| **THU-70250043-0, Pattern Recognition (Spring 2021)** | **Homework: 1** |

<div align="center">

Bayesian Methods

</div>

Lecturer: Changshui Zhang   `zcs@mail.tsinghua.edu.cn`

Hong Zhao   `vzhao@tsinghua.edu.cn`

Student: Jingxuan Yang   `yangjx20@mails.tsinghua.edu.cn`

Bayes Decision

1. Let the conditional densities for a two-category one-dimensional problem be given by the Cauchy distribution described as follows:

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

1.1. By explicit integration, check that the distributions are indeed normalized.

解: 直接积分可得, $\forall\, i = 1, 2$, 有

$$
\begin{aligned}
\int_{-\infty}^{\infty} p(x|\omega_i)\mathrm{d}x &= \int_{-\infty}^{\infty} \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\dfrac{x - a_i}{b}\right)^2}\mathrm{d}x \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1 + u^2}\mathrm{d}u, \quad u = \frac{x - a_i}{b} \\
&= \frac{1}{\pi} \arctan(u)\Big|_{-\infty}^{\infty} \\
&= \frac{1}{\pi}\left(\frac{\pi}{2} + \frac{\pi}{2}\right) \\
&= 1
\end{aligned}
\tag{2}
$$

故此分布函数是归一化的.

1.2. Assuming $P(\omega_1) = P(\omega_2)$, show that $P(\omega_1|x) = P(\omega_2|x)$ if $x = (a_1 + a_2)/2$, i.e., the minimum error decision boundary is a point midway between the peaks of the two distributions, regardless of b.

解: 由 Bayes 公式可知

$$P(\omega_1|x) = \frac{p(x|\omega_1)P(\omega_1)}{p(x)} \tag{3}$$

$$P(\omega_2|x) = \frac{p(x|\omega_2)P(\omega_2)}{p(x)} \tag{4}$$

令 $P(\omega_1|x) = P(\omega_2|x)$ 可得

$$\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \frac{p(x|\omega_2)P(\omega_2)}{p(x)} \tag{5}$$

又 $P(\omega_1) = P(\omega_2)$, 则

$$p(x|\omega_1) = p(x|\omega_2) \tag{6}$$

即

$$\frac{1}{\pi b} \cdot \frac{1}{1 + \left(\dfrac{x - a_1}{b}\right)^2} = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\dfrac{x - a_2}{b}\right)^2} \tag{7}$$

整理得

$$|x - a_1| = |x - a_2| \tag{8}$$

易知 $x = (a_1 + a_2)/2$ 满足上式.

1.3. Assuming $P(\omega_1) = P(\omega_2)$, plot $P(\omega_1|x)$ for the case $a_1 = 3$, $a_2 = 5$ and $b = 1$.

解: 由 Bayes 公式可得

$$
\begin{aligned}
P(\omega_1|x) &= \frac{p(x|\omega_1)P(\omega_1)}{p(x)} \\
&= \frac{p(x|\omega_1)P(\omega_1)}{p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)} \\
&= \frac{p(x|\omega_1)}{p(x|\omega_1) + p(x|\omega_2)} \\
&= \frac{\dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\dfrac{x - a_1}{b}\right)^2}}{\dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\dfrac{x - a_1}{b}\right)^2} + \dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\dfrac{x - a_2}{b}\right)^2}} \\
&= \frac{1 + \left(\dfrac{x - a_2}{b}\right)^2}{1 + \left(\dfrac{x - a_1}{b}\right)^2 + 1 + \left(\dfrac{x - a_2}{b}\right)^2}
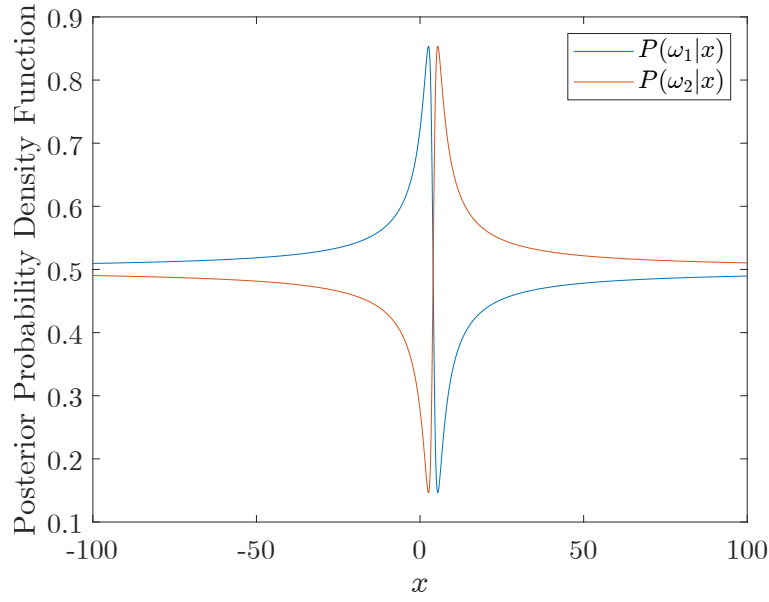\end{aligned}
\tag{9}
$$

图 1: 后验概率

易知

$$P(\omega_2|x) = 1 - P(\omega_1|x) \tag{10}$$

代入已知数值可绘图如图 1 所示.

