模式识别第一次作业

孙浩淼 201928013229100

Email: sunhm15@gmail.com

1 问题一

a) 证明此规则下的误差概率为

$$P(error) = P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)dx + P(\omega_2)\int_{\theta}^{\infty} p(x|\omega_2)dx$$

证： 如果类别为 ω_1，错误率

$$P(error|\omega_1) = \int_{-\infty}^{\theta} p(x|\omega_1)dx$$

如果类别为 ω_2，错误率

$$P(error|\omega_2) = \int_{\theta}^{\infty} p(x|\omega_2)dx$$

因此，总的错误率

$$\begin{aligned} P(error) &= P(\omega_1)P(error|\omega_1) + P(\omega_2)P(error|\omega_2) \\ &= P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)dx + P(\omega_2)\int_{\theta}^{\infty} p(x|\omega_2)dx \end{aligned}$$

b) 通过微分运算，证明最小化 P(error) 的一个必要条件为 θ 满足

$$p(\theta|\omega_1)P(\omega_1) = p(\theta|\omega_2)P(\omega_2)$$

证： $P(error)$ 最小化的必要条件为导数为 0，即

$$\frac{\partial P(error)}{\partial \theta} = P(\omega_1)p(\theta|\omega_1) - P(\omega_2)p(\theta|\omega_2) = 0$$

即，

$$p(\theta|\omega_1)P(\omega_1) = p(\theta|\omega_2)P(\omega_2)$$

c) θ 可以唯一确定吗?

解: 不能唯一确定，例如

$$
\begin{aligned}
p(x|\omega_1) &= 0.1 \qquad x \in [0,10] \\
p(x|\omega_2) &= 0.1 \qquad x \in [5,15] \\
P(\omega_1) &= P(\omega_2)
\end{aligned}
$$

则，$\forall\theta \in [5,10]$ 都满足上述式子。

d) 给出一个例子说明满足此时的一个 θ 有可能使误差最大化。

解: 假设 ω_1, ω_2 满足

$$
\begin{aligned}
p(x|\omega_1) &= 1 - |x| \qquad x \in [0,10] \\
p(x|\omega_2) &= 2 - 4|x| \qquad x \in [-0.5, 0.5] \\
P(\omega_1) &= P(\omega_2)
\end{aligned}
$$

由公式可以得出 $\theta = \pm\frac{1}{3}$，而

$$
\begin{aligned}
P(error|\theta = -\frac{1}{3}) &= \frac{25}{36} \\
P(error|\theta = \frac{1}{3}) &= \frac{5}{12}
\end{aligned}
$$

其中 $\theta = -\frac{1}{3}$ 时，误差最大，而 $\theta = \frac{1}{3}$ 时误差最小。

2 问题二

证: 决策风险可以显示如下,

$$
C(\omega_j|x) = \begin{cases} \lambda_r & j = c+1 \\ \lambda_s(1 - P(\omega_j|x)) & j \neq i \end{cases}
$$

因此，做出 ω_i 决策的充分必要条件为

$$
\begin{cases} \lambda_s(1 - P(\omega_j|x)) \geq \lambda_s(1 - P(\omega_i|x)) & \forall\omega_j \neq \omega_i \\ \lambda_s(1 - P(\omega_i|x)) \leq \lambda_r \end{cases}
$$

即,

$$
\begin{cases} P(\omega_j|x) \leq P(\omega_i|x) & \forall\omega_j \neq \omega_i \\ P(\omega_i|x) \geq 1 - \dfrac{\lambda_r}{\lambda_s} \end{cases}
$$

如果 $\lambda_r = 0$，则 $C(\omega_{c+1}|x) = 0 < C(\omega_j|x)$，即始终做出决策为 ω_{c+1}，如果 $\lambda_r > \lambda_s$，则 $C(\omega_{c+1}|x) = \lambda_r > \lambda_c > C(\omega_j|x)$，即不会做出 ω_{c+1} 的决策。

3 问题三

a) 解释为什么作为先验概率 $P(\omega_1)$ 函数的总贝叶斯风险一定要下凹?

证: 设 $\lambda_{ij}=\lambda(\alpha_i|\omega_j)$，$R_i$ 为分类器判断为 ω_i 类别的决策区域，则总贝叶斯风险可以表示如下，

$$R=\int\limits_{R_1}[\lambda_{11}P(\omega_1)p(x|\omega_1)+\lambda_{12}P(\omega_2)p(x|\omega_2)]dx+\int\limits_{R_2}[\lambda_{21}P(\omega_1)p(x|\omega_1)+\lambda_{22}P(\omega_2)p(x|\omega_2)]dx$$

