模式识别课后作业一

2020E8013282019 李一帆

October 7, 2020

1 问题 1

由题目可知，类别 ω_j 判断为 α_i 的决策代价可以表示为：

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0, & i = j \\ \lambda_s, & i \neq j \\ \lambda_r, & \text{reject} \end{cases}$$

条件风险 $R(\alpha_i|\mathrm{x})$ 可以表示为：

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

通过推导可知条件风险 $R_i(\mathrm{x})$ 可以表示为：

$$R_i(\mathrm{x}) = \begin{cases} \sum_{j\neq i}^{c} \lambda(\alpha_i|\omega_i)P(\omega_j|\mathrm{x}) = \lambda_s[1 - P(\omega_i|\mathrm{x})], & i = 1, ..., c \\ \lambda_r, & \text{reject} \end{cases}$$

因为在 $i = 1, 2, ..., c$ 情况下，如果对应的条件风险更小，则不会被拒绝，反之则被拒绝，即 $\lambda_s(1 - P(\omega_i|x)) < \lambda_r \Rightarrow P(\omega_i|x) > 1 - \lambda_r/\lambda_s$ 。根据条件风险最小化原则，有如下公式成立：

$$\arg\min_i R_i(\mathrm{x}) = \begin{cases} \arg\max_i P(\omega_i|\mathrm{x}), & if \max_i P(\omega_i|\mathrm{x}) > 1 - \lambda_r/\lambda_s \\ \text{reject}, & \text{otherwise} \end{cases}$$

当 $\lambda_r = 0$ 时，有 $\max_i P(\omega_i|\mathrm{x}) < 1$，第一个条件恒不成立，因此被拒绝。

当 $\lambda_r > \lambda_s$ 时，有 $1 - \lambda_r/\lambda_s < 0 \Rightarrow \max_i P(\omega_i|\mathrm{x}) > 1 - \lambda_r/\lambda_s$，即第一个条件恒成立，因此拒绝的条件会被无效化，起不到拒绝的效果。

2 问题 2

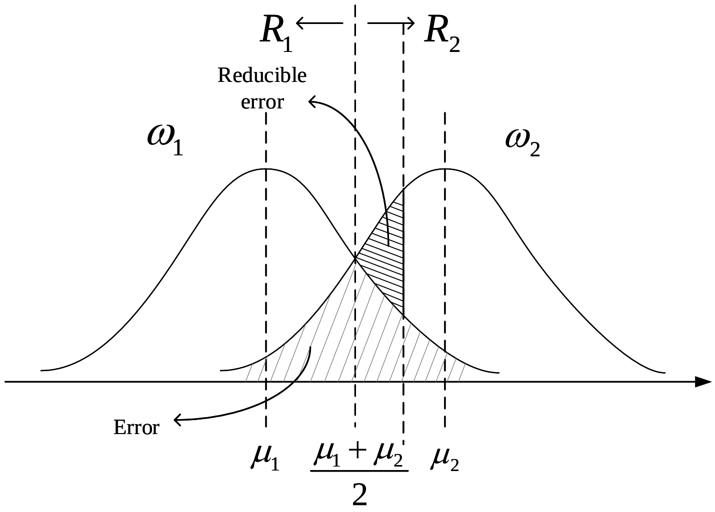

Figure 1: 错误率.

2.1 第 1 小问

由 Figure 1 所示，当决策边界位于两概率密度函数的交点处时，即 $\frac{\mu_1+\mu_2}{2}$，误差概率最小（reducible error=0）。根据定义，此刻的最小误差概率 P_e 为对阴影部分的积分，推导过程如下所示。

$$\begin{aligned}
P_e &= P(\mathrm{x} \in R_2, \omega_1) + P(\mathrm{x} \in R_1, \omega_2) \\
&= P(\mathrm{x} \in R_2|\omega_1)P(\omega_1) + P(\mathrm{x} \in R_1|\omega_2)P(\omega_2) \\
&= \frac{1}{2}(\int\limits_{R_2} P(\mathrm{x} \in R_2|\omega_1)d\mathrm{x} + \int\limits_{R_1} P(\mathrm{x} \in R_1|\omega_2)d\mathrm{x}) \\
&= \frac{1}{2} \cdot 2 \int\limits_{R_2} P(\mathrm{x} \in R_2|\omega_1)d\mathrm{x} \\
&= \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mathrm{x}-\mu_1)^2}{2\sigma^2}} d\mathrm{x} \\
&= \sigma \int_{\frac{\mu_1+\mu_2}{2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{\mathrm{x}-\mu_1}{\sigma})^2} d(\frac{\mathrm{x}-\mu_1}{\sigma}) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{\mu_2-\mu_1}{2\sigma}}^{+\infty} e^{-\frac{1}{2}\mu^2} d\mu
\end{aligned}$$

其中，由于 $\mu_2 > \mu_1$，所以积分下限 $\frac{\mu_2-\mu_1}{2}$ 应写为 $\frac{|\mu_1-\mu_2|}{2}$。

