Pattern Recognition
University of Chinese Academy of Sciences
Fall 2023
Shiming Xiang, Gaofeng Meng

Homework 1

Chenkai GUO
2023.10.1

1. 请描述最小错误率贝叶斯决策的计算步骤（包含已知条件以及求解任务）；给出一种两类情形下的最小错误率贝叶斯决策规则。

最小错误率贝叶斯决策

已知条件：（1）类别：$\omega_i, i = 1, \ldots, c$

（2）特征向量：$\mathbf{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d$

（3）先验概率：$P(\omega_i), \sum_{i=1}^{c} P(\omega_i) = 1$

（4）概率密度函数/条件概率/似然：$P(\mathbf{x}|\omega_i)$
（一般假设为正态分布或其他分布）

求解任务：对一个给定观测样本 $\mathbf{x}$ 给出其最合理的分类类别，即 $P(\omega_i|\mathbf{x})$

计算步骤：根据贝叶斯公式计算后验概率：$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\mathbf{x}|\omega_j)P(\omega_j)}$

决策规则：选择后验概率最大的类别作为输出，即若 $P(\omega_i|\mathbf{x}) = \max_{j=1,2,\ldots,c} P(\omega_j|\mathbf{x})$，则 $\mathbf{x} \in \omega_i$

※ 考虑两类情形下，即 $\boldsymbol{\omega} = (\omega_1, \omega_2)$：

形式 1：根据贝叶斯公式可得：

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

若 $\frac{p(\mathbf{x}|\omega_1)P(\omega_1)}{P(\mathbf{x})} > \frac{p(\mathbf{x}|\omega_2)P(\omega_2)}{P(\mathbf{x})}$，即当 $p(\mathbf{x}|\omega_1)P(\omega_1) > p(\mathbf{x}|\omega_2)P(\omega_2)$ 时，$\mathbf{x} \in \omega_1$；反之 $\mathbf{x} \in \omega_2$

形式 2：若似然比 $l(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1}{p(\mathbf{x}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$，则 $\mathbf{x} \in \omega_1$；反之 $\mathbf{x} \in \omega_2$

形式 3：若负对数似然比 $-\ln(l(\mathbf{x})) = -\ln(\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}) = -\ln(p(\mathbf{x}|\omega_1)) + \ln(p(\mathbf{x}|\omega_2)) < -\ln(\frac{P(\omega_2)}{P(\omega_1)}) = \ln(P(\omega_1)) - \ln(P(\omega_2))$，则 $\mathbf{x} \in \omega_1$；反之 $\mathbf{x} \in \omega_2$

2. 请描述最小风险贝叶斯决策的计算步骤（包含已知条件以及求解任务）；给出一种两类情形下的最小风险贝叶斯决策规则。

最小风险贝叶斯决策

已知条件：(*1*) 类别：$\omega_i, i = 1, \ldots, c$

(*2*) 特征向量：$\mathbf{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d$

(*3*) 先验概率：$P(\omega_i), \sum_{i=1}^{c} P(\omega_i) = 1$

(*4*) 概率密度函数/条件概率/似然：$P(\mathbf{x}|\omega_i)$
(一般假设为正态分布或其他分布)

(*5*) 决策空间：$\boldsymbol{\alpha}$，$\alpha_i, i = 1, \ldots, a$

(*6*) 损失函数：$\lambda(\alpha_i|\omega_j)$，表示当类别为 ω_j 时所采取的决策 α_i 多引起的损失，简记为 λ_{ij}

求解任务：给定一个观测样本 $\mathbf{x}$，如何对其分类使得其风险/损失最小

计算步骤：(*1*) 根据贝叶斯公式计算后验概率：$P(\omega_i|\mathbf{x})$

(*2*) 计算条件风险：$R(\alpha_i|\mathbf{x}) = E[\lambda(\alpha_i|\omega_j)] = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x})$
(条件风险 $R(\alpha_i|\mathbf{x})$ 为随机变量 $\mathbf{x}$ 的函数)

