# Assignment 1

## Question 1

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 of rejection is not too high, rejection may be a desirable action. Let

$$\lambda\left(\alpha_i \mid w_j\right)=\begin{cases}0, & i=j \quad i, j=1, \ldots, c \\ \lambda_r, & i=c+1 \\ \lambda_s, & \text{otherwise}\end{cases}$$

where $\lambda_r$ is the loss incurred for choosing the $(c+1)$ the action, rejection, and $\lambda_s$ is the loss incurred for making a substitution error. Show that the minimum risk is obtained if we decide $w_i$ if $P\left(w_i \mid x\right) \geq P\left(w_j \mid x\right)$ for all $j$ and if $P\left(w_i \mid x\right) \geq 1-\frac{\lambda_r}{\lambda_s}$, and reject otherwise. What happens if $\lambda_r=0$? And what happens if $\lambda_r>\lambda_s$?

### *Answer1:*

(1 )对于$i=1, \ldots, c$ 而言：

$$\begin{aligned} R\left(\alpha_i \mid \mathbf{x}\right) & =\sum_{j=1}^c \lambda\left(\alpha_i \mid \omega_j\right) P\left(\omega_j \mid \mathbf{x}\right) \\ & =\lambda_s \sum_{j=1, j \neq i}^c P\left(\omega_j \mid \mathbf{x}\right) \\ & =\lambda_s\left[1-P\left(\omega_i \mid \mathbf{x}\right)\right]\end{aligned}$$

对于$i=c+1$而言：

$$R\left(\alpha_{c+1} \mid \mathbf{x}\right)=\lambda_r$$

所以当选择$\omega_i$ 时，

$$R\left(\alpha_i \mid \mathbf{x}\right) \leq R\left(\alpha_{c+1} \mid \mathbf{x}\right)$$

$$\begin{aligned} \lambda s[1-P(wi \mid x)] & \leq \lambda r \\ 1-P(wi \mid x) & \leq \frac{\lambda r}{\lambda s} \\ P(wi \mid x) & \geqslant 1-\frac{\lambda r}{\lambda s}\end{aligned}$$

因此当$R\left(\alpha_i \mid \mathbf{x}\right) \leq R\left(\alpha_{c+1} \mid \mathbf{x}\right)$即$P\left(\omega_i \mid \mathbf{x}\right) \geq 1-\frac{\lambda_r}{\lambda_e}$时选择$wi$，不满足则拒绝识别。

(2) 如果 $\lambda_r=0$, 将一直拒识。

(3) 如果 $\lambda_r>\lambda_s$, 将永不拒识。

# Question 2

Let $p\left(x \mid w_i\right) \sim \mathcal{N}\left(\mu_i, \sigma^2\right)$ for a two-category one-dimensional problem with $p\left(w_1\right)=p\left(w_2\right)=0.5$ (a) Show that the minimum probability of error is given by

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

where $\alpha=\frac{\left|\mu_1-\mu_2\right|}{2 \sigma}$ (b) Use the inequality

$$P_e \leq \frac{1}{\sqrt{2 \pi} a} e^{-\frac{a^2}{2}}$$

to show that $P_e$ goes to zero as $\frac{\left|\mu_1-\mu_2\right|}{\sigma}$ goes to infinity.

## *Answer2:*

(a) 证明：

利用似然比计算决策面：对于一个二分类问题而言，如果$x$属于$w1$，则似然比满足如下条件：

$$\frac{p\left(x \mid \omega_1\right)}{p\left(x \mid \omega_2\right)}>\frac{P\left(\omega_2\right)}{P\left(\omega_1\right)}$$

$\because P(w_1)=P(w_2)=1/2$，且 $P\left(x \mid \omega_i\right)$ 服从 $N\left(\mu_l, \sigma^2\right)$，即：

$$P\left(x \mid \omega_i\right)=\frac{1}{\sqrt{2 \pi} \sigma} e^{-\frac{(x-\mu)^2}{2 \sigma}}$$

