4) For the data displayed in the following table,

x	y
-2.1	-4.2
-0.9	-2.3
0.2	-0.1
1.2	2.1
2.4	3.9

compute the least-squares parameter fit for a model of the form $\hat{y} = w_1 + w_2x$.

Solution:

$$\begin{aligned}
w_2 &= \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sum_{i=1}^{N} x_i^2 - N\overline{x}^2} = \\
&= \frac{-2.1 \times (-4.2) + (-0.9) \times (-2.3) + 0.2 \times (-0.1) + 1.2 \times 2.1 + 2.4 \times 3.9 - 5 \times 0.16 \times (-0.12)}{(-2.1)^2 + (-0.9)^2 + 0.2^2 + 1.2^2 + 2.4^2 - 5 \times 0.0256} = \\
&= 1.853 \\
w_1 &= \overline{y} - w_2\overline{x} = -0.12 + 1.853 \times 0.16 = -0.416
\end{aligned}$$

(Q5) For the data displayed in the following table,

x	y
-2.1	-4.2
-0.9	-2.3
0.2	-0.1
1.2	2.1
2.4	3.9

compute the least-squares parameter fit for a model of the form $\hat{y} = w_1 x + w_2 x^2$.

Solution:

In matrix form, the prediction problem can be written as:

$$\mathbf{y} = \mathbf{X}\mathbf{w}$$

where:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} -4.2 \\ -2.3 \\ -0.1 \\ 2.1 \\ 3.9 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_1 & x_1^2 \\ x_2 & x_2^2 \\ x_3 & x_3^2 \\ x_4 & x_4^2 \\ x_5 & x_5^2 \end{bmatrix} = \begin{bmatrix} -2.1 & 4.41 \\ -0.9 & 0.81 \\ 0.2 & 0.4 \\ 1.2 & 1.44 \\ 2.4 & 5.76 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

The least square solution in the non-linear case is given by:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Start by evaluating $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} -2.1 & -0.9 & 0.2 & 1.2 & 2.4 \\ 4.41 & 0.81 & 0.40 & 1.44 & 5.76 \end{bmatrix} \begin{bmatrix} -2.1 & 4.41 \\ -0.9 & 0.81 \\ 0.2 & 0.4 \\ 1.2 & 1.44 \\ 2.4 & 5.76 \end{bmatrix} = \begin{bmatrix} 12.46 & 5.642 \\ 5.642 & 55.5154 \end{bmatrix}$$

Its inverse:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{12.46 \times 55.5154 - 5.642 \times 5.642}\begin{bmatrix} 55.5154 & -5.642 \\ -5.642 & 12.46 \end{bmatrix} = \begin{bmatrix} 0.0841 & -0.0085 \\ -0.0085 & 0.0189 \end{bmatrix}$$

Next we compute $\mathbf{X}^T\mathbf{y}$:

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} -2.1 & -0.9 & 0.2 & 1.2 & 2.4 \\ 4.41 & 0.81 & 0.40 & 1.44 & 5.76 \end{bmatrix}\begin{bmatrix} -4.2 \\ -2.3 \\ -0.1 \\ 2.1 \\ 3.9 \end{bmatrix} = \begin{bmatrix} 22.750 \\ 5.0630 \end{bmatrix}$$

Finally we obtain the weight vector as:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 0.0841 & -0.0085 \\ -0.0085 & 0.0189 \end{bmatrix}\begin{bmatrix} 22.750 \\ 5.0630 \end{bmatrix} = \begin{bmatrix} 1.8706 \\ -0.0989 \end{bmatrix}$$

So $w_1 = 1.8706$ and $w_2 = -0.0989$