1.4. Assuming $P(\omega_1) = P(\omega_2)$, how do $P(\omega_1|x)$ and $P(\omega_2|x)$ behave as $x \to -\infty$? $x \to +\infty$? Explain.

解: 由图 1 可知, 当 $x \to \pm\infty$ 时, $P(\omega_1|x) = P(\omega_2|x) = \dfrac{1}{2}$, 证明如下.

对式 (9) 取极限可得

$$\lim_{x \to \pm\infty} P(\omega_1|x) = \lim_{x \to \pm\infty} \frac{1 + \left(\dfrac{x - a_2}{b}\right)^2}{1 + \left(\dfrac{x - a_1}{b}\right)^2 + 1 + \left(\dfrac{x - a_2}{b}\right)^2} = \frac{1}{2} \tag{11}$$

则有

$$\lim_{x \to \pm\infty} P(\omega_2|x) = \frac{1}{2} \tag{12}$$

2. Consider Neyman-Pearson criteria for two Cauchy distributions in one dimension described in Problem 1. Assume a zero-one error loss, and for simplicity $a_2 > a_1$, the same "width" b, and equal priors.

2.1. Suppose the maximum acceptable error rate for classifying a pattern that is actually in ω_1 as if it were in ω_2 is E_1. Determine the decision boundary in terms of the variables given.

解: 本题要求考虑 Neyman-Pearson 准则, 则边界并不一定是单边边界, 以下为大部分错误解法.

设边界点为 t, 由 $a_2 > a_1$ 可知

$$R_1 = (-\infty, t), \quad R_2 = (t, \infty) \tag{13}$$

又先验概率相等, 则

$$P(\omega_1) = p(\omega_2) = \frac{1}{2} \tag{14}$$

故第一类错误率

$$
\begin{aligned}
E_1 &= \int_t^\infty p(x|\omega_1) P(\omega_1) \mathrm{d}x \\
&= \frac{1}{2} \int_t^\infty \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\dfrac{x - a_1}{b}\right)^2} \mathrm{d}x \\
&= \frac{1}{2\pi} \int_{(t-a_1)/b}^\infty \frac{1}{1 + u^2} \mathrm{d}u, \quad u = \frac{x - a_1}{b} \\
&= \frac{1}{2\pi} \arctan(u) \Big|_{(t-a_1)/b}^\infty \\
&= \frac{1}{2\pi} \left[\frac{\pi}{2} - \arctan\left(\frac{t - a_1}{b}\right) \right]
\end{aligned}
\tag{15}
$$

即

$$\arctan\left(\frac{t - a_1}{b}\right) = \frac{\pi}{2} - 2\pi E_1 \tag{16}$$

两边取正切可得

$$\frac{t - a_1}{b} = \tan\left(\frac{\pi}{2} - 2\pi E_1\right) = \cot(2\pi E_1) \tag{17}$$

故

$$t = a_1 + b \cot(2\pi E_1) = a_1 + \frac{b}{\tan(2\pi E_1)} \tag{18}$$

正确解法应为, 考虑两类错误

$$
\begin{aligned}
P_1(e) &= \int_{\mathcal{R}_2} p(x|\omega_1) \mathrm{d}x = E_1 \\
P_2(e) &= \int_{\mathcal{R}_1} p(x|\omega_2) \mathrm{d}x
\end{aligned}
\tag{19}
$$

最小化 $P_2(e)$ 等价于

$$\min \gamma = P_2(e) + \lambda(P_1(e) - E_1) \tag{20}$$

对 λ 和 $\mathcal{R}_2$ 边界求导等于 0, 有

$$\lambda = \frac{p(x|\omega_2)}{p(x|\omega_1)} = \frac{b^2 + (x - a_1)^2}{b^2 + (x - a_2)^2} \tag{21}$$

$$\int_{\mathcal{R}_2} p(x|\omega_1) \mathrm{d}x = E_1 \tag{22}$$

则决策边界为

$$t_{1,2} = \frac{(a_1 - \lambda a_2) \pm \sqrt{\lambda(a_1 - a_2)^2 - b^2(\lambda - 1)^2}}{1 - \lambda} \tag{23}$$

2.2. For this boundary, what is the error rate for classifying ω_2 as ω_1?

解: 第二类错误率为

$$
\begin{aligned}
E_2 &= \int_{-\infty}^{\top} p(x|\omega_2) P(\omega_2) \mathrm{d}x \\
&= \frac{1}{2} \int_{-\infty}^{\top} \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\dfrac{x - a_2}{b}\right)^2} \mathrm{d}x \\
&= \frac{1}{2\pi} \int_{-\infty}^{(t-a_2)/b} \frac{1}{1 + u^2} \mathrm{d}u \\
&= \frac{1}{2\pi} \left[\arctan\left(\frac{t - a_2}{b}\right) + \frac{\pi}{2} \right]
\end{aligned}
\tag{24}
$$

