

模式识别作业一、二

数据科学与计算机学院 17大数据与人工智能

17341015 陈鸿崢

问题 1 (§2 Q2). 假设两个等概率的一维密度具有如下形式: 对任给 $i = 1, 2$ 及 $0 < b_i$, $p(x | \omega_i) \propto e^{-|x-a_i|/b_i}$ 。

- (a) 写出每个密度的解析表达式, 即对任意的 a_i 和正的 b_i , 将每个函数归一化
- (b) 计算似然比, 作为4个变量的函数
- (c) 绘出在 $a_1 = 0, b_1 = 1, a_2 = 1, b_2 = 2$ 时的似然比 $p(x | \omega_1)/p(x | \omega_2)$ 的曲线图

解答. (a) 设比例系数为 k_i , 由概率的基本性质有

$$\begin{aligned} & \int_{-\infty}^{\infty} k_i e^{-\frac{|x-a_i|}{b_i}} dx \\ &= \int_{-\infty}^{a_i} k_i e^{\frac{x-a_i}{b_i}} dx + \int_{a_i}^{\infty} k_i e^{-\frac{x-a_i}{b_i}} dx \\ &= 2k_i b_i \\ &= 1 \end{aligned}$$

进而求得 $k_i = 1/(2b_i)$, 故解析表达式为

$$p(x | \omega_i) = \frac{1}{2b_i} e^{-|x-a_i|/b_i}$$

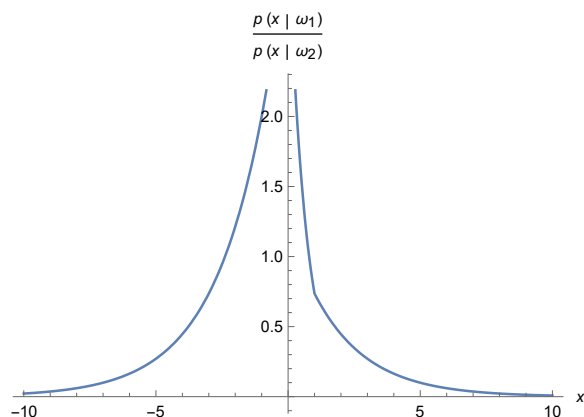
(b)

$$\frac{p(x | \omega_1)}{p(x | \omega_2)} = \frac{b_2}{b_1} e^{-\frac{|x-a_1|}{b_1} + \frac{|x-a_2|}{b_2}}$$

(c) 将 $a_1 = 0, b_1 = 1, a_2 = 1, b_2 = 2$ 代入(b)求得的式子化简得

$$\frac{p(x | \omega_1)}{p(x | \omega_2)} = 2e^{-|x| + \frac{|x-1|}{2}}$$

图像如下



问题 2 (§2 Q7). 考虑两个一维柯西分布的 *Neyman-Pearson* 准则:

$$p(x | \omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x-a_i}{b}\right)^2}, \quad i = 1, 2$$

在 0-1 误差损失下, 且为了简化, 设 $a_2 > a_1$, 宽度 b 相同, 且先验概率相等。

- 假设当一样本实际属于 ω_1 却被误认为 ω_2 的模式分类时的最大可接受误差率为 E_1 , 用所给变量确定判决边界。
- 对于此边界, 将 ω_2 错分为 ω_1 的误差率是多少?
- 在 0-1 损失率下的总误差率是多少?
- 将你的结论应用于特殊情况: $b = 1$ 且 $a_1 = -1, a_2 = 1$ 且 $E_1 = 0.1$
- 将你的结论与贝叶斯误差率 (即没有 *Neyman-Pearson* 条件) 作比较。

解答. (a) 因先验概率相等, 故 $p(x | \omega_1) = 1/2$. 设判别边界为 x^* , 则

$$\begin{aligned} E_1 &= \int_{x^*}^{\infty} p(x | \omega_1) P(\omega_1) dx \\ &= \frac{1}{2} \int_{x^*}^{\infty} \frac{1}{\pi b} \frac{1}{1 + \left(\frac{x-a_1}{b}\right)^2} dx \\ &= \frac{1}{2\pi b} \int_{u^*}^{\infty} \frac{b}{1 + u^2} du \quad u = \frac{x-a_1}{b} \\ &= \frac{1}{2\pi} \arctan u \Big|_{u^*}^{\infty} \end{aligned}$$

移项得到

$$2\pi E_1 = \frac{\pi}{2} - \arctan \frac{x^* - a_1}{b}$$

两侧同时取 \tan , 整理得

$$x^* = a_1 + \frac{b}{\tan(2\pi E_1)}$$

(b) 由题(a)的结果

$$\begin{aligned}
 E_2 &= \int_{-\infty}^{x^*} p(x | \omega_2) P(\omega_2) dx \\
 &= \frac{1}{\pi b} \int_{-\infty}^{x^*} \frac{1}{1 + \left(\frac{x-a_2}{b}\right)^2} P(\omega_2) dx \\
 &= \frac{1}{2\pi b} \int_{-\infty}^{u^*} \frac{b}{1 + u^2} du \quad u = \frac{x - a_2}{b} \\
 &= \frac{1}{2\pi} \arctan u \Big|_{-\infty}^{u^*} \\
 &= \frac{1}{4} + \frac{1}{2\pi} \arctan \frac{x^* - a_2}{b}
 \end{aligned}$$

