# 第一题

与行为 $\alpha_i$ 相关联的损失为:

$$R(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\boldsymbol{x}) \tag{1}$$

$R(\alpha_i|\boldsymbol{x})$ 是和行为 $\alpha_i$ 有关的条件风险，行为由决策结果指定，则总风险由

$$R = \int R(\alpha(\boldsymbol{x})|\boldsymbol{x}) p(x)\, dx \tag{2}$$

给出，其中 $\alpha(\boldsymbol{x})$ 定义了通过每种可能的观测应该采取的行为。显然，当我们选择 $\alpha(\boldsymbol{x})$ 使得 $R(\alpha(\boldsymbol{x}))$ 对于每一个 $\boldsymbol{x}$ 尽量小，则总风险将被最小化。

根据题目，误差函数定义为:

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0, & i = j \quad i, j = 1, \ldots, c \\ \lambda_r, & i = c + 1 \\ \lambda_s, & otherwise \end{cases} \tag{3}$$

将 (3) 式代入 (1) 式得:

$$R_i(x) = \begin{cases} \lambda_s[1 - P(\omega_j|\boldsymbol{x})], & i = 1, \ldots, c \\ \lambda_r, & reject \end{cases} \tag{4}$$

因此当拒识风险大于判断风险即 $\lambda_r > \lambda_s[1 - P(\omega_j|\boldsymbol{x})]$ 时，我们值得做出判定。此时，$i = 1, \ldots, c$，为了最小化平均误差概率我么需要选取 $i$ 使得后验概率 $P(\omega_j|\boldsymbol{x})$ 最大，即

$$arg\,\underset{i}{min} = \begin{cases} arg\,\underset{i}{max}\, P(\omega_j|\boldsymbol{x}), & if\ \ \underset{i}{max}\, P(\omega_j|\boldsymbol{x}) > 1 - \lambda_r/\lambda_s \\ reject, & otherwise \end{cases} \tag{5}$$

证毕。当拒识代价 $\lambda_r = 0$且判断风险$\lambda_s > 0$ 时，拒识风险永远不会大于判断风险，永远会拒识。当 $\lambda_r > \lambda_s$ 时，拒识代价永远大于判定，不会发生拒识的现象。

## 第二题

(a)

条件概率密度的分布函数为：

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu_i}{\sigma})^2} \tag{6}$$

最小化误差概率条件下的贝叶斯决策规则为：

$$\text{如果} P(\omega_1|x) > P(\omega_2|x), \text{判别为}\omega_1, \text{否则判别为}\omega_2 \tag{7}$$

我们只考虑二分类问题，假设分类器以某种方式将空间分为两个区域 $R_1$ 和 $R_2$。错误以两种可能的形式出现：真实类别为 $\omega_1$ 而观测值落入 $R_2$ 或真实类别为 $\omega_2$ 而观测值落入 $R_1$ 所以误差概率为：

$$\begin{aligned} P_e &= P(x \in R_2, \omega_1) + P(\boldsymbol{x} \in R_1, \omega_2) \\ &= P(x \in R_2|\omega_1)P(\omega_1) + P(\boldsymbol{x} \in R_1, \omega_2)P(\omega_2) \\ &= \int\limits_{R_2} p(\boldsymbol{x}|\omega_1)P(\omega_1)dx + \int\limits_{R_1} p(\boldsymbol{x}|\omega_2)P(\omega_2)dx \end{aligned} \tag{8}$$

因为 $P(\omega_1) = P(\omega_2)$，根据贝叶斯公式及式 (7)，可得最小误差分类的判决边界为过两个高斯分布的中点且垂直于 x 轴的直线。所以误

差为（不失一般性，我们假设 $\mu_1 \leq \mu_2$）：

$$\begin{aligned}
P_e &= \frac{1}{2\sqrt{2\pi}\sigma}[\int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} e^{-\frac{1}{2}(\frac{x-\mu_2}{\sigma})^2}dx + \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma})^2}dx] \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma})^2}dx \quad \text{换元可得：} \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{\mu_2-\mu_1}{2\sigma}}^{+\infty} e^{-\frac{1}{2}\mu^2}dx
\end{aligned} \tag{9}$$

证毕。(因为证明过程假设 $\mu_1 \leq \mu_2$，所以免于使用绝对值符号。)