将等式 $P(\omega_2)=1-P(\omega_1)$ 和 $\int\limits_{R_2}p(x|\omega_2)dx=1-\int\limits_{R_1}p(x|\omega_2)dx$ 带入并化简可以得到，

$$\begin{aligned}R=&\lambda_{22}+(\lambda_{12}-\lambda_{22})\int\limits_{R_1}p(x|\omega_2)dx\\&+[(\lambda_{11}-\lambda_{22})+(\lambda_{21}-\lambda_{11})\int\limits_{R_2}p(x|\omega_1)dx+(\lambda_{22}-\lambda_{12})\int\limits_{R_1}p(x|\omega_2)]P(\omega_1)\end{aligned}$$

可以看出，对于任意的 $0<P^*(\omega_1)<1$，存在一个策略使得决策风险最小。而如果固定策略 (R_1,R_2 固定不变) 时，随着 $P(\omega_1)$ 的不断变化，总的贝叶斯风险显示一条过点 $(P^*(\omega_1),R(P^*(\omega_1)))$ 的直线，实际的贝叶斯曲线风险的曲线随着决策区域的改变，一定会位于该直线的下方。

b) 利用极小化极大准则，对于两个未知先验概率的一维正态分布进行决策

解: 对于 0-1 风险决策问题 ($\lambda_{11}=\lambda_{22}=0,\lambda_{12}=\lambda_{21}=1$) 而言，根据极大极小准则，最优边界点需要满足，

$$\int\limits_{R_1}p(x|\omega_2)dx=\int\limits_{R_2}p(x|\omega_1)dx$$

将决策区域 x^* 代入得，

$$\int\limits_{-\infty}^{x^*}p(x|\omega_2)dx=\int\limits_{x^*}^{\infty}p(x|\omega_1)dx$$

因为两个类别的分布均为正态分布，根据正态分布的性质，上式可以等价于，

$$\frac{\mu_2-x^*}{\sigma_2}=\frac{x^*-\mu_1}{\sigma_1}$$

化简可以得到分类平面为，

$$x^*=\frac{\sigma_1\mu_2+\sigma_2\mu_1}{\sigma_1+\sigma_2}$$

c) 求解问题 b 中对应的总贝叶斯风险。

解： 代入 a) 中的风险决策公式可以得到，

$$R = \int\limits_{R_1} p(x|\omega_2)dx = \frac{1 - erf[\frac{\mu_2\sigma_1 - \mu_1\sigma_2}{\sigma_1(\sigma_1+\sigma_2)}]}{2}$$

d) 假设 $p(x|\omega_1) \sim N(0,1), p(x|\omega_2) \sim N(0.5, 0.25)$, 试求出 x^* 和总贝叶斯风险

解： 根据 b)、c) 两个问题的结果可以得出，

$$x^* = \frac{\sigma_1\mu_2 + \sigma_2\mu_1}{\sigma_1 + \sigma_2} = 0.4$$

$$R = \frac{1 - erf[\frac{\mu_2\sigma_1 - \mu_1\sigma_2}{\sigma_1(\sigma_1+\sigma_2)}]}{2} = \frac{1 - erf(0.4)}{2} = 0.2858$$

4 问题四

a) 证明最小误差可以表示为

$$P_e = \frac{1}{\sqrt{2\pi}} \int\limits_a^{\infty} e^{-\frac{\mu^2}{2}} d\mu$$

其中，$a = \frac{|\mu_1 - \mu_2|}{2\sigma}$

证： 错误率表示如下：

$$P_e = P(\omega_1) \int\limits_{-\infty}^{\theta} p(x|\omega_1)dx + P(\omega_2) \int\limits_{\theta}^{\infty} p(x|\omega_2)dx$$

由于 $P(\omega_1) = P(\omega_2), \sigma_1 = \sigma_2$，分类边界为

$$x^* = \frac{\mu_1 + \mu_2}{2}$$

对应的错误率为：

$$P_e = \frac{1}{2}(\int\limits_{-\infty}^{x^*} p(x|\omega_1)dx + \int\limits_{x^*}^{\infty} p(x|\omega_2)dx = \int\limits_{x^*}^{\infty} p(x|\omega_2)dx = \frac{1}{\sqrt{2\pi}} \int\limits_a^{\infty} e^{-\frac{\mu^2}{2}} d\mu$$

其中，

$$a = \frac{|x^* - \mu_1|}{\sigma_1} = \frac{|\mu_1 - \mu_2|}{2\sigma}$$

b) 利用不等式，

$$P_e = \frac{1}{\sqrt{2\pi}} \int\limits_a^{\infty} e^{-\frac{\mu^2}{2}} d\mu \leq \frac{1}{\sqrt{2\pi}a} e^{-\frac{a^2}{2}}$$

证明随着 $\frac{\mu_1 - \mu_2}{\sigma}$ 趋向于无穷，P_e 趋近于 0

证：

$$\lim_{a\to\infty} P_e = \lim_{a\to\infty}\frac{1}{\sqrt{2\pi}}\int_a^\infty e^{-\frac{\mu^2}{2}}d\mu \le \lim_{a\to\infty}\frac{1}{\sqrt{2\pi}a}e^{-\frac{a^2}{2}} = 0$$

