HW3

1. Assune that in a c-class classification problem, we have k features $X_1, X_2, \ldots, X_k$ that are independent conditioned on the class label and $X_j|\omega_i \sim \text{Gamma}(p_i, \lambda_j)$, i.e. $p_{X_j|\omega_i}(x_j|\omega_i) = \frac{1}{\Gamma(p_i)} \lambda_j^{p_i} x_j^{p_i-1} e^{-\lambda_j x_j}, p_i, \lambda_j > 0$. (30 pts)

   (a) Determine the Bayes' optimal classifier's decision rule making the general assumption that the prior probability of the classes are different.

(a) $\because P(w_i|X) \propto p(x|w_i)P(w_i)$

and features $X_1, X_2, \cdots, X_k$ are independent

$\therefore p(X|w_i) = p(x_1, x_2, \cdots x_k | w_i) = \prod_{j=1}^{k} p_{x_j|w_i}(x_j|w_i)$

$\therefore$ for each w_i, we should calculate:

$$p(X|w_i)P(w_i) = P(w_i) \prod_{j=1}^{k} \left(\frac{1}{\Gamma(p_i)} \lambda_j^{p_i} x_j^{p_i-1} e^{-\lambda_j x_j} \right)$$

for easy to calculate:

$$\log(P(w_i|X)) \propto \log(P(w_i)) + \sum_{j=1}^{k} \left[-\log(\Gamma(p_i)) + p_i \log(\lambda_j) + (p_i - 1)\log(x_j) - \lambda_j x_j \right]$$

   (b) When are the decision boundaries linear functions of $x_1, x_2, \ldots, x_k$?

(b) $\because$ in the decision boundaries

$P(w_a|X) = P(w_b|X)$

$\therefore$ according to (a)

$$\log(P(w_a)) + \sum_{j=1}^{k} \left[-\log(\Gamma(p_a)) + p_a \log(\lambda_j) + (p_a - 1)\log(x_j) - \cancel{\lambda_j x_j} \right]$$

$$= \log(P(w_b)) + \sum_{j=1}^{k} \left[-\log(\Gamma(p_b)) + p_b \log(\lambda_j) + (p_b - 1)\log(x_j) - \cancel{\lambda_j x_j} \right]$$

$$\therefore \log\left(\frac{P(w_a)}{P(w_b)}\right) + \sum_{j=1}^{k} \left[\log\left(\frac{\Gamma(p_b)}{\Gamma(p_a)}\right) + (p_a - p_b)\log(\lambda_j) \right] + \sum_{j=1}^{k} \left[(p_a - p_b)\log(x_j) \right] = 0$$

$\therefore$ only when $p_a = p_b$, the parameter of $\log(x_j) = 0$

but in this case, the entire expression does not contain x_j

the resulting difference will be a constant $\neq 0$

make it an invalid decision boundary

Cannot be eliminate

$\therefore$ only when λ_j take different values aross different class w

$$\log\left(\frac{P(w_a)}{P(w_b)}\right) + \sum_{j=1}^{k} \left[\log\left(\frac{\Gamma(p_b)}{\Gamma(p_a)}\right) + (p_a - p_b)\log(\lambda_j) \right] + \sum_{j=1}^{k} \left[(p_a - p_b)\log(x_j) - \underline{(\lambda_{ja} - \lambda_{jb}) x_j} \right] = 0$$

In this case, if $p_a = p_b$, the decision boundary will be a linear equation

(c) Assuming that $p_1 = 4, p_2 = 2, c = 2, k = 4, \lambda_1 = \lambda_3 = 1, \lambda_2 = \lambda_4 = 2$, and that the prior probabilites of each class are equal, classify $\mathbf{x} = (0.1, 0.2, 0.3, 4)$.

(c) $\because$ prior probabilities are equal

$\therefore P(w_i|x) \propto p_{x_j|w_i}(x_j|w_i)$

$\therefore$ for w_1: $\sum_{j=1}^{4} [-\log(\Gamma(4)) + 4\log(\lambda_j) + 3\log(x_j) - \lambda_j x_j] = -21.61$

for w_2: $\sum_{j=1}^{4} [-\log(\Gamma(2)) + 2\log(\lambda_j) + \log(x_j) - \lambda_j x_j] = -9.76$

$\because -21.61 < -9.75$

$\therefore$ classify x as w_2

(d) Assuming that $p_1 = 3.2, p_2 = 8, c = 2, k = 1, \lambda_1 = 1$, and that the prior probabilites of each class are equal, find the decision boundary $x = x^*$. Also, find the probability of type-1 and type-2 errors.