2.3. What is the overall error rate under zero-one loss?

解: 总错误率为

$$E = E_1 + E_2 = E_1 + \frac{1}{2\pi}\left[\arctan\left(\frac{t - a_2}{b}\right) + \frac{\pi}{2}\right] \tag{25}$$

2.4. Apply your results to the specific case $b = 1$ and $a_1 = -1, a_2 = 1$ and $E_1 = 0.1$.

解: 代入数值得边界点为

$$t = a_1 + \frac{b}{\tan(2\pi E_1)} = -1 + \frac{1}{\tan(2\pi \times 0.1)} \approx 0.376 \tag{26}$$

则总错误率为

$$
\begin{aligned}
E &= E_1 + \frac{1}{2\pi}\left[\arctan\left(\frac{t - a_2}{b}\right) + \frac{\pi}{2}\right] \\
&= 0.1 + \frac{1}{2\pi}\left[\arctan\left(\frac{0.376 - 1}{1}\right) + \frac{\pi}{2}\right] \\
&\approx 0.261
\end{aligned}
\tag{27}
$$

2.5. Compare your result to the Bayes error rate (i.e. without the Neyman-Pearson conditions).

解: 对于 Bayes 决策, 分界点为

$$t = \frac{a_1 + a_2}{2} = \frac{-1 + 1}{2} = 0 \tag{28}$$

故 Bayes 错误率为

$$
\begin{aligned}
E_B &= \int_{-\infty}^{\top} p(x|\omega_2)P(\omega_2)\mathrm{d}x + \int_{t}^{\infty} p(x|\omega_1)P(\omega_1)\mathrm{d}x \\
&= \frac{1}{2\pi}\left[\frac{\pi}{2} - \arctan\left(\frac{t-a_1}{b}\right)\right] + \frac{1}{2\pi}\left[\arctan\left(\frac{t-a_2}{b}\right) + \frac{\pi}{2}\right] \\
&= \frac{1}{2} - \frac{1}{2\pi}\arctan\left(\frac{0+1}{1}\right) + \frac{1}{2\pi}\arctan\left(\frac{0-1}{1}\right) \\
&= \frac{1}{4}
\end{aligned}
\tag{29}
$$

由 $E_B < E$ 可知 Bayes 决策错误率小于 Neyman-Pearson 决策错误率.

3. In many pattern classification problems one has the option either to assign the pattern to one of c classes, or to *reject* it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$
\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j, \quad i,j = 1,2,\cdots,c \\ \lambda_r & i = c+1 \\ \lambda_s & \text{otherwise,} \end{cases}
\tag{30}
$$

where λ_r is the loss incurred for choosing the $(c+1)$th action, rejection, and λ_s is the loss incurred for making a substitution error. Show that the minimum risk is obtained if we decide ω_i if $P(\omega_i|\boldsymbol{x}) \geqslant P(\omega_j|\boldsymbol{x})$ for all j and if $P(\omega_i|\boldsymbol{x}) \geqslant 1 - \lambda_r/\lambda_s$, and reject otherwise. What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?

解: 令

$$
m = \underset{j=1,2,\cdots,c}{\operatorname{argmax}} P(\omega_j|x)
\tag{31}
$$

即 ω_m 表示后验概率最大的类别, 分类讨论.

当选择 ω_m 时, 风险为

$$
R_m = \lambda_s \sum_{j\neq m} P(\omega_j|x) = \lambda_s[1 - P(\omega_m|x)]
\tag{32}
$$

当选择其他类别 ω_k, $k \neq m$ 时, 风险为

$$
R_k = \lambda_s \sum_{j\neq k} P(\omega_j|x) = \lambda_s[1 - P(\omega_k|x)] \geqslant \lambda_s[1 - P(\omega_m|x)]
\tag{33}
$$

当选择拒绝时, 风险为

$$
R_r = \lambda_r
\tag{34}
$$

所以最小风险为

$$
\begin{aligned}
R &= \min\{R_m, R_k, R_r\} \\
&= \min\{\lambda_s[1 - P(\omega_m|x)], \lambda_r\}
\end{aligned}
\tag{35}
$$

当 $\lambda_s[1 - P(\omega_m|x)] \leqslant \lambda_r$, 即

$$P(\omega_m|x) \geqslant 1 - \frac{\lambda_r}{\lambda_s} \tag{36}$$

时, 选择 ω_m 风险最小, 否则选择拒绝风险最小.

若 $\lambda_r = 0$, 则选择拒绝风险最小, 会拒绝所有样本.

若 $\lambda_r > \lambda_s$, 则选择最大后验概率类别 ω_m 风险最小, 不会拒绝任何样本.

4. Consider a two-category classification problem in two dimensions. Given the following conditions, calculate the Bayes decision boundary respectively and plot them in the coordinate space (with computer or by hand). You should indicate the mean point of each category.

4.1. We set $p(\boldsymbol{x}|\omega_1) \sim N(\boldsymbol{0}, \boldsymbol{I})$, $p(\boldsymbol{x}|\omega_2) \sim N\left((1,1)^\top, \boldsymbol{I}\right)$ and $P(\omega_1) = P(\omega_2) = 0.5$.