(b)

已知

$$P_e \leq \frac{1}{\sqrt{2\pi}a}e^{-\frac{a^2}{2}} \tag{10}$$

当 a 趋于无穷大时，式 (10) 中分母趋于无穷大，分子趋于 0，所以 $P_e$ 趋于 0。

## 第三题

(a)

$d$ 维多元正态密度函数的形式为：

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] \tag{11}$$

其中 $\boldsymbol{x}$ 是一个 $d$ 维列向量，$\boldsymbol{\mu}$ 是 $\boldsymbol{x}$ 的 $d$ 维均指向量，$\boldsymbol{\Sigma}$ 是 $\boldsymbol{x}$ 的 $d \times d$ 维协方差矩阵。

(b)

当每一类的协方差矩阵不相等时，判别函数为：

$$g_i(\boldsymbol{x}) = \boldsymbol{x}^t W_i \boldsymbol{x} + \boldsymbol{w}_i^t \boldsymbol{x} + \omega_{i0} \tag{12}$$

其中

$$W_i = -\frac{1}{2}\boldsymbol{\Sigma_i}^{-1}$$

$$\boldsymbol{w}_i = \boldsymbol{\Sigma_i}^{-1}\boldsymbol{\mu_i}$$

$$\omega_{i0} = -\frac{1}{2}\boldsymbol{\mu_i^t}\boldsymbol{\Sigma_i}^{-1}\boldsymbol{\mu_i} - \frac{1}{2}\ln|\boldsymbol{\Sigma_i}| + \ln P(\omega_i)$$

当每一类的协方差矩阵都相同时（设均为 $\boldsymbol{\Sigma}$），判别函数为：

$$g_i(\boldsymbol{x}) = \boldsymbol{w_i^t}\boldsymbol{x} + \omega_{i0} \tag{13}$$

其中

$$\boldsymbol{w_i} = \Sigma^{-1}\boldsymbol{\mu_i}\mu_i$$

$$\omega_{i0} = -\frac{1}{2}\boldsymbol{\mu_i^t}\Sigma^{-1}\boldsymbol{\mu_i} + \ln P(\omega_i)$$

(c)

当协方差矩阵奇异时，协方差矩阵的逆不存在且行列式的值为 0，不可计算。说明特征有冗余信息，可以通过 PCA、SVD 分解降维，使用降维后的协方差矩阵。

## 第四题

令两个判别函数式 (13) 相等，得到的边界面方程为：

$$\boldsymbol{w^t}(\boldsymbol{x} - \boldsymbol{x_0}) = 0 \tag{14}$$

其中

$$\boldsymbol{w} = \boldsymbol{\Sigma^{-1}}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \tag{15}$$

$$\boldsymbol{x}_0 = -\frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\ln[P(\omega_i)/P(\omega_j)]}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^t\boldsymbol{\Sigma^{-1}}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \tag{16}$$

欲使得判定面不过 $\boldsymbol{\mu_1}$ 与 $\boldsymbol{\mu_2}$ 之间的区域，临界条件为零点是 $\boldsymbol{\mu_1}$ 或

$\boldsymbol{\mu_2}$ 两点。若零点不在 $\boldsymbol{\mu_1}$ 和 $\boldsymbol{\mu_2}$ 之间，则有：

$$\boldsymbol{w}_t(\boldsymbol{\mu}_1 - \boldsymbol{x}_0) * \boldsymbol{w}_t(\boldsymbol{\mu}_2 - \boldsymbol{x}_0) \ > \ 0 \tag{17}$$

解得

$$|\ln \frac{P(\omega_1)}{P(\omega_2)}| > \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{18}$$

## 第五题

(a)

由于样本是独立同分布的，则有：

$$P(z_{i_1}, \ldots, z_{i_n}|P(\omega_i)) = \prod_{k=1}^{n} P(z_{i_k}|P(\omega_i)) \tag{19}$$

因为对于任意第 $k$ 个样本

$$P(z_{i_k} = 1|\omega_i) = P(\omega_i) = P(\omega_i)^{z_{i_k}}$$

$$P(z_{i_k} = 0|\omega_i) = 1 - P(\omega_i) = (1 - P(\omega_i))^{1-z_{i_k}}$$

所以

$$P(z_{i_1}, \ldots, z_{i_n}|P(\omega_i)) = \prod_{k=1}^{n} P(\omega_i)^{z_{i_k}}(1 - P(\omega_i))^{1-z_{i_k}} \tag{20}$$