(*3*) 选择风险最小的决策：$a = \arg\min_{j=1,\ldots,a} R(\alpha_j|\mathbf{x})$

决策规则：最小化期望风险：$\min_{\alpha} R(\alpha) = \min_{\alpha} \int R(\alpha(\mathbf{x})|\mathbf{x}) \cdot p(\mathbf{x})\ dx$
对特定 $\mathbf{x}$ 的类别决策则为选择条件风险 $R(\alpha_i|\mathbf{x})$ 最小的决策

※ 考虑两类情形下，即 $\boldsymbol{\omega} = (\omega_1, \omega_2)$,：

情况一：不考虑拒实，即 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$

计算类别的条件风险得*:*

$$R(\alpha_1|\mathbf{x}) = \lambda_{11} P(\omega_1|\mathbf{x}) + \lambda_{12} P(\omega_2|\mathbf{x})$$
$$R(\alpha_2|\mathbf{x}) = \lambda_{21} P(\omega_1|\mathbf{x}) + \lambda_{22} P(\omega_2|\mathbf{x})$$

形式 *1*：若 $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$，则 $\alpha = \alpha_1$；反之 $\alpha = \alpha_2$

形式 *2*:若 $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x}) \Leftrightarrow \lambda_{11} P(\omega_1|\mathbf{x}) + \lambda_{12} P(\omega_2|\mathbf{x}) < \lambda_{21} P(\omega_1|\mathbf{x}) + \lambda_{22} P(\omega_2|\mathbf{x})$ $\Leftrightarrow (\lambda_{11} - \lambda_{21}) P(\omega_1|\mathbf{x}) < (\lambda_{22} - \lambda_{12}) P(\omega_2|\mathbf{x})$，不失一般性，假设 $\lambda_{11} < \lambda_{21}, \lambda_{22} < \lambda_{12}$，则得到，若 $\frac{P(\omega_1|\mathbf{x})}{P(\omega_2|\mathbf{x})} > \frac{\lambda_{22} - \lambda_{12}}{\lambda_{11} - \lambda_{21}}$，此时 $\alpha = \alpha_1$；反之 $\alpha = \alpha_2$

形式 *3*：因为 $\frac{P(\omega_1|\mathbf{x})}{P(\omega_2|\mathbf{x})} = \frac{P(\mathbf{x}|\omega_1) P(\omega_1)}{P(\mathbf{x}|\omega_2) P(\omega_2)}$，则若似然比 $l(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \cdot \frac{\lambda_{22} - \lambda_{12}}{\lambda_{11} - \lambda_{21}}$，则 $\alpha = \alpha_1$；反之 $\alpha = \alpha_2$

形式 *4*：考虑 *Zero-one Loss*：$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \quad i, j = 1, 2, ..., c$

此时 $R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x}) = \sum_{i \neq j} P(\omega_j|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x})$ (此时，最小风险贝叶斯决策等价于最小错误率贝叶斯决策)；对于二类决策，若 $P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x})$，则 $\mathbf{x} \in \omega_1, \alpha = \alpha_1$，反之 $\mathbf{x} \in \omega_2, \alpha = \alpha_2$

情况二：考虑拒实，即 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,reject)$，$\lambda(\alpha_i|\omega_j)=\begin{cases}0, & i=j\\ \lambda_s, & i\neq j\\ \lambda_r, & reject\end{cases}$

计算类别的条件风险得: $R(\alpha_i|\mathbf{x})=\sum\limits_{j=1}^{c}\lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$

此时：$R_i(\mathbf{x})\triangleq R(\alpha_i|\mathbf{x})=\begin{cases}\lambda_s[1-P(\omega_i\mid\mathbf{x})], & i=1,2\\ \lambda_r, & reject\end{cases}$，决策规则为 $\arg\min\limits_i R_i(\mathbf{x})$

3. 对于 c 类问题，假定各类条件概率密度函数均为多元正态分布。在最小错误率贝叶斯决策的框架下，请写出其判别函数；请分别指出在什么情况下可以获得最小距离分类器，在什么情况下可以得到线性判别函数。

