《模式识别》第1 次作业

姓名：谷绍伟　　　学号：202418020428007

1：(简答题) 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。

答： 计算步骤：已知条件：

- 类别：$\omega_i, i = 1, \ldots, c$
- 特征矢量 $\mathbf{x} = [x_1, \ldots, x_d] \in \mathbb{R}^d$
- 先验概率 $P(\omega_i)$, $\sum_{i=1}^{c} P(\omega_i) = 1$
- 概率密度函数 (条件概率) $p(\mathbf{x} \mid \omega_i)$

任务：如果观测到一个样本 x，应该将其分到哪一类使得分类最合理。
计算步骤：

- 利用贝叶斯公式计算后验概率：

$$P(\omega_i|x) = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{\sum_{j=i}^{c} p(\mathbf{x}|\omega_j)p(\omega_j)}, \quad \sum_{i=1}^{c} P(\omega_i|\mathbf{x}) = 1$$

- 在各类决策中选择后验概率最大的决策：$\mathbf{x} \in \omega_i$，若

$$p(\omega_i|\mathbf{x}) = \arg\max_{j=1,\ldots,c} P(\omega_j|\mathbf{x})$$

两类情形下的最小错误率贝叶斯决策规则：考察没有拒识的两类情形，即 $a = c = 2$，此时有：

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

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

决策规则：

$$\mathbf{x} \in \omega_1, if \quad p(\mathbf{x}|\omega_1)p(\omega_1) > p(\mathbf{x}|\omega_2)p(\omega_2), else \quad \mathbf{x} \in \omega_2$$

2：(简答题) 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

答： 计算步骤：已知条件：

- 类别：$\omega_i, i = 1, \dots, c$
- 特征矢量 $\mathbf{x} = [x_1, \dots, x_d] \in \mathbb{R}^d$
- 先验概率 $P(\omega_i)$, $\sum_{i=1}^{c} P(\omega_i) = 1$
- 概率密度函数 (条件概率) $p(\mathbf{x} \mid \omega_i)$
- 决策空间包含 a 个决策 $\alpha_i, i = 1, 2, ..., a$
- 损失函数 $\lambda(\alpha_i|\omega_j)$, 表示当类别为 ω_j 所采取的决策 α_i 所引起的损失，简记为 $\lambda_i j$

任务：如果观测到一个样本 x，应该将其分到哪一类使得其风险最小？
计算步骤：

- 利用贝叶斯公式计算后验概率：$P(\omega_j|x), j = 1, 2, \dots, c$
- 利用决策计算风险：

$$R(\alpha_i|\mathbf{x}) = E[\lambda(\alpha_i|\omega_j)] = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}), \quad i = 1, 2, \dots, a$$

- 在各类决策中选择风险最小的决策：

$$a = \arg\min_{j=1,\dots,a} R(\alpha_j|\mathbf{x})$$

两类情形下的最小错误率贝叶斯决策规则：考察没有拒识的两类情形，即 $a = c = 2$，此时有：

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

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

决策规则：

$$\alpha = \begin{cases} \alpha_1, & R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x}) \\ \alpha_2, & R(\alpha_1|\mathbf{x}) \geq R(\alpha_2|\mathbf{x}) \end{cases}$$

3：(简答题) 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。

答：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判别函数为

$$\begin{aligned} g_i(\mathbf{x}) &= \ln p(\mathbf{x}|\omega_i) + \ln p(\omega_i) \\ &= -\frac{1}{2}(\mathbf{x} - \mu_\mathbf{i})^T \boldsymbol{\Sigma_i^{-1}} (\mathbf{x} - \mu_\mathbf{i}) - \frac{d}{2} ln(2\pi) - \frac{1}{2} ln|\boldsymbol{\Sigma_i}| + \ln p(\omega_i), \quad (i = 1, 2, \dots, c) \end{aligned}$$