(d) $\because c=2$ and prior probabilities of each class are equal

$\therefore P(w_1) = P(w_2) = 0.5$

$\therefore$ according to (b)

$-\log(\Gamma(3.2)) + 3.2\log(1) + 2.2\log(x) = -\log(\Gamma(8)) + 8\log(1) + 7\log(x)$

$\therefore 4.8\log(x^*) = \log(\Gamma(8)) - \log(\Gamma(3.2)) - 4.8\log(1)$

$\therefore x^* = 4.91$

$\therefore P(\text{type } 1) = P(\text{classify as } w_2 \mid \text{true class is } w_1)$

$= P(X > x^* | w_1)$

$= 1 - P(X \le x^* | w_1)$

$= 1 - \int_{-\infty}^{4.91} \frac{1}{\Gamma(3.2)} x^{2.2} e^{-x} dx$

$= 1 - \frac{\gamma(3.2, 4.91)}{\Gamma(3.2)}$

$= 0.1572$

$P(\text{type } 2) = P(\text{classify as } w_1 \mid \text{true class is } w_2)$

$= P(X < x^* | w_2)$

$= \frac{\gamma(8, 4.91)}{\Gamma(8)}$

$= 0.1243$

(e) Assuming that $p_1 = p_2 = 4, c = 2, k = 2, \lambda_1 = 8, \lambda_2 = 0.3$, and $P(\omega_1) = 1/4, P(\omega_2) = 3/4$, find the decision boundary $f(x_1, x_2) = 0$.

(e) according to (b)

$$\log\left(\frac{P(w_1)}{P(w_2)}\right) + \sum_{j=1}^{k}\left[\underbrace{\cancel{\log\left(\frac{\Gamma(p_1)}{\Gamma(p_2)}\right)}}_{=0} + \underbrace{\cancel{\log\left(\frac{\lambda_j p_1}{\lambda_j p_2}\right)}}_{=0}\right] + \sum_{j=1}^{k}\left[\underbrace{\cancel{(p_1-p_2)}}_{=0}\log(x_j)\right] = 0$$

$\because p_1 = p_2 = 4$

$\therefore \log\left(\frac{1/4}{3/4}\right) = \log\left(\frac{1}{3}\right) \neq 0$ contradiction

2. Assume that in a c-class classification problem, there are k conditionally independent features and $X_i|\omega_j \sim \text{Lap}(m_{ij}, \lambda_i)$, i.e. $p_{X_i|\omega_j}(x_i|\omega_j) = \frac{\lambda_i}{2}e^{-\lambda_i|x_i - m_{ij}|}, \lambda_i > 0, i \in \{1, 2, \ldots, k\}, j \in \{1, 2, \ldots, c\}$. Assuming that the prior class probabilities are equal, show that the minimum error rate classifier is also a minimum weighted Manhattan distance (or weighted $\mathscr{L}_1$-distance) classifier. When does the minimum error rate classifier becomes the minimum Manhattan distance classifier? (15 pts)

① In MAP classifier:

Decide w_i if $P(w_i|x) > P(w_j|x)$ $\forall j \neq i$

$\because P(w_i|X) \propto p(X|w_i)P(w_i)$

and the prior class probabilities are equal

$\therefore P(w_i|X) \propto p(X|w_i)$

$\because$ features are independent

$\therefore p(X|w_i) = p(x_1, x_2, \cdots, x_k|w_i) = \prod_{i=1}^{k} \frac{1}{2}\lambda_i e^{-\lambda_i|x_i - m_{ij}|}$

$\therefore \log P(w_i|x) \propto \log p(X, w_i) = \sum_{i=1}^{k}\left[\log\frac{\lambda_i}{2} - \lambda_i|x_i - m_{ij}|\right]$

$$= -\sum_{j=1}^{k}\lambda_i|x_i - m_{ij}| + \underbrace{\sum_{i=1}^{k}\log\frac{\lambda_i}{2}}_{\text{constant}}$$

$$\therefore \arg\max_j \log P(w_i|x) = \arg\max_j\left(-\sum_{i=1}^{k}\lambda_i|x_i - m_{ij}|\right)$$

$$= \arg\min_j \sum_{i=1}^{k}\lambda_i|x_i - m_{ij}|$$

we can see λ_i as the weight of x_i

so the minimum error rate classifier is a mininum weighted $\mathscr{L}_1$-distance classifier