在最小错误率贝叶斯决策的框架下：假设各类条件概率密度函数（似然）为多元正态分布，即 $p(\mathbf{x}|\omega_i)\sim N(\mu_i,\Sigma_i), i=1,2,\ldots,c$，采用的判别函数为二次判别函数 *(Quadratic Discriminant Function, QDF)*:

$$
\begin{aligned}
g_i(\mathbf{x})&=\ln\left(p(\mathbf{x}\mid\omega_i)\right)+\ln\left(P(\omega_i)\right)\\
&=-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln(\mid\boldsymbol{\Sigma}_i\mid)+\ln\left(P(\omega_i)\right)\\
&\quad(i=1,2,\ldots,c)
\end{aligned}
$$

情形一：假设 $\boldsymbol{\Sigma}_i=\sigma^2\mathbf{I}\in\mathbb{R}^{d\times d}, i=1,2,\ldots,c$

此时，QDF 可化为

$$
\begin{aligned}
g_i(\mathbf{x})&=-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)\underbrace{-\frac{d}{2}\ln(2\pi)-\frac{1}{2}\ln(\sigma^{2d})}_{Const.}+\ln\left(P(\omega_i)\right)\\
&\Leftrightarrow-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)+\ln\left(P(\omega_i)\right)\\
&=-\frac{1}{2\sigma^2}\underbrace{\|\mathbf{x}-\boldsymbol{\mu}_i\|^2}_{ED}+\ln\left(P(\omega_i)\right)
\end{aligned}
$$

(*1*) 当各类先验概率相等时，即 $P(\omega_i)=P(\omega_j)$ 时，$g_i(\mathbf{x})=-\frac{1}{2\sigma^2}\|\mathbf{x}-\mu_i\|^2+\ln\left(P(\omega_i)\right)$ $\Leftrightarrow g_i(\mathbf{x})=-\frac{1}{2\sigma^2}\|\mathbf{x}-\mu_i\|^2$，此时最小错误率贝叶斯规则即为选取样本 $\mathbf{x}$ 对各类均值向量 $\boldsymbol{\mu}_i$ 距离最近的一类作为其分类，这种分类器被称为最小距离分类器

(*2*) 当各类先验概率不相等时，即 $P(\omega_i)\neq P(\omega_j)$ 时，此时：

$$
\begin{aligned}
g_i(\mathbf{x})&=-\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)+\ln\left(P(\omega_i)\right)\\
&=-\frac{1}{2\sigma^2}(\mathbf{x}^T\mathbf{x}-2\boldsymbol{\mu}_i^T\mathbf{x}+\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i)+\ln\left(P(\omega_i)\right)
\end{aligned}
$$

$$\Leftrightarrow \underbrace{\left(\frac{1}{\sigma^2}\boldsymbol{\mu}_i^T\right)}_{\mathbf{w}_i^T}\mathbf{x} + \underbrace{\left(-\frac{1}{2\sigma^2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i + \ln\left(P(\omega_i)\right)\right)}_{w_{i0}}$$

此时可以得到 $g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{i0}$，判别函数 $g_i(\mathbf{x})$ 为 $\mathbf{x}$ 的线性函数，因此也称其为线性判别函数 *(Linear Discriminant Function, LDF)*，这种分类器被称为线性分类器

4. 针对概率密度函数参数估计问题，请描述最大似然估计的计算步骤（包含已知条件以及求解任务）。

最大似然估计（MLE）

已知条件：(1) 类条件概率密度 $p(\mathbf{x}|\omega_i)$ 的函数形式（如：服从 n 维正态分布）

(2) 类条件概率密度 $p(\mathbf{x}|\omega_i)$ 独立同分布，且不同类别的参数是独立的

(3) 样本集：$\mathbf{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$

(4) MLE 假设 $\boldsymbol{\theta}$ 是未知但固定的，即 $P(\boldsymbol{\theta}) = 1$

求解任务：求解能够使获得样本集 $\mathbf{D}$ 可能性最大的参数 θ，即 $\arg\max\limits_{\theta} P(\mathbf{D}|\boldsymbol{\theta})$

