模式识别：作业 #1

201928014629008
牛李金梁

2019 年 10 月 2 日

Question 1

$$P(error|x) = \begin{cases} P(\omega_1|x), \text{if we decide } \omega_2 \\ P(\omega_1|x), \text{if we decide } \omega_1 \end{cases}$$

(a)

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

(b)

$$\begin{aligned} \frac{dP(error)}{d\theta} &= P(\omega_1)(p(\theta|\omega_1) - p(-\infty|\omega_1)) + P(\omega_2)(p(\infty|\omega_2) - p(\theta|\omega_2)) \\ &= P(\omega_1)p(\theta|\omega_1) - P(\omega_2)p(\theta|\omega_2) \end{aligned}$$

最小化 $P(error)$ 的一个必要条件就是 $\frac{dP(error)}{d\theta} = 0$，即，

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

(c)

(b) 中等式无法唯一确定 θ。因为 2 个概率密度可以存在多个交点满足这一条件，从而可以确定多个 θ。

(d)

$$\frac{dP(error)^2}{d^2\theta} = P(\omega_1)\frac{dp(\theta|\omega_1)}{d\theta} - P(\omega_2)\frac{dp(\theta|\omega_2)}{d\theta}$$

要使 $P(error)$ 取最大，等价于 $\frac{dP(error)}{d\theta} = 0$ 且 $\frac{dP(error)^2}{d^2\theta} < 0$。
因此可以取 $P(\omega_1) = P(\omega_2) = \frac{1}{2}$，$\theta$= 0，$P(x|\omega_1) \sim N(-1,1)$，$P(x|\omega_2) \sim N(1,1)$

Question 2

$$R(\alpha_i|\vec{x}) = \sum_{j=1}^{n} \lambda(\alpha_i|\omega_j) P(\omega_j|\vec{x})$$

根据上式，$i = 1, 2, \dots, n$ 时，

$$\begin{aligned} R_i(\vec{x}) &= \lambda_s \sum_{j \neq i} P(\omega_j|\vec{x}) \\ &= \lambda_s [1 - P(\omega_i|\vec{x})] \end{aligned}$$

拒识时，

$$R_i(\vec{x}) = \lambda_r$$

通过最小决策风险决策为 i 时，有，

$$\lambda_s[1 - P(\omega_i|\vec{x})] \leq \lambda_r$$

即，

$$P(\omega_i|\vec{x}) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

否则拒识。
当 $\lambda_r = 0$ 时，会一直拒识，$\lambda_r > \lambda_s$ 时，不会出现拒识。

Question 4

(a)

因为 $P(\omega_1) = P(\omega_2)$
所以最小错误判别依据 $P(\omega_i|x)$ 只与 $p(x|\omega_i)$ 有关
又 $p(x|\omega_i) \sim N(\mu_i, \sigma^2)$
所以判别条件为：如果 $|x - \mu_1| < |x - \mu_2|$，选择 ω_1，否则选择 ω_2。
不妨设 $\mu_1 < \mu_2$

$$\begin{aligned} P_e &= P(|x - \mu_1| > |x - \mu_2|\,|\omega_1)P(\omega_1) + P(|x - \mu_1| < |x - \mu_2|\,|\omega_2)P(\omega_2) \\ &= \frac{1}{2\sqrt{2\pi}\sigma} \left(\int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \mathsf{exp} \left[-\frac{1}{2} \Big(\frac{x - \mu_1}{\sigma} \Big)^2 \right] dx + \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} \mathsf{exp} \left[-\frac{1}{2} \Big(\frac{x - \mu_2}{\sigma} \Big)^2 \right] dx \right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} \mathsf{exp} \left[-\frac{1}{2} \Big(\frac{x - \mu_1}{\sigma} \Big)^2 \right] dx \end{aligned}$$

令 $\frac{x-\mu_1}{\sigma} = \mu$，上式可化为，

$$\frac{1}{\sqrt{2\pi}} \int_a^{\infty} \mathsf{e}^{-\frac{\mu^2}{2}} d\mu$$

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

(b)

由 P_e 定义知，$P_e \geq 0$
则存在不等式 $0 \leq P_e \leq \lim\limits_{a\to\infty} \frac{1}{\sqrt{2\pi}a} e^{-\frac{a^2}{2}}$
而 $\frac{|\mu_1-\mu_2|}{\sigma} \to \infty$ 可以得到 $a \to \infty$
又 $\lim\limits_{a\to\infty} \frac{1}{\sqrt{2\pi}a} e^{-\frac{a^2}{2}} = 0$
由夹逼准则可得 $\frac{|\mu_1-\mu_2|}{\sigma}$ 趋近 ∞ 时，P_e 趋近于 0。

Question 5

(a)

概率的归一化条件为 $\int p(x)dx = 1$，这是一个隐含的约束。
因而有 $a_0 = 1$ 且 $b_0(x) = 1$ 对于所有的 x
因此，由拉格朗日乘子法有，

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

(b)