2.2 第 2 小问

由第 1 小问可知，$P_e \geq \frac{1}{\sqrt{2\pi}}\int_{\frac{|\mu_2-\mu_1|}{2\sigma}}^{+\infty} e^{-\frac{1}{2}\mu^2}d\mu$，因此，结合第 2 小问所给出信息有如下不等式成立。

$$\frac{1}{\sqrt{2\pi}}\int_a^{+\infty} e^{-\frac{1}{2}\mu^2}d\mu \leq P_e \leq \frac{1}{\sqrt{2\pi}a}e^{-\frac{1}{2}a^2}$$

其中，$a = \frac{|\mu_2-\mu_1|}{2\sigma}$。因为有如下极限成立：

$$\lim_{a\to+\infty}\frac{1}{\sqrt{2\pi}}\int_a^{+\infty} e^{-\frac{1}{2}\mu^2}d\mu = \lim_{a\to+\infty}[1-\Phi(a)] = 0$$

$$\lim_{a\to+\infty}\frac{1}{\sqrt{2\pi}a}e^{-\frac{1}{2}a^2} = 0$$

因此，根据夹逼准则可知，$\lim\limits_{a\to+\infty} P_e = 0$。

3 问题 3

3.1 第 1 小问

条件概率密度 $P(\mathrm{x}|\omega_\mathrm{i})$ 可以表示为如下形式。

$$P(\mathrm{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}\exp\left[-\frac{1}{2}(\mathrm{x}-\mu_i)^T\Sigma_i^{-1}(\mathrm{x}-\mu_i)\right]$$

3.2 第 2 小问

(a) 在类协方差矩阵 Σ_i 不等的情况下，最小误差判别函数 $g_i(\mathrm{x})$ 可以表示为

$$g_i(\mathrm{x}) = -\frac{1}{2}(\mathrm{x}-\mu_i)^t\Sigma_i^{-1}(\mathrm{x}-\mu_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

(b) 在类协方差矩阵相等的情况下，可以细分为两种情况。

Case 1：$\Sigma_i = \sigma^2\mathrm{I}$，此时的最小误差判别函数 $g_i(\mathrm{x})$ 可以表示为如下形式。

$$g_i(\mathrm{x}) = -\frac{\|\mathrm{x}-\mu_i\|^2}{2\sigma^2} + \ln P(\omega_i)$$

Case 2：$\Sigma_i = \Sigma$，此时最小误差判别函数 $g_i(\mathrm{x})$ 可以表示为如下形式。

$$g_i(\mathrm{x}) = -\frac{1}{2}(\mathrm{x}-\mu_i)^t\Sigma_i^{-1}(\mathrm{x}-\mu_i) + \ln P(\omega_i)$$

3.3 第 3 小问

方法 1：利用 PCA 对协方差矩阵进行降维处理。

方法 2：在协方差矩阵的基础上加上一个 $\varepsilon\mathrm{I}$，其中，ε 足够小。

4 问题 4

在高斯分布协方差矩阵相同的情况下，决策面与坐标轴的交点 x_0 可以表示为如下方程。

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln\left[P(\omega_i)/P(\omega_j)\right]}{(\mu_i - \mu_j)^t \Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$$

如果先验概率相等，决策面经过两均值向量的中点处。如果不等，则边界会移向先验概率较小的均值，如果两先验概率之差足够大，判定面可以不落在两均值向量之间。

5 问题 5

5.1 第 1 小问

在 $z_{ik} = 1$ 的情况下，第 k 个样本的类别为 ω_i，先验概率为 $P(\omega_i)$。

$$P(\omega_i)^{z_{ik}} \cdot (1 - P(\omega_i))^{1-z_{ik}} = P(\omega_i)$$

在 $z_{ik} = 0$ 的情况下，第 k 个样本的类别不为 ω_i，先验概率为 $1 - P(\omega_i)$。

$$P(\omega_i)^{z_{ik}} \cdot (1 - P(\omega_i))^{1-z_{ik}} = 1 - P(\omega_i)$$

当 z_{ik} 为不确定值时，第 k 个样本类别不确定，对应的先验概率可以合并为 $P(\omega_i)^{z_{ik}}(1 - P(\omega_i))$。又根据独立性原则，最终的似然可以表示为如下形式。

$$P(z_{ik}, ..., z_{in}|P(\omega_i)) = \prod_{k=1}^{n} P(\omega_i)^{z_{ik}}(1 - P(\omega_i))^{1-z_{ik}}$$

5.2 第 2 小问

将似然 $P(z_{ik}, ..., z_{in}|P(\omega_i))$ 取对数后得 $l(P(\omega_i))$。

$$l(P(\omega_i)) = \sum_{k=1}^{n} \left[z_{ik} \ln P(\omega_i) + (1 - z_{ik}) \ln(1 - P(\omega_i))\right]$$

将取对数后的似然对先验 $P(\omega_i)$ 进行求导运算，如下所示。

$$\nabla_{P(\omega_i)} l = \sum_{k=1}^{n} \left[\frac{z_{ik}}{P(\omega_i)} + \frac{z_{ik} - 1}{1 - P(\omega_i)}\right] = \sum_{k=1}^{n} \frac{z_{ik} - P(\omega_i)}{P(\omega_i)(1 - P(\omega_i))}$$