② when $\lambda_1 = \lambda_2 = \cdots = \lambda_k$

$$\arg\max_j \log P(w_i|x) = \arg\min_j \lambda\sum_{i=1}^{k}|x_i - m_{ij}|$$

$$= \arg\min_j \sum_{i=1}^{k}|x_i - m_{ij}|$$

the minimum error rate classifier becomes the minimum Manhattan distance classifier

3. The class-conditional density functions of a discrete random variable X for four pattern classes are shown below: (20 pts)

x	$p(x\|\omega_1)$	$p(x\|\omega_2)$	$p(x\|\omega_3)$	$p(x\|\omega_4)$
1	1/3	1/2	1/6	2/5
2	1/3	1/4	1/3	2/5
3	1/3	1/4	1/2	1/5

The loss function $\lambda(\alpha_i|\omega_j)$ is summarized in the following table, where action α_i means decide pattern class ω_i:

	ω_1	ω_2	ω_3	ω_4
α_1	0	2	3	4
α_2	1	0	1	8
α_3	3	2	0	2
α_4	5	3	1	0

Assume $P(\omega_1) = 1/10, P(\omega_2) = 1/5, P(\omega_3) = 1/2, P(\omega_4) = 1/5.$

(a) Compute the conditional risk for each action as:
$R(\alpha_i|x) = \sum_{j=1}^{4} \lambda(\alpha_i|\omega_j)p(\omega_j|x)$

(a) when $x=1$: $p(x_1) = \sum_{j=1}^{4} p(x|w_j)P(w_j)$

$$= p(x|w_1)P(w_1) + p(x|w_2)P(w_2) + p(x|w_3)P(w_3) + p(x|w_4)P(w_4)$$

$$= \frac{1}{3}\times\frac{1}{10} + \frac{1}{2}\times\frac{1}{5} + \frac{1}{6}\times\frac{1}{2} + \frac{2}{5}\times\frac{1}{5}$$

$$= 89/300$$

$$R(\alpha_1|x_1) = \sum_{j=1}^{4} \lambda(\alpha_1|w_j)p(w_j|x_1)$$

$$= \sum_{j=1}^{4} \lambda(\alpha_1|w_j)\frac{p(x_1|w_j)P(w_j)}{p(x_1)}$$

$$= \frac{1}{p(x_1)}[\lambda(\alpha_1|w_1)p(x|w_1)P(w_1) + \lambda(\alpha_1|w_2)p(x|w_2)P(w_2) + \lambda(\alpha_1|w_3)p(x|w_3)P(w_3) + \lambda(\alpha_1|w_4)p(x|w_4)P(w_4)]$$

$$= \frac{300}{89}(0\times\frac{1}{3}\times\frac{1}{10} + 2\times\frac{1}{2}\times\frac{1}{5} + 3\times\frac{1}{6}\times\frac{1}{2} + 4\times\frac{2}{5}\times\frac{1}{5})$$

$$= \frac{300}{89}\times\frac{77}{100}$$

$$= 2.596$$

Similarly: $R(\alpha_2|x_1) = \frac{300}{89}(1\times\frac{1}{3}\times\frac{1}{10} + 0 + 1\times\frac{1}{6}\times\frac{1}{2} + 8\times\frac{2}{5}\times\frac{1}{5}) = \frac{300}{89}\times\frac{227}{300} = 2.551$

$$R(\alpha_3|x_1) = \frac{300}{89}(3\times\frac{1}{3}\times\frac{1}{10} + 2\times\frac{1}{2}\times\frac{1}{5} + 0 + 2\times\frac{2}{5}\times\frac{1}{5}) = \frac{300}{89}\times\frac{23}{50} = 1.551$$

$$R(\alpha_4|x_1) = \frac{300}{89}(5\times\frac{1}{3}\times\frac{1}{10} + 3\times\frac{1}{2}\times\frac{1}{5} + 1\times\frac{1}{6}\times\frac{1}{2} + 0) = \frac{300}{89}\times\frac{11}{20} = 1.854$$