计算步骤：(1) 计算似然函数或对数似然函数：

$$l(\boldsymbol{\theta}) = P(\mathbf{D}|\boldsymbol{\theta}) = P(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathbf{x}_i|\boldsymbol{\theta})$$

$$H(\boldsymbol{\theta}) = \ln l(\boldsymbol{\theta}) = \ln \prod_{i=1}^{n} p(\mathbf{x}_i|\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln\left(p(\mathbf{x}_i|\boldsymbol{\theta})\right)$$

(2) 求解似然函数或对数似然函数的极值，得到估计参数集：

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \Leftrightarrow \nabla_{\boldsymbol{\theta}}(l(\boldsymbol{\theta})) = \left[\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_2}, \cdots, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_m}\right]^T = 0$$

$$\frac{\partial H(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0 \Leftrightarrow \nabla_{\boldsymbol{\theta}}(H(\boldsymbol{\theta})) = \sum_{i=1}^{n} \nabla_{\boldsymbol{\theta}} \ln\left(p(\mathbf{x}_i|\boldsymbol{\theta})\right) = 0$$

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T$$

5. 针对样本的类条件概率密度函数估计问题，请描述贝叶斯估计的计算步骤（包含已知条件以及求解任务）。

贝叶斯估计

已知条件：(1) 待估参数的先验概率：$p(\boldsymbol{\theta})$

(2) 样本集：$\mathbf{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$（并假设数据为独立采样获取）

求解任务：使后验概率 $p(\boldsymbol{\theta}|\mathbf{D})$ 最大，即 $\arg\max\limits_{\theta} P(\boldsymbol{\theta}|\mathbf{D})$，并最终基于有限的样本集 $\mathbf{D}$ 估计后验概率密度函数 $p(\mathbf{x}|\mathbf{D})$

计算步骤：(1) 计算后验概率密度函数：

$$p(\boldsymbol{\theta}|\mathbf{D}) = \frac{p(\mathbf{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = \frac{\prod_{i=1}^{n} p(\mathbf{x}_i|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \prod_{i=1}^{n} p(\mathbf{x}_i|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} = \alpha \prod_{i=1}^{n} p(\mathbf{x}_i|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

(2) 计算平均估计量：$\hat{\boldsymbol{\theta}} = \int_{\boldsymbol{\theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{D})\ d\boldsymbol{\theta}$

（*3*）估计后验数据分布：

$$\begin{tikzcd} p(\boldsymbol{\theta}) \arrow[r] & p(\boldsymbol{\theta}|\mathbf{D}) \arrow[r] \arrow[rr, bend left] & p(\mathbf{x}|\boldsymbol{\theta}) \arrow[r] & p(\mathbf{x}|\mathbf{D}) \end{tikzcd}$$

$$\begin{aligned} p(\mathbf{x}|\mathbf{D}) &= \int_{\boldsymbol{\theta}} p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{D}) d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} \frac{p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{D})}{p(\mathbf{D})} d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} \frac{p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{D}) p(\boldsymbol{\theta}, \mathbf{D})}{p(\mathbf{D})} d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} {\color{blue} p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{D})} {\color{red} p(\boldsymbol{\theta}|\mathbf{D})} d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} {\color{blue} p(\mathbf{x}|\boldsymbol{\theta})} {\color{red} p(\boldsymbol{\theta}|\mathbf{D})} d\boldsymbol{\theta} \end{aligned}$$

（给定参数 $\boldsymbol{\theta}$ 时，样本分布于训练集 $\mathbf{D}$ 无关）

6. 请指出最大似然估计和贝叶斯估计的不同之处。

不同点：（*1*）对估计参数的假设不同：*MLE* 假设待估计的参数是未知但固定的变量，贝叶斯估计则假设待估计的参数是一个随机变量也因此，*MLE* 为点估计，贝叶斯估计为对分布估计

（*2*）研究的对象不同：因为对待估计参数的假设不同，*MLE* 重点研究样本空间，而贝叶斯估计则认为样本是固定的，则重点研究待估计参数的分布