

第三章作业

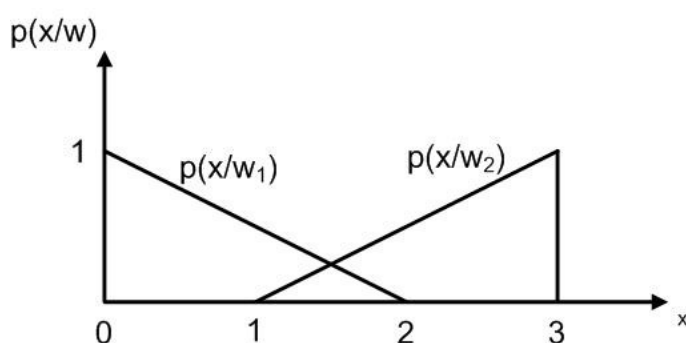
1. 已知两个一维模式类别的类概率密度函数为,

$$p(x|\omega_1) = \begin{cases} x & 0 \leq x < 1 \\ 2-x & 1 \leq x < 2 \\ 0 & \text{其他} \end{cases}, \quad p(x|\omega_2) = \begin{cases} x-1 & 1 \leq x < 2 \\ 3-x & 2 \leq x \leq 3 \\ 0 & \text{其他} \end{cases}$$

先验概率 $p(\omega_1) = 0.6, p(\omega_2) = 0.4$,

- (1) 求 Bayes 最小错误率判决函数;
- (2) 求总错误概率 p_e ;
- (3) 判断样本 $\{x_1=1.35, x_2=1.45, x_3=1.55, x_4=1.65\}$ 各属于哪一类?

2. 两个一维模式类别, 其概率密度函数如下图:



(a) 其先验概率相等, 试导出其贝叶斯最小错误率判别函数。

(b) 求出判别界面的位置。

3. 假设在某个地区细胞识别中正常 (w_1) 和异常 (w_2) 两类先验概率分别为 $p(w_1) = 0.8$, $p(w_2) = 0.2$, 现有一待识别的细胞, 其观察值为 x , 从类条件概率密度分布曲线上查得 $p(x|w_1) = 0.25$, $p(x|w_2) = 0.6$, 并且已知 $\lambda_{11} = 0$, $\lambda_{12} = 0.5$, $\lambda_{21} = 1$, $\lambda_{22} = 0$ 试对该细胞 x 用基于最小风险的贝叶斯决策。

4. 在字符检测中, 假定类型 ω_1 为字符, 类型 ω_2 为非字符, 已知先验概率

$P(\omega_1) = 0.6$ 和 $P(\omega_2) = 0.4$, 现在有两个待识样本 x_1 和 x_2 , 其类概率密度分别为:

$$p(x|\omega_1): \quad 0.8, \quad 0.1$$

$$p(x|\omega_2): \quad 0.2, \quad 0.9$$

- 1) 试用贝叶斯最小误判概率准则判决两个样本各属于哪一个类型;
- 2) 如果 λ_{12} 表示属于 ω_1 类判决 ω_2 所造成的损失, 正确判断的损失 $\lambda_{11} = \lambda_{22} = 0$, 如果 $\lambda_{12} = 4$, 试用贝叶斯最小风险准则判决两个样本均属于第一类, 误判损失 λ_{21} 应该如何设计? 请分析两种分类结果的异同及原因。

5. 随机变量 x 服从 Erlang 概率密度函数:

$$p(x, \theta) = \theta^2 x \exp(-\theta x) u(x)$$

其中 $u(x)$ 是单位阶跃函数, $u(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}$, 给定 N 个测量值 x_1, \dots, x_N , 计算 θ 最大似然估计。

6. 给定数据样本 $X = \{3, 4, 5, 5, 6, 12, 14, 14, 15, 16, 17, 18\}$, 采用 Parzen 窗估计在 5 和 14 处的密度函数 $p(x)$, 窗宽 $h_N = 4$ 。试分别计算出采用方窗和正态窗 ($\mu = 0, \sigma = 1$) 估计结果。

7. 给定两类数据样本 $\omega_1 = \{-3, -2, -1, -1, 0, 1, 1, 2, 2, 3\}$, $\omega_2 = \{1, 2, 3, 3, 4, 5, 5, 6, 6, 7\}$, 试用最小错误率贝叶斯决策判别观测样本 $x=2$ 属于哪一类? 此处两类

出现的概率 $p(w_1)$ 和 $p(w_2)$ 相等, 采用方窗函数 $\varphi(u) = \begin{cases} 1 & \text{如果 } |u| \leq 1/2 \\ 0 & \text{其它} \end{cases}$ 的

Parzen 窗估计密度方法, 窗宽 $h_N = 5$ 估计类条件概率。

● 创新思考题

下面是 10 首李白和 10 首杜甫的 5 言诗的题目，尝试利用这些诗里面的句子作为训练样本集。



杜甫

《绝句》、《春望》、《春夜喜雨》、《登岳阳楼》、《旅夜书怀》、《月夜》、《八阵图》、《月夜忆舍弟》、《归雁》、《武侯庙》、《绝句二首》



李白

《静夜思》、《夜宿山寺》、《独坐敬亭山》、《越女词》、《题情深树,寄象公》、《渌水曲》、《秋浦歌十七之十五》、《估客行》、《玉阶怨》、《劳劳亭》

问题：思考如何利用统计判别方法做出如下判决：1) “白水绕东城”该句出自谁的诗？2) 该句诗可能出现在第几句？

注意：1. 同学可以增加训练诗词，但不能采用包括该句诗词的诗作为训练样本。2. 可以采用不同的特征，比如韵律等等。