证毕。

(b)

使用最大似然估计估计先验概率，等价于求似然估计取对数的最大

值，即求

$$\ln P(z_{i_1},\dots,z_{i_n}|P(\omega_i)) = \ln P(\omega_i)\sum_{k=1}^{n} z_{i_k} + \ln(1-P(\omega_i))\sum_{k=1}^{n}(1-z_{i_k}) \tag{21}$$

的最大值。计算偏导数得

$$\frac{\partial L}{\partial P(\omega_i)} = \sum_{k=1}^{n}(\frac{z_{i_k}}{P(\omega_i)} - \frac{1-z_{i_k}}{1-P(\omega_i)}) \tag{22}$$

令偏导为 0，化简

$$\sum_{k=1}^{n}(\frac{z_{i_k}}{P(\omega_i)} - \frac{1-z_{i_k}}{1-P(\omega_i)}) = \sum_{k=1}^{n}\frac{z_{i_k}-P(\omega_i)}{P(\omega_i)(1-P(\omega_i))} \tag{23}$$

由于式 (23) 中 $P(\omega_i)$ 和 $1-P(\omega_i)$ 均不可能为 0，则只能分子为 0，即

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik} \tag{24}$$

证毕。

## 第六题

(a) 根据期望定义公式，得

$$\begin{aligned}
\hat{\mu}_{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1} x_k \\
&= \frac{1}{n+1}(\sum_{i=1}^{n} x_k + x_{n+1}) \\
&= \frac{1}{n+1}(n\hat{\mu}_n + x_{n+1}) \\
&= \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)
\end{aligned} \tag{25}$$

根据方差定义公式，得

$$C_{n+1} = \frac{1}{n}\sum_{k=1}^{n+1}(x_k - \hat{\mu}_{n+1})(x_k - \hat{\mu}_{n+1})^t$$

$$= \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu}_{n+1})(x_k - \hat{\mu}_{n+1})^t + (x_{n+1} - \hat{\mu}_{n+1})(x_{n+1} - \hat{\mu}_{n+1})^t$$

将式 (25) 代入，且令 $\frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)$ 为 $Y$，得

$$= \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu}_n - Y)(x_k - \hat{\mu}_n - Y)^t + (x_{n+1} - \hat{\mu}_{n+1})(x_{n+1} - \hat{\mu}_{n+1})^t$$

令 $(x_{n+1} - \hat{\mu}_{n+1})(x_{n+1} - \hat{\mu}_{n+1})^t$ 为 $Z$，得

$$= \frac{1}{n}\sum_{k=1}^{n}[(x_k - \hat{\mu}_n)(x_k - \hat{\mu}_n)^t - (x_k - \hat{\mu}_n)Y^t - Y(x_k - \hat{\mu}_n)^t + YY^t] + Z$$

$$= \frac{n-1}{n}C_n + \frac{1}{n}\sum_{k=1}^{n}[(\hat{\mu}_n - x_k)Y^t + Y(\hat{\mu}_n - x_k)^t + YY^t] + Z$$

因为 $\sum_{k=1}^{n}(\hat{\mu}_n - x_k) = n * \hat{\mu}_n - \sum_{k=1}^{n} x_k = 0$，则

$$= \frac{n-1}{n}C_n + YY^t + Z$$

将 $Y$、$Z$ 代入，得

$$C_{n+1} = \frac{n-1}{n}C_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)(x_{n+1} - \hat{\mu}_n)^t$$

(b)

期望递归计算公式为

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)$$

每次递归是 $O(d)$ 的复杂度，递归 $n$ 次，则总体复杂度为 $O(nd)$。方

差递归计算公式为

$$C_{n+1} = \frac{n-1}{n} C_n + \frac{1}{n+1} (x_{n+1} - \hat{\mu}_n)(x_{n+1} - \hat{\mu}_n)^t$$

每次递归是 $O(d^2)$ 的复杂度，递归 $n$ 次，则总体复杂度为 $O(nd^2)$。