令导数为 0 可以求得对先验 $P(\omega_i)$ 的最大似然估计。

$$\nabla_{P(\omega_i)} l = 0 \Rightarrow \sum_{k=1}^{n} z_{ik} - nP(\omega_i) = 0 \Rightarrow \hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik}$$

6 问题 6

6.1 第 1 小问

(a) 将均值的迭代公式进行展开，如下所示。

$$
\begin{aligned}
\hat{\mu}_{n+1} &= \hat{\mu}_n + \frac{1}{n+1}(\mathrm{x}_{n+1} - \hat{\mu}_n) = \frac{n}{n+1}\hat{\mu}_n + \frac{1}{n+1}\mathrm{x}_{n+1} = \frac{n}{n+1} \cdot (\frac{n-1}{n}\hat{\mu}_{n-1} + \frac{1}{n}\mathrm{x}_n) \\
&= \frac{n}{n+1} \cdot (\frac{n-2}{n}\hat{\mu}_{n-2} + \frac{1}{n}\mathrm{x}_{n-1} + \frac{1}{n}\mathrm{x}_n) + \frac{1}{n+1}\mathrm{x}_{n+1} \\
&= \frac{n-2}{n+1}\hat{\mu}_{n-2} + \frac{1}{n+1}(\mathrm{x}_{n-1} + \mathrm{x}_n + \mathrm{x}_{n+1}) \\
&= \cdots = \frac{1}{n+1}\hat{\mu}_1 + \frac{1}{n+1}(\mathrm{x}_2 + \cdots \mathrm{x}_n + \mathrm{x}_{n+1}) \\
&= \frac{1}{n+1}(\mathrm{x}_1 + \mathrm{x}_2 + \cdots \mathrm{x}_n + \mathrm{x}_{n+1}) = \frac{1}{n+1}\sum_{k=1}^{n+1}\mathrm{x}_k
\end{aligned}
$$

因此，$\hat{\mu}_n = \frac{1}{n}\sum\limits_{k=1}^{n}\mathrm{x}_k$ 成立。

(b) 将上面得到的均值迭代公式代入到方差计算公式中可以有如下推导。

$$
\begin{aligned}
C_{n+1} &= \frac{1}{n}\sum_{i=1}^{n+1}\left[x_i - (\mu_n + \frac{x_{n+1}-\mu_n}{n+1})\right]\left[x_i - (\mu_n + \frac{x_{n+1}-\mu_n}{n+1})\right]^t \\
&= \frac{1}{n}\sum_{i=1}^{n+1}\left[(x_i-\mu_n) - (\frac{x_{n+1}-\mu_n}{n+1})\right]\left[(x_i-\mu_n) - (\frac{x_{n+1}-\mu_n}{n+1})\right]^t \\
&= \frac{1}{n}\sum_{i=1}^{n+1}\left[(x_i-\mu_n)(x_i-\mu_n)^t - 2(x_i-\mu_{n-1})(\frac{x_{n+1}-\mu_n}{n+1})^t + (\frac{x_{n+1}-\mu_n}{n+1})(\frac{x_{n+1}-\mu_n}{n+1})^t\right] \\
&= \frac{n+1}{n}(\frac{x_{n+1}-\mu_n}{n+1})(\frac{x_{n+1}-\mu_n}{n+1})^t + \frac{1}{n}\sum_{i=1}^{n+1}\left[(x_i-\mu_n)(x_i-\mu_n)^t - 2(x_i-\mu_{n-1})(\frac{x_{n+1}-\mu_n}{n+1})^t\right] \\
&= \frac{1}{n(n+1)}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t + \frac{1}{n}\left[(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t - 2\frac{(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t}{n+1}\right] \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\mu_n)(x_i-\mu_n)^t - 2\frac{(x_i-\mu_{n-1})(x_{n+1}-\mu_n)^t}{n+1}\right] \\
&= (\frac{1}{n(n+1)} + \frac{1}{n}\cdot\frac{n-1}{n+1})(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t + \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_n)(x_i-\mu_n)^t \\
&- 2\frac{x_{n+1}-\mu_n}{n(n+1)}\sum_{i=1}^{n}(x_i-\mu_{n-1}) \\
&= \frac{1}{n+1}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t + \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_n)(x_i-\mu_n)^t \\
&= \frac{1}{n+1}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t + \frac{n-1}{n}\cdot\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_n)(x_i-\mu_n)^t \\
&= \frac{1}{n+1}(x_{n+1}-\mu_n)(x_{n+1}-\mu_n)^t + \frac{n-1}{n}C_n
\end{aligned}
$$

6.2 第 2 小问

计算均值迭代每一步需要 d 次运算，迭代 n 次则需要 nd 次运算；计算方差迭代每一步计算 $x_{n+1} - m_n$ 需要 d 次运算，计算 $(x_{n+1} - m_n)(x_{n+1} - m_n)^t$ 需要 d^2 次运算，因此每一步需要 d^2 次运算，迭代 n 次则需要 nd^2 次运算。