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

令 $\frac{\partial H_s}{\partial p(x)} = 0$，有

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

即，

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

Question 6

决策面方程形式为，

$$\vec{w}^t(\vec{x} - \vec{x_0}) = 0$$

其中，

$$\vec{w} = \Sigma^{-1}(\vec{\mu_1} - \vec{\mu_2})$$

$$\overrightarrow{x_0}=\frac{1}{2}\left(\overrightarrow{\mu_1}+\overrightarrow{\mu_2}\right)-\frac{\ln\frac{P(\omega_1)}{P(\omega_2)}}{\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)}$$

决策面不在 $\overrightarrow{\mu_1}$、$\overrightarrow{\mu_2}$ 间，有 $\overrightarrow{w}^t(\overrightarrow{\mu_1}-\overrightarrow{x_0})$ 与 $\overrightarrow{w}^t(\overrightarrow{\mu_2}-\overrightarrow{x_0})$ 同号。

$$\begin{aligned}\overrightarrow{w}^t(\overrightarrow{\mu_1}-\overrightarrow{x_0})&=(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left[\overrightarrow{\mu_1}-\frac{1}{2}\left(\overrightarrow{\mu_1}+\overrightarrow{\mu_2}\right)+\frac{\ln\frac{P(\omega_1)}{P(\omega_2)}}{\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)\right]\\&=\frac{1}{2}(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)+\ln\frac{P(\omega_1)}{P(\omega_2)}\end{aligned}$$

同理，

$$\overrightarrow{w}^t(\overrightarrow{\mu_2}-\overrightarrow{x_0})=-\frac{1}{2}(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)+\ln\frac{P(\omega_1)}{P(\omega_2)}$$

① 二者均大于 0，

$$(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)<2\ln\frac{P(\omega_1)}{P(\omega_2)}$$
$$(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)>-2\ln\frac{P(\omega_1)}{P(\omega_2)}$$

当 $P(\omega_1)>P(\omega_2)$ 时有解。
② 二者均大于 0，

$$(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)<-2\ln\frac{P(\omega_1)}{P(\omega_2)}$$
$$(\overrightarrow{\mu_1}-\overrightarrow{\mu_2})^t\Sigma^{-1}\left(\overrightarrow{\mu_1}-\overrightarrow{\mu_2}\right)>2\ln\frac{P(\omega_1)}{P(\omega_2)}$$

当 $P(\omega_1)\leq P(\omega_2)$ 时有解。

编程题

环境为 python3.6
所需依赖：numpy、scipy、matplotlib

Question 1

a 题调用的是 np.random.random_integers() 这一函数。
b 题直接使用上述函数取随机数，不过该函数默认的是闭区间，因此 n 的取值需从 1 开始。

考虑到中心极限定理的独立同分布前提。本实验先随机取出 x_l、x_u 与 n，之后进行若干（如 10000）次采样，每次从均匀分布 $U\left(x_l,x_u\right)$ 中产生 n 个样本，再对这些样本求平均作为本次采样的结果。最终将这些样本绘制成直方图。

从结果图中可以看出，大量独立随机变量的平均近似为高斯分布，且样本越多越接近高斯分布。

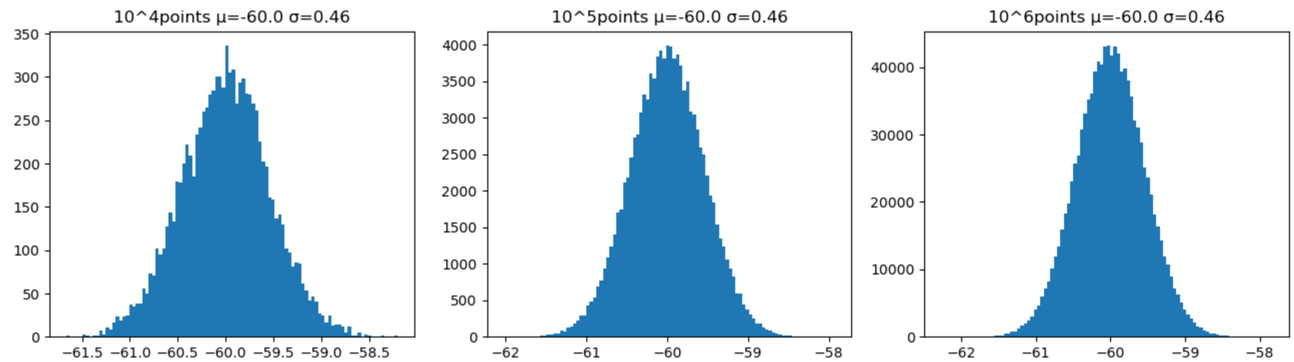

图 1: 结果

Question 2

a 题调用的是 np.random.multivariate_normal() 这一函数，该函数中的维数是由均值向量决定的，因此做了一个判断来保证维数和均值维度相同。

b 题画出了二元正态分布的 3D 图形以及贝叶斯决策面。画 3D 图形需要先生成定义域中的散点，通过 np.meshgrid() 构建网格然后求出对应的 z 值。贝叶斯决策面则利用公式先求出 x_0 和 w，然后利用上述方法绘制。结果如下：