(c) 综合题(a)和题(b)有

$$E = E_1 + E_2 = E_1 + \frac{1}{4} + \frac{1}{2\pi} \arctan \frac{x^* - a_2}{b}$$

(d) 将值代入(c)式得到 $x^* = 0.3764$, $E = 0.2613$

(e) 贝叶斯的决策边界为 $x_B^* = 0$, 进而贝叶斯误差率为

$$\begin{aligned}
 E_B &= 2 \cdot \frac{1}{2\pi b} \int_0^\infty \frac{1}{1 + \left(\frac{x-a_1}{b}\right)^2} dx \\
 &= \arctan u \Big|_0^\infty \\
 &= 0.25 < E
 \end{aligned}$$

即确实贝叶斯误差率会小于基于Neyman-Pearson准则的误差率

问题 3 (§2 Q9). 使用第7题给出的条件密度, 设类别的先验概率相等。

(a) 证明最小误差概率为

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left| \frac{a_2 - a_1}{2b} \right|$$

(b) 绘出它随 $|a_2 - a_1|/(2b)$ 变化的曲线图。

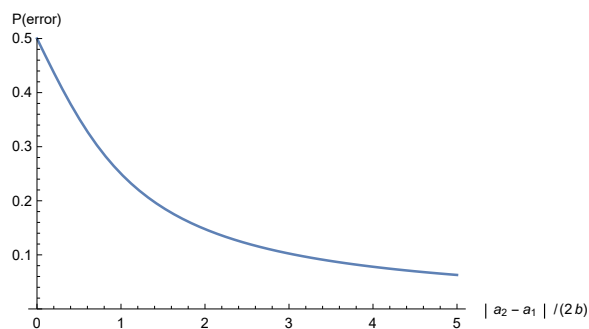
(c) $P(\text{error})$ 的最大值是多少? 在什么条件下可以达到此值? 试说明原因。

解答. (a) 不妨设 $a_2 > a_1$, 判决边界为 $(a_1 + a_2)/2$, 则误差概率为

$$\begin{aligned}
 P(\text{error}) &= \int_{-\infty}^{(a_1+a_2)/2} p(x | \omega_2) P(\omega_2) dx + \int_{(a_1+a_2)/2}^{\infty} p(x | \omega_1) P(\omega_1) dx \\
 &= \frac{1}{2\pi b} \left(\int_{-\infty}^{(a_1+a_2)/2} \frac{1}{1 + \left(\frac{x-a_2}{b}\right)^2} dx + \int_{(a_1+a_2)/2}^{\infty} \frac{1}{1 + \left(\frac{x-a_1}{b}\right)^2} dx \right) \\
 &= \frac{1}{2\pi b} \left(\arctan u \Big|_{-\infty}^{(a_1-a_2)/(2b)} + \arctan u \Big|_{(-a_1+a_2)/(2b)}^{\infty} \right) \\
 &= \frac{1}{2} - \frac{1}{\pi} \arctan \frac{a_2 - a_1}{2b}
 \end{aligned}$$

对于 $a_1 \leq a_2$ 的情形类似，故得证

(b) 如下图所示



(c) 当 $|a_2 - a_1|/(2b) = 0$ 时 $P(error)$ 达到最大值 $1/2$ ，也即两个概率分布相同 ($a_1 = a_2$)，或者两个概率分布都为常量值 ($a_1 \neq a_2, b = \infty$)

问题 4 (§2 Q23). 考虑三维正态分布 $p(\mathbf{x} | \omega) \sim N(\boldsymbol{\mu}, \Sigma)$ ，其中

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix}$$

- 求点 $\mathbf{x}_0 = [0.5 \ 0 \ 1]^T$ 处的概率密度。
- 构造白化变换 $A_\omega = \phi\Lambda^{-1/2}$ ，计算分别表示本征向量和本征值的矩阵 ϕ 和 Λ ；接下来，将此分布转换为以原点为中心协方差矩阵为单位阵的分布，即 $p(\mathbf{x} | \omega) \sim N(0, I)$ 。
- 将整个同样的转换过程应用于点 \mathbf{x}_0 以产生一变换点 \mathbf{x}_ω 。
- 通过详细计算，证明原分布中从 \mathbf{x}_0 到均值 $\boldsymbol{\mu}$ 的Mahalanobis距离与变换后的分布中从 \mathbf{x}_ω 到 $\mathbf{0}$ 的Mahalanobis距离相等。
- 概率密度在一个一般的线性变换下是否保持不变？换句话说，对于某线性变换 T ，是否有 $p(\mathbf{x}_0 | N(\boldsymbol{\mu}, \Sigma)) = p(T^T \mathbf{x}_0 | N(T^T \boldsymbol{\mu}, T^T \Sigma T))$ ？解释原因。
- 证明当把一个一般的白化变换 $A_\omega = \phi\Lambda^{-1/2}$ 应用于一个高斯分布时可保证最终分布的协方差与单位阵 I 成比例，检查变换后的分布是否仍然具有归一化特性。