解: 忽略类别无关项, 判别函数为

$$g_i(\boldsymbol{x}) = -\frac{1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top(\boldsymbol{x} - \boldsymbol{\mu}_i) \tag{37}$$

决策面满足

$$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x}) \tag{38}$$

故

$$-\frac{1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top(\boldsymbol{x} - \boldsymbol{\mu}_i) = -\frac{1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^\top(\boldsymbol{x} - \boldsymbol{\mu}_j) \tag{39}$$

化简得

$$-2\boldsymbol{x}^\top \boldsymbol{\mu}_i + \boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i = -2\boldsymbol{x}^\top \boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top \boldsymbol{\mu}_j \tag{40}$$

即

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top \boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i - \boldsymbol{\mu}_j^\top \boldsymbol{\mu}_j) = 0 \tag{41}$$

代入数据得

$$x_1 + x_2 - 1 = 0 \tag{42}$$

决策面如图 2 所示.

4.2. We set $p(\boldsymbol{x}|\omega_1) \sim N(\boldsymbol{0}, \boldsymbol{I})$, $p(\boldsymbol{x}|\omega_2) \sim N\left((1,1)^\top, \boldsymbol{I}\right)$ and $P(\omega_1) = 0.2, P(\omega_2) = 0.8$.

解: 忽略类别无关项, 判别函数为

$$g_i(\boldsymbol{x}) = -\frac{1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top(\boldsymbol{x} - \boldsymbol{\mu}_i) + \ln P(\omega_i) \tag{43}$$

决策面满足

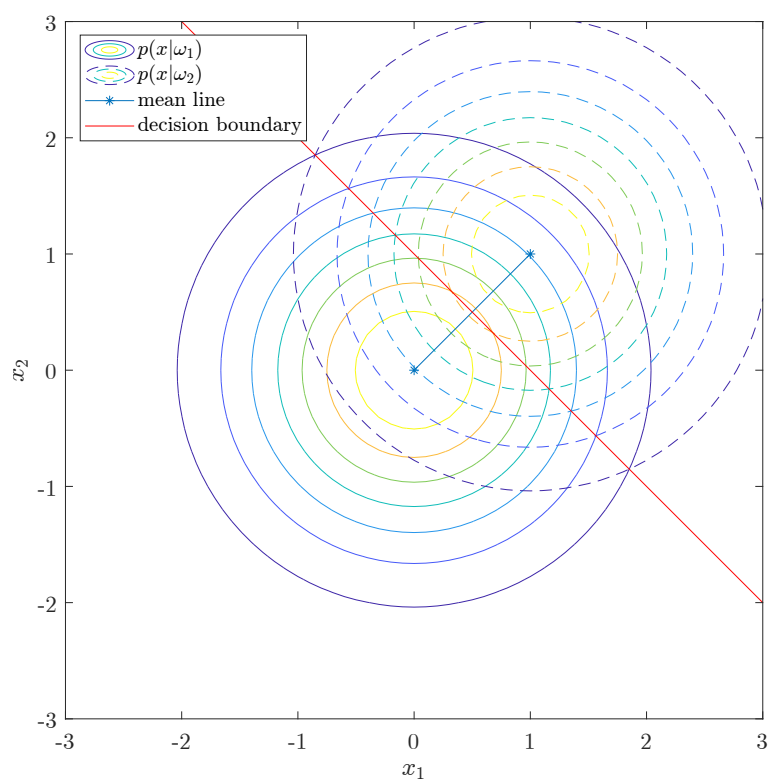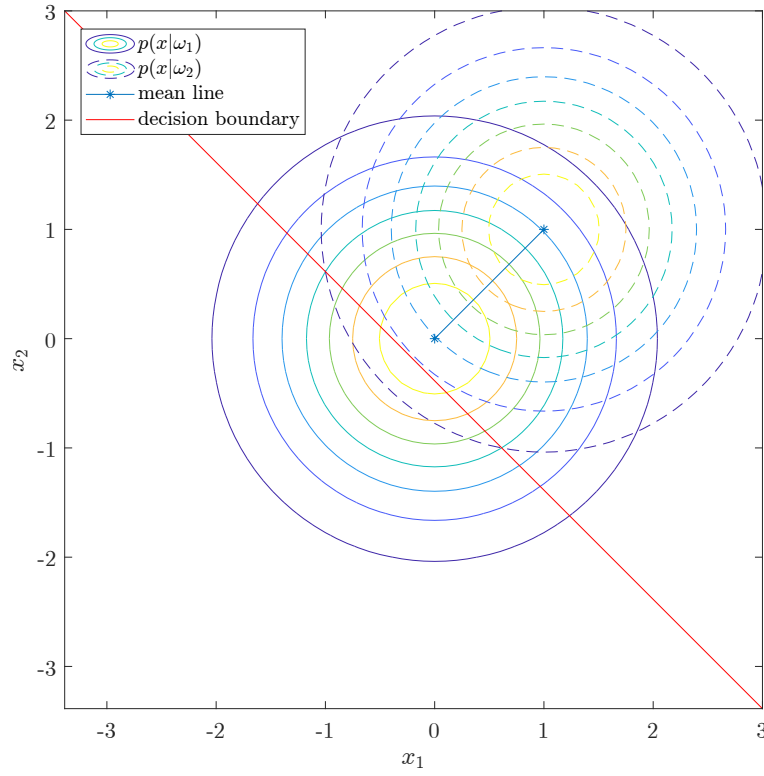$$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x}) \tag{44}$$