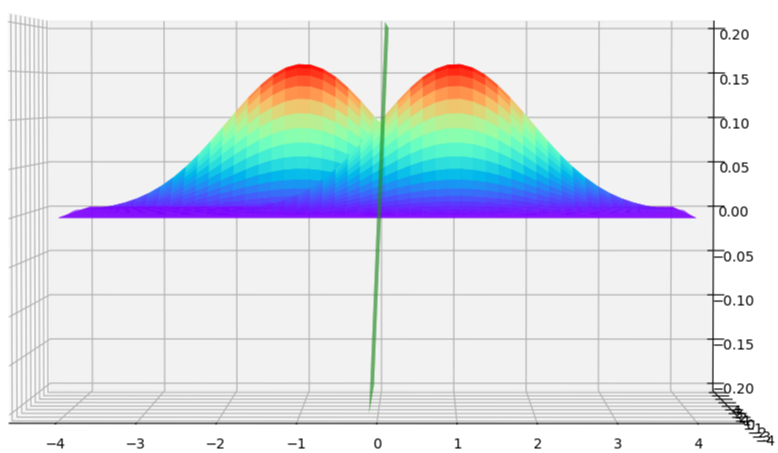

图 2: 两个二元正态分布及其贝叶斯决策面

通过上图可以看出，贝叶斯决策面在 $z=0$ 的投影是 $x=0$，因此对于 c 题，通过 a 中函数生成两类各 50 个点后，绘制出决策面，通过判决函数与 0 比较计算其分类结果。经验误差指的是误分点的百分比，通过计算大约在 0.12 上下。

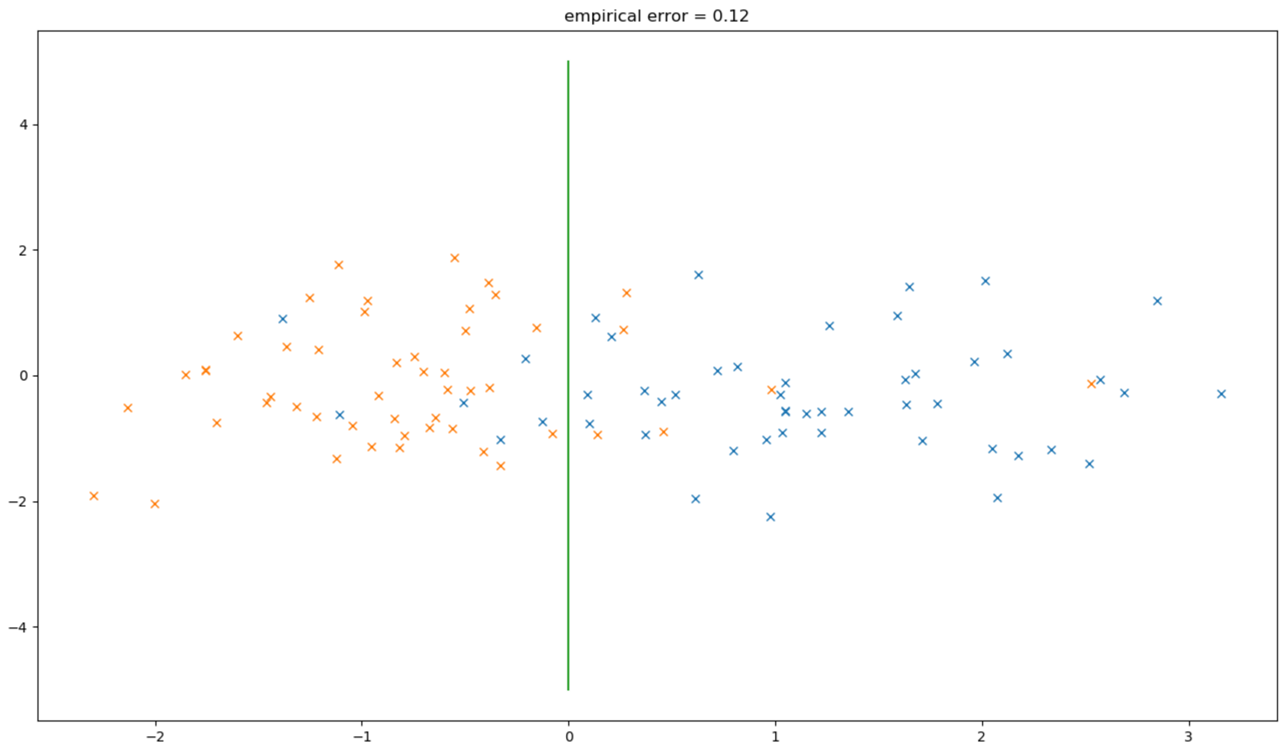

图 3： 二元正态分布及其经验误差