5 问题五

a) 利用拉格朗日待定银子，推导合成的函数式

$$H_s = -\int p(x)[\ln p(x) - \sum_{k=0}^{q}\lambda_k b_k(x)]dx - \sum_{k=0}^{q} a_k$$

说明为什么对所有的 x，有 $a_0 = 1, b_0(x) = 1$ 成立。

证： 由于概率密度的归一化条件，存在约束，

$$\int p(x)dx = 1$$

利用拉格朗日乘子，我们可以得到，

$$\begin{aligned} H_s &= -\int p(x)\ln p(x)dx + \lambda_0[\int p(x)dx - 1] + \sum_{k=1}^{q}\lambda_k[\int b_k(x)p(x)dx - a_k] \\ &= -\int p(x)[\ln p(x) - \sum_{k=0}^{q}\lambda_k b_k(x)]dx - \sum_{k=0}^{q} a_k\lambda_k \end{aligned}$$

其中对于所有的 x 而言，有 $a_0 = b_0 = 1$。

b) 根据 H_s 对 $p(x)$ 求导，证明最大熵分布遵循

$$p(x) = \exp[\sum_{k=0}^{q}\lambda_k b_k - 1]$$

证： 为了使得熵最小，需要使得熵的导数为 0，即

$$\frac{\partial H_s}{\partial p(x)} = -\int[\ln p(x) - \sum_{k=0}^{q}\lambda_k b_k + 1]dx = 0$$

积分号内值应该为 0，即

$$\ln p(x) = \sum_{k=0}^{q}\lambda_k b_k - 1$$

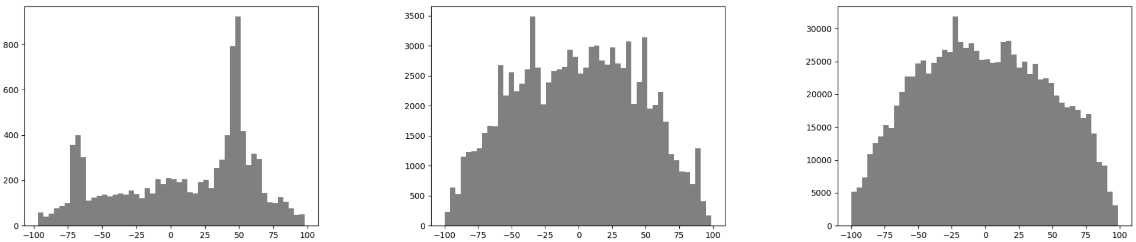
等价于

$$p(x) = \exp[\sum_{k=0}^{q}\lambda_k b_k - 1]$$

6 编程一

编程环境 Python 3.7+numpy+random(python 自带)+matplotlib(画图)

实验结果 实验结果显示图1所示，其中均值、方差显示如下表1所示：



(a) n=1e4 (b) n=1e5 (c) n=1e6

图 1 上机实验一结果图片对比

n	μ	σ
1e4	10.44	49.40
1e5	-0.97	46.22
1e6	-1.27	47.91

表 1 上机实验一结果数据

根据实验结果可以看出，随着采样数目的增多，该分布近似于正态分布 $N(0,50)$。

7 编程二

编程环境 Python 3.7+numpy(内含多维高斯分布生成器)+random(python 自带)+matplotlib(画图)

实验结果 实验结果显示图2所示。

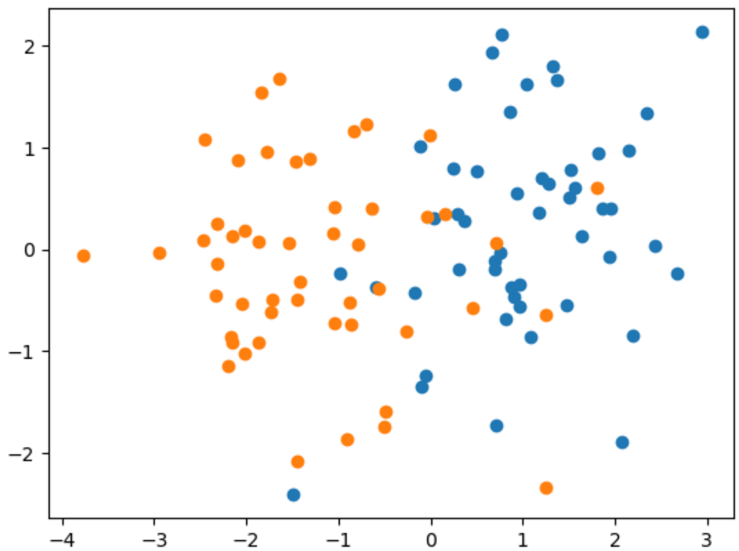


图 2 二维高斯采样

分类平面为

$$x_1 = 0$$

对应的错误率为：0.13