图 2: 决策面

图 3: 决策面

故

$$-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top(\boldsymbol{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i) = -\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^\top(\boldsymbol{x}-\boldsymbol{\mu}_j) + \ln P(\omega_j) \tag{45}$$

化简得

$$-2\boldsymbol{x}^\top\boldsymbol{\mu}_i + \boldsymbol{\mu}_i^\top\boldsymbol{\mu}_i - 2\sigma^2 \ln P(\omega_i) = -2\boldsymbol{x}^\top\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top\boldsymbol{\mu}_j - 2\sigma^2 \ln P(\omega_j) \tag{46}$$

即

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_i^\top\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^\top\boldsymbol{\mu}_j) + \sigma^2 \ln \frac{P(\omega_i)}{P(\omega_j)} = 0 \tag{47}$$

代入数据得

$$x_1 + x_2 - 1 + \ln 4 = 0 \tag{48}$$

决策面如图 3 所示.

4.3. We set $p(\boldsymbol{x}|\omega_1) \sim N\left(\mathbf{0}, \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\right)$, $p(\boldsymbol{x}|\omega_2) \sim N\left((1,1)^\top, \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\right)$ and $P(\omega_1) = P(\omega_2) = 0.5$.

解: 忽略类别无关项, 判别函数为

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) \tag{49}$$
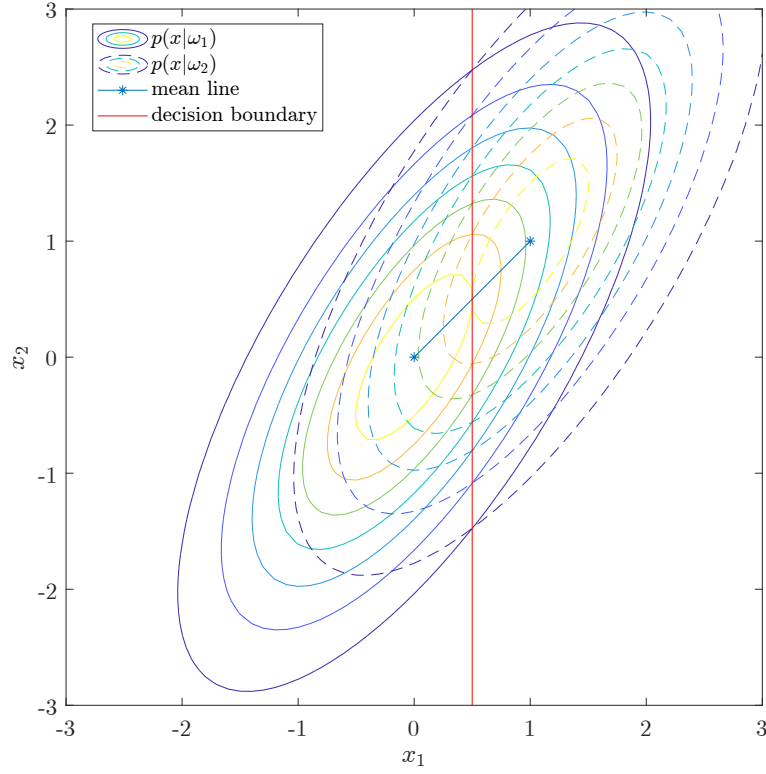
图 4: 决策面

决策面满足

$$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x}) \tag{50}$$

故

$$-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_j) \tag{51}$$

化简得

$$-2\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i = -2\boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j \tag{52}$$

即

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j) = 0 \tag{53}$$

代入数据得

$$x_1 - \frac{1}{2} = 0 \tag{54}$$

决策面如图 4 所示.

4.4. We set $p(\boldsymbol{x}|\omega_1) \sim N\left(\boldsymbol{0}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)$, $p(\boldsymbol{x}|\omega_2) \sim N\left((1,1)^\top, \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\right)$ and $P(\omega_1) = P(\omega_2) = 0.5$.