Similarly:

when $x=2$: $p(x_2)=\sum_{j=1}^{4} p(x|w_j)P(w_j) = 33/100$

$$R(\alpha_1|x_2) = \frac{100}{33}\left(0\times\frac{1}{3}\times\frac{1}{10}+2\times\frac{1}{4}\times\frac{1}{5}+3\times\frac{1}{3}\times\frac{1}{2}+4\times\frac{2}{5}\times\frac{1}{5}\right) = \frac{100}{33}\times\frac{23}{25} = 2.788$$

$$R(\alpha_2|x_2) = \frac{100}{33}\left(1\times\frac{1}{3}\times\frac{1}{10}+0\times\frac{1}{4}\times\frac{1}{5}+1\times\frac{1}{3}\times\frac{1}{2}+8\times\frac{2}{5}\times\frac{1}{5}\right) = \frac{100}{33}\times\frac{21}{25} = 2.545$$

$$R(\alpha_3|x_2) = \frac{100}{33}\left(3\times\frac{1}{3}\times\frac{1}{10}+2\times\frac{1}{4}\times\frac{1}{5}+0\times\frac{1}{3}\times\frac{1}{2}+2\times\frac{2}{5}\times\frac{1}{5}\right) = \frac{100}{33}\times\frac{9}{25} = 1.091$$

$$R(\alpha_4|x_2) = \frac{100}{33}\left(5\times\frac{1}{3}\times\frac{1}{10}+3\times\frac{1}{4}\times\frac{1}{5}+1\times\frac{1}{3}\times\frac{1}{2}+0\times\frac{2}{5}\times\frac{1}{5}\right) = \frac{100}{33}\times\frac{29}{60} = 1.465$$

when $x=3$: $p(x_3)=\sum_{j=1}^{4} p(x|w_j)P(w_j) = 28/75$

$$R(\alpha_1|x_3) = \frac{75}{28}\left(0\times\frac{1}{3}\times\frac{1}{10}+2\times\frac{1}{4}\times\frac{1}{5}+3\times\frac{1}{2}\times\frac{1}{2}+4\times\frac{1}{5}\times\frac{1}{5}\right) = \frac{75}{28}\times\frac{101}{100} = 2.705$$

$$R(\alpha_2|x_3) = \frac{75}{28}\left(1\times\frac{1}{3}\times\frac{1}{10}+0\times\frac{1}{4}\times\frac{1}{5}+1\times\frac{1}{2}\times\frac{1}{2}+8\times\frac{1}{5}\times\frac{1}{5}\right) = \frac{75}{28}\times\frac{181}{300} = 1.616$$

$$R(\alpha_3|x_3) = \frac{75}{28}\left(3\times\frac{1}{3}\times\frac{1}{10}+2\times\frac{1}{4}\times\frac{1}{5}+0\times\frac{1}{2}\times\frac{1}{2}+2\times\frac{1}{5}\times\frac{1}{5}\right) = \frac{75}{28}\times\frac{7}{25} = 0.75$$

$$R(\alpha_4|x_3) = \frac{75}{28}\left(5\times\frac{1}{3}\times\frac{1}{10}+3\times\frac{1}{4}\times\frac{1}{5}+1\times\frac{1}{2}\times\frac{1}{2}+0\times\frac{1}{5}\times\frac{1}{5}\right) = \frac{75}{28}\times\frac{17}{30} = 1.518$$

(b) Compute the overall risk R as:
$R = \sum_{i=1}^{3} R(\alpha(x_i)|x_i)p(x_i)$
where $\alpha(x_i)$ is the decision rule minimizing the conditional risk for x_i.

(b) $\because R(\alpha(x_i)|x_i) = \min_i R(\alpha_i|x_i)$

$\therefore R(\alpha(x_1)|x_1) = R(\alpha_3|x_1) = 1.551$

$R(\alpha(x_2)|x_2) = R(\alpha_3|x_2) = 1.091$

$R(\alpha(x_3)|x_3) = R(\alpha_3|x_3) = 0.75$

$\therefore R = \sum_{i=1}^{3} R(\alpha(x_i)|x_i)p(x_i)$

$= R(\alpha_3|x_1)p(x_1) + R(\alpha_3|x_2)p(x_2) + R(\alpha_3|x_3)p(x_3)$

$= 1.551\times\frac{89}{300} + 1.091\times\frac{33}{100} + 0.75\times\frac{28}{75}$

≈ 1.10016