解答. (a) 我们有 $d = 3$ 维，

$$|\Sigma| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{vmatrix} = 21$$

注意到 Σ 可表示为分块矩阵，我们有

$$\Sigma = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & A \end{bmatrix} \quad A^{-1} = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5/|\det A| & -2/|\det A| \\ -2/|\det A| & 5/|\det A| \end{bmatrix} = \begin{bmatrix} 5/21 & -2/21 \\ -2/21 & 5/21 \end{bmatrix}$$

进而

$$\Sigma^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & A^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5/21 & -2/21 \\ 0 & -2/21 & 5/21 \end{bmatrix}$$

故有平方Mahalanobis距离

$$\begin{aligned} & (\mathbf{x}_0 - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}) \\ &= \begin{bmatrix} 0.5 - 1 \\ 0 - 2 \\ 1 - 2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5/21 & -2/21 \\ 0 & -2/21 & 5/21 \end{bmatrix} \begin{bmatrix} 0.5 - 1 \\ 0 - 2 \\ 1 - 2 \end{bmatrix} \\ &= 1.06 \end{aligned}$$

将上述数值代入得到概率密度

$$p(\mathbf{x}_0 | \omega) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x}_0 - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}) \right] = 0.008155$$

(b) 先求出 Σ 的特征值，考虑特征方程 $|\Sigma - \lambda I| = 0$

$$\begin{vmatrix} 1 - \lambda & 0 & 0 \\ 0 & 5 - \lambda & 2 \\ 0 & 2 & 5 - \lambda \end{vmatrix} = (1 - \lambda)(3 - \lambda)(7 - \lambda) = 0$$

对于不同特征值，回代求其特征向量，并可求得其通解形式

$$\lambda_1 = 1: \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \implies \mathbf{x} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

$$\lambda_2 = 3: \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 3 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \implies \mathbf{x} = x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

$$\lambda_3 = 7: \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 7 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \implies \mathbf{x} = x_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

进而有三个正交基

$$\begin{aligned} \phi_1 &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \\ \phi_2 &= \begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^T \\ \phi_3 &= \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T \end{aligned}$$

最终得到正交特征向量矩阵

$$\phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

以及白化变换

$$A_\omega = \phi \Lambda^{-1/2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{7} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{bmatrix}$$

进而有

$$Y = A_\omega^T(\mathbf{x} - \boldsymbol{\mu}) \sim N(0, I)$$

(c) 将 \mathbf{x}_0 代入有

$$\mathbf{x}_\omega = A_\omega^T(\mathbf{x}_0 - \boldsymbol{\mu}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 1/\sqrt{14} & 1/\sqrt{14} \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -1/\sqrt{6} \\ -3/\sqrt{14} \end{bmatrix}$$

(d) 由(a)知原距离 $r^2 = 1.06$ ，而新的距离

$$r_\omega^2 = \mathbf{x}_\omega^T \mathbf{x}_\omega = 1.06$$

因此两者相等

(e) 设线性变换后的向量为 $\mathbf{x}' = T^T \mathbf{x}$ ，进而有均值

$$\boldsymbol{\mu}' = \sum_{k=1}^n \mathbf{x}'_k / n = \sum_{k=1}^n T^T \mathbf{x}_k / n = T^T \sum_{k=1}^n \mathbf{x}_k / n = T^T \boldsymbol{\mu}$$

和协方差

$$\Sigma' = \sum_{k=1}^n (\mathbf{x}'_k - \boldsymbol{\mu}')(\mathbf{x}'_k - \boldsymbol{\mu}')^T = T^T \left[\sum_{k=1}^n (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T \right] T = T^T \Sigma T$$

又 $|\Sigma'| = |T^T \Sigma T| = |\Sigma|$ ，故

$$p(\mathbf{x}_0 | N(\boldsymbol{\mu}, \Sigma)) = p(T^T \mathbf{x}_0 | N(T^T \boldsymbol{\mu}, T^T \Sigma T))$$

(f) 由课本图2.8上面的公式，我们有

$$p(A_\omega^T \mathbf{x}) \sim N(A_\omega^T \boldsymbol{\mu}, A_\omega^T \Sigma A_\omega)$$

进而可得协方差矩阵

$$\begin{aligned} A_\omega^T \Sigma A_\omega &= (\phi \Lambda^{-1/2})^T \phi \Lambda \phi^T (\phi \Lambda^{-1/2}) \\ &= \Lambda^{-1/2} (\phi^T (\phi \Lambda \phi^T) \phi) \Lambda^{-1/2} \\ &= I \end{aligned}$$

故变换后的分布依然具有归一化特性