解: 忽略类别无关项, 判别函数为

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| \tag{55}$$

决策面满足

$$g_i(\boldsymbol{x}) = g_j(\boldsymbol{x}) \tag{56}$$

故

$$-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_j) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_j| \tag{57}$$

化简得

$$\boldsymbol{x}^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{x} - 2\boldsymbol{x}^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i + \ln|\boldsymbol{\Sigma}_i| = \boldsymbol{x}^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{x} - 2\boldsymbol{x}^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j + \ln|\boldsymbol{\Sigma}_j| \tag{58}$$

即

$$\boldsymbol{x}^\top (\boldsymbol{\Sigma}_j^{-1} - \boldsymbol{\Sigma}_i^{-1})\boldsymbol{x} + 2(\boldsymbol{\mu}_i \boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\mu}_j \boldsymbol{\Sigma}_j^{-1})^\top \boldsymbol{x} - (\boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\mu}_j) - (\ln|\boldsymbol{\Sigma}_i| - \ln|\boldsymbol{\Sigma}_j|) = 0 \tag{59}$$

代入数据得

$$\frac{1}{2}x_1^2 - \frac{1}{2}x_2^2 + x_1 + 2x_2 - \frac{3}{2} = 0 \tag{60}$$

化简得

$$x_1^2 - x_2^2 + 2x_1 + 4x_2 - 3 = 0 \tag{61}$$

即

$$(x_1 + 1)^2 - (x_2 - 2)^2 = 0 \tag{62}$$

即

$$|x_1 + 1| = |x_2 - 2| \tag{63}$$

决策面如图 5 所示.

Naive Bayes

5. Consider a simple learning problem of determining whether Alice and Bob from CA will go to hiking or not, $Y : Hike \in \{T, F\}$ given the weather conditions $X_1 : Sunny \in \{T, F\}$ and $X_2 : Windy \in \{T, F\}$ by a Naive Bayes classifier. Using training data, we estimated the parameters $P(Hike) = 0.5$, $P(Sunny|Hike) = 0.9$, $P(Windy|\neg Hike) = 0.8$, $P(Windy|Hike) = 0.3$ and $P(Sunny|\neg Hike) = 0.4$. Assume that the true distributions of X_1, X_2, and Y satisfy the Naive Bayes assumption of conditional independence with the above parameters.
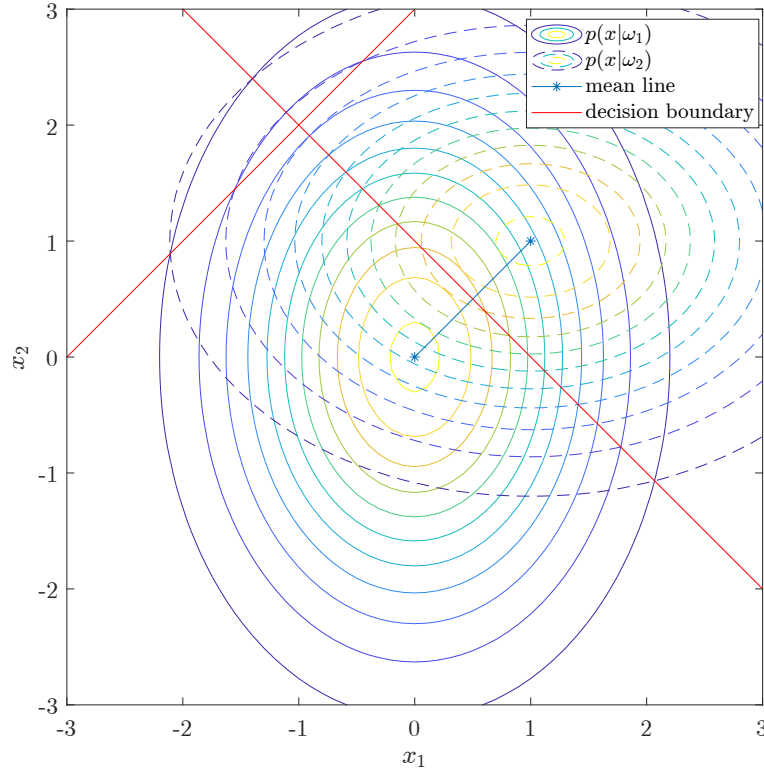
图 5: 决策面

5.1. Assume Sunny and Windy are truly independent given Hike. Write down the Naive Bayes decision rule for this problem using both attributes Sunny and Windy.

解: 已知 $P(Hike) = 0.5$, 则 $P(\neg Hike) = P(Hike) = 0.5$.

由 Bayes 公式结合条件独立的假设可得

$$
\begin{aligned}
P(Hike|Sunny, Windy) &= \frac{0.9 \times 0.3}{0.9 \times 0.3 + 0.4 \times 0.8} = 0.4576 \\
P(Hike|Sunny, \neg Windy) &= \frac{0.9 \times 0.7}{0.9 \times 0.7 + 0.4 \times 0.2} = 0.8873 \\
P(Hike|\neg Sunny, Windy) &= \frac{0.1 \times 0.3}{0.1 \times 0.3 + 0.6 \times 0.8} = 0.0588 \\
P(Hike|\neg Sunny, \neg Windy) &= \frac{0.1 \times 0.7}{0.1 \times 0.7 + 0.6 \times 0.2} = 0.3684
\end{aligned}
\tag{64}
$$

所以

$$
\begin{aligned}
P(Hike|Sunny, Windy) &< P(\neg Hike|Sunny, Windy) \\
P(Hike|Sunny, \neg Windy) &> P(\neg Hike|Sunny, \neg Windy) \\
P(Hike|\neg Sunny, Windy) &< P(\neg Hike|\neg Sunny, Windy) \\
P(Hike|\neg Sunny, \neg Windy) &< P(\neg Hike|\neg Sunny, \neg Windy)
\end{aligned}
\tag{65}
$$

因此 Bayes 决策规则为

$$(Sunny, Windy) \rightarrow \neg Hike$$
$$(Sunny, \neg Windy) \rightarrow Hike$$
$$(\neg Sunny, Windy) \rightarrow \neg Hike \tag{66}$$
$$(\neg Sunny, \neg Windy) \rightarrow \neg Hike$$

5.2. Given the decision rule above, what is the expected error rate of the Naive Bayes classifier? (The expected error rate is the probability that each class generates an observation where the decision rule is incorrect.)