$\boldsymbol{\Sigma_i} = \sigma \boldsymbol{I}, i = 1, 2, \dots, c$。当先验概率相等时，即 $p(\omega_i) = p(\omega_j)$，此时判别函数可以简化为 $g_i(\boldsymbol{x}) = -\frac{1}{2\sigma^2}||\boldsymbol{x} - \boldsymbol{\mu_i}||_2^2$，最小错误率贝叶斯规则被简化为 $\underset{i=1,\dots,a}{\arg\min}\,||\boldsymbol{x} - \boldsymbol{\mu_i}||_2^2$，称为最小距离分类器。

先验概率不相等时，即为 $p(\omega_i) \neq p(\omega_j)$，则

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

由于每一类的判别函数均包含 $\mathbf{x}^T\mathbf{x}$，与下标无关，因此可以进一步简化为线性判别函数：

$$g_i(\boldsymbol{x}) = \frac{1}{\sigma^2}\mu_i^T\mathbf{x} - \frac{\mu_i^T\mu_i}{2\sigma^2} + \ln\left(P(\omega_i)\right) = \boldsymbol{\omega_i^T x} + \omega_{i0}$$

其中 $\boldsymbol{\omega_i} = \frac{\mu_i}{\sigma^2}$, $\omega_{i0} = -\frac{\mu_i^T\mu_i}{2\sigma^2} + \ln\left(P(\omega_i)\right)$

4：(简答题) 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)。

答：最大似然估计的基本原理是对于随机抽取的 n 个样本，寻找最合理的参数估计使得该模型中能抽取到这 n 个样本的概率最大。

设 n 次独立随机采样获得的样本集 $D = \{\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n}\}$，可以得到其联合概率：

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

称为参数 θ 相对于样本集 D 的似然函数。最大似然估计就是求解 θ 使得似然函数最大，即

$$\underset{\theta \in \Theta}{\arg\max}\,\theta$$

为了计算方便一般采用对数似然函数：

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

$$\arg\max l(\theta) = \arg\max H(\theta)$$

当 $l(\theta)$ 可微时，$\frac{\partial l(\theta)}{\partial \theta} = 0$，对于多维情形，$\theta = [\theta_1, \theta_2, \dots, \theta_m]^T$，则梯度向量为 $\mathbf{0}$，得到：

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

用梯度上升法求解：

$$\theta = \theta + \eta\frac{\partial l(\theta)}{\partial \theta}$$

5： (简答题) 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)。

答：

参数先验分布 $p(\theta)$：是指在没有任何数据时，有关参数 θ 的分布情况（根据领域知识或经验）。

给定样本集 $D=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}$，数据独立采样，且服从数据分布：

$$p(D\mid\theta)=p(\mathbf{x}_1,\mathbf{x}_2...,\mathbf{x}_n\mid\theta)=\prod_{i=1}^{n}p(\mathbf{x}_i\mid\theta)$$

利用贝叶斯公式计算参数的后验分布 $p(\theta|D)$：

$$p(\theta\mid D)=\frac{p(D\mid\theta)p(\theta)}{p(D)}$$

其中 $p(\theta|D)$ 中融合了先验知识和数据信息，$p(D)$ 是与参数无关的归一化因子，根据全概率公式：

$$p(D)=\int_{\theta}p(D\mid\theta)p(\theta)d\theta$$

对于一般情况，计算 $p(D)$ 十分困难，对 D 进行采样，可得贝叶斯参数估计中的**后验概率密度函数：**

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

得到 $p(\theta|D)$，可以进一步获得参数 θ 的估计值：如对 $p(\theta|D)$ 进行采样、最大后验概率估计（MAP）、后验数据分布等。

6： (简答题) 请指出最大似然估计和贝叶斯估计的不同之处。

答：最大似然估计是将待估计的参数当作未知但固定的变量，其任务是根据观测数据估计其在参数空间中的取值。

贝叶斯估计将待估计的参数视为一个随机变量，其中的一个核心任务是根据观测数据对参数的分布进行估计。