解: 由 Bayes 决策规则可知, 错误率为

$$\begin{aligned} E &= P(Hike)[1 - P(Sunny, \neg Windy|Hike)] + P(\neg Hike)P(Sunny, \neg Windy|\neg Hike) \\ &= 0.5 \times (1 - 0.9 \times 0.7) + 0.5 \times 0.4 \times 0.2 \\ &= 0.225 \end{aligned} \tag{67}$$

5.3. What is the joint probability that Alice and Bob go to hiking and the weather is sunny and windy, that is $P(Sunny, Windy, Hike)$?

解: 由条件乘法公式与条件独立性可知, 联合概率密度为

$$\begin{aligned} P(Sunny, Windy, Hike) &= P(Hike)P(Sunny, Windy|Hike) \\ &= P(Hike)P(Sunny|Hike)P(Windy|Hike) \\ &= 0.5 \times 0.9 \times 0.3 \\ &= 0.135 \end{aligned} \tag{68}$$

5.4. Next, suppose that we gather more information about weather conditions and introduce a new feature denoting whether the weather is X_3: Rainy or not. Assume that each day the weather in CA can be either Rainy or Sunny. That is, it can not be both Sunny and Rainy (similarly, it can not be $\neg Sunny$ and $\neg Rainy$). In the above new case, are any of the Naive Bayes assumptions violated? Why (not)? What is the joint probability that Alice and Bob go to hiking and the weather is sunny, windy and not rainy, that is $P(Sunny, Windy, \neg Rainy, Hike)$?

解: 确有 Naive Bayes 的假设不成立, 事实上

$$P(Sunny, Windy, Rainy|Hike) = 0 \tag{69}$$

而

$$\begin{aligned} &P(Sunny|Hike)P(Windy|Hike)P(Rainy|Hike) \\ &= P(Sunny|Hike)P(Windy|Hike)P(\neg Sunny|Hike) \\ &= 0.9 \times 0.3 \times 0.1 \\ &= 0.027 \end{aligned} \tag{70}$$

所以

$$P(Sunny, Windy, Rainy | Hike) \neq P(Sunny | Hike)P(Windy | Hike)P(Rainy | Hike) \tag{71}$$

故 Naive Bayes 的条件独立性假设不成立.

由题干假设以及条件乘法公式可知, 联合概率密度为

$$\begin{aligned}
P(Sunny, Windy, \neg Rainy, Hike) &= P(Sunny, Windy, Hike) \\
&= P(Hike)P(Sunny, Windy | Hike) \\
&= P(Hike)P(Sunny | Hike)P(Windy | Hike) \\
&= 0.5 \times 0.9 \times 0.3 \\
&= 0.135
\end{aligned} \tag{72}$$

5.5. What is the expected error rate when the Naive Bayes classier uses all three attributes? Will the performance of Naive Bayes be improved by observing the new attribute Rainy? Explain why.

解: 考虑三种因素的 Bayes 决策规则为

$$\begin{aligned}
(Sunny, Windy, \neg Rainy) &\rightarrow \neg Hike \\
(Sunny, \neg Windy, \neg Rainy) &\rightarrow Hike \\
(\neg Sunny, Windy, Rainy) &\rightarrow \neg Hike \\
(\neg Sunny, \neg Windy, Rainy) &\rightarrow \neg Hike
\end{aligned} \tag{73}$$

错误率为

$$\begin{aligned}
\tilde{E} &= P(Hike)[1 - P(Sunny, \neg Windy, \neg Rainy | Hike)] + P(\neg Hike)P(Sunny, \neg Windy, \neg Rainy | \neg Hike) \\
&= P(Hike)[1 - P(Sunny, \neg Windy | Hike)] + P(\neg Hike)P(Sunny, \neg Windy | \neg Hike) \\
&= 0.5 \times (1 - 0.9 \times 0.7) + 0.5 \times 0.4 \times 0.2 \\
&= 0.225
\end{aligned} \tag{74}$$

由 $\tilde{E} = E$ 可知, 增加新要素 Rainy 并不能改进 Naive Bayes 的性能, 因为 Rainy 信息可以直接从 Sunny 信息获得, 本质上并未真正增加新的信息.

ROC Analysis

6. The purpose of this assignment is to learn the theory and practical applications of ROC analysis. Therefore you should first read a paper about ROC curves [1] that is available in this homework zip (especially pay attention to section 1-5 and 7-8). In the following task you will work on a given theoretical example of two

| 表 1: 分类器 1 | | | | 表 2: 分类器 2 | | | |
Class	Score	FP Rate	TP Rate	Class	Score	FP Rate	TP Rate
1	0.6	0	1/4	0	0.8	1/4	0
0	0.51	1/4	1/4	1	0.68	1/4	1/4
1	0.5	1/4	2/4	0	0.53	2/4	1/4
1	0.4	1/4	3/4	1	0.4	2/4	2/4
0	0.33	2/4	3/4	1	0.22	2/4	3/4
0	0.3	3/4	3/4	0	0.11	3/4	3/4
1	0.22	3/4	1	0	0.1	1	3/4
0	0.2	1	1	1	0.04	1	1

classifier (C_1 and C_2). We have a set of samples that we wish to classify in one of two classes and a ground truth class of each sample (denoted as 0 and 1). For each sample a classifier gives a score based on which we can determine to which class should be the sample belong to (score closer to 0 means class 0, score closer to 1 means class 1). Below are the results for 8 samples, their ground truth values (S_{id}) and the score values for both classifiers (S_{C_1} and S_{C_2}).

$$
\begin{aligned}
S_{\mathrm{id}} &= \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \\
S_{C_1} &= \begin{bmatrix} 0.5 & 0.3 & 0.6 & 0.22 & 0.4 & 0.51 & 0.2 & 0.33 \end{bmatrix} \\
S_{C_2} &= \begin{bmatrix} 0.04 & 0.1 & 0.68 & 0.22 & 0.4 & 0.11 & 0.8 & 0.53 \end{bmatrix}
\end{aligned}
\tag{75}
$$

For the example above calculate (by hand) and draw the ROC curves for classifier C_1 as well as classifier C_2. Also calculate the area under the curve (AUC) for both classifier. For the classifier C_1 select a decision threshold (working points) $\theta_1 = 0.33$. For the classifier C_2 select a decision threshold $\theta_2 = 0.1$. Based on theory from [1] decide which classifier is better in the selected working points and motivate your decision. Which working point is optimal for each classifier?

解: 分别对两个分类器的分数进行排序, 计算相应的真阳性率和假阳性率, 分类器 1 的计算结果如表 1 所示, 分类器 2 的计算结果如表 2 所示.

根据计算结果, 绘制两个分类器的 ROC 曲线分别如图 6 和图 7 所示.

由图 6 和图 7 可得两个分类器的 AUC 分别为

$$
\begin{aligned}
\mathrm{AUC}_1 &= 1 - 5 \times \frac{1}{4} \times \frac{1}{4} = \frac{11}{16} \\
\mathrm{AUC}_2 &= 7 \times \frac{1}{4} \times \frac{1}{4} = \frac{7}{16}
\end{aligned}
\tag{76}
$$

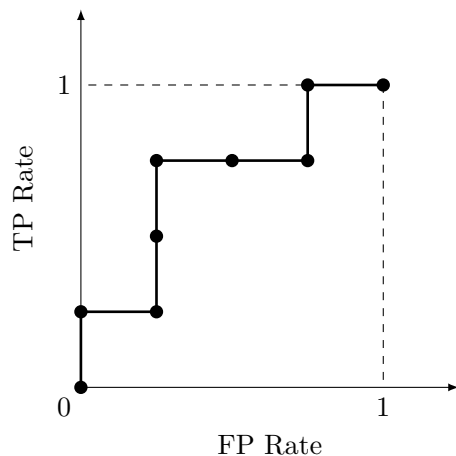由 $\theta_1 = 0.33$ 和 $\theta_2 = 0.1$ 可知分类器 1 和分类器 2 的工作点分别为 (2/4, 3/4) 和 (1, 3/4), 两个工作点的真阳
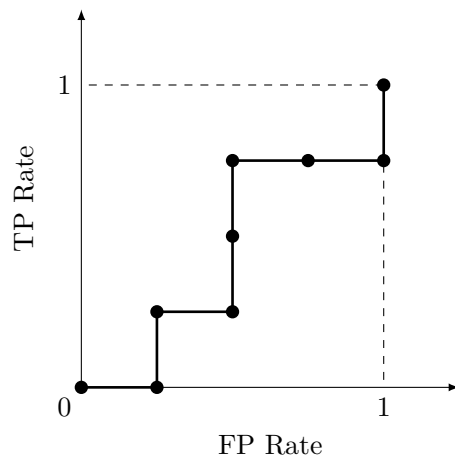
图 6: 分类器 1 ROC 曲线

图 7: 分类器 2 ROC 曲线

性率相等, 但是分类器 1 的假阳性率更小, 故在选定工作点处分类器 1 更好.

越靠近 ROC 曲线左上角的工作点越好, 故对分类器 1 来说, 工作点 (1/4, 3/4) 最好, 相应的阈值为 $\theta_1 = 0.4$; 对分类器 2 来说, 工作点 (2/4, 3/4) 最好, 相应的阈值为 $\theta_2 = 0.22$.

参考文献

[1] T. Fawcett. An introduction to ROC analysis. Pattern Recognition Letters, 27(8):861-874, 2006.