则根据似然比决策，对于一个给定的特征$x$，判决为$w1$ 时有：

$$\begin{gathered}
\frac{\frac{1}{\sigma \sqrt{2 \pi}} e^{-\frac{1}{2 \sigma^2}\left(x-\mu_1\right)^2}}{\frac{1}{\sigma \sqrt{2 \pi}} e^{-\frac{1}{2 \sigma^2}\left(x-\mu_2\right)^2}}>\frac{1/2}{1/2} \\
\frac{1}{\sigma \sqrt{2 \pi}} e^{-\frac{1}{2 \sigma^2}\left(x-\mu_1\right)^2}>\frac{1}{\sigma \sqrt{2 \pi}} e^{-\frac{1}{2 \sigma^2}\left(x-\mu_2\right)^2} \\
\left(x-\mu_1\right)^2>\left(x-\mu_2\right)^2 \\
x^2-2 x \mu_1+\mu_1^2>x^2-2 x \mu_2+\mu_2^2 \\
\left(\mu_2-\mu_1\right)\left(2 x-\left(\mu_2+\mu_1\right)\right)<0 \\
x<\frac{\left(\mu_1+\mu_2\right)}{2}, x \text{判决为} w_1 \\
x>\frac{\left(\mu_1+\mu_2\right)}{2}, x \text{判决为} w_2
\end{gathered}$$

则决策边界为：

$$x=\frac{u_1+u_2}{2}$$

将二分类器划分为$R_1$和$R_2$两个区域，则错误分类可能有以下两种形式出现：

1. 真实类别为$w_1$ 而被分为$R_2$ 2. 真实类别为$w_2$而被分为$R_1$

因此误差概率为：

$$\begin{aligned}
P(\text{error}) &= P\left(\mathbf{x} \in \mathcal{R}_2, \omega_1\right) + P\left(\mathbf{x} \in \mathcal{R}_1, \omega_2\right) \\
&= P\left(\mathbf{x} \in \mathcal{R}_2 \mid \omega_1\right) P\left(\omega_1\right) + P\left(\mathbf{x} \in \mathcal{R}_1 \mid \omega_2\right) P\left(\omega_2\right) \\
&= \int_{\mathcal{R}_2} p\left(\mathbf{x} \mid \omega_1\right) P\left(\omega_1\right) d\mathbf{x} + \int_{\mathcal{R}_1} p\left(\mathbf{x} \mid \omega_2\right) P\left(\omega_2\right) d\mathbf{x} \\
&= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}\sigma} \left( \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} dx + \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}} dx \right) \\
&\text{另 } \mu = \frac{x-\mu_i}{\sigma}, i = 1, 2 \\
&\text{将 } \frac{\mu_1+\mu_2}{2} \text{带入} \frac{x-\mu_i}{\sigma}, i = 1, 2, \text{则得新边界} \frac{\mu 1-\mu 2}{2\sigma} \\
&= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \left( \int_{\frac{\mu 1-\mu 2}{2\sigma}}^{\infty} e^{-\frac{u^2}{2}} du + \int_{-\infty}^{\frac{\mu 1-\mu 2}{2\sigma}} e^{-\frac{u^2}{2}} du \right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\left|\frac{\mu_2-\mu_1}{2\sigma}\right|}^{\infty} e^{-\frac{\mu^2}{2}} du \\
&= \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{\mu^2}{2}} du, \text{其中} \alpha = \left|\frac{\mu 2-\mu 1}{2\sigma}\right|
\end{aligned}$$

即，最小误差概率为：

$$\begin{aligned}
P_e &= \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{\mu^2}{2}} d\mu \\
\alpha &= \frac{|\mu_1-\mu_2|}{2\sigma}
\end{aligned}$$

(b) 证明：当$P_e$为0时，$\left|\frac{\mu 1+\mu 2}{\sigma}\right|$趋于无穷大

$$\begin{aligned}
&\because P_e = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{\mu^2}{2}} d\mu, \text{ 且 } P_e \leq \frac{1}{\sqrt{2\pi} a} e^{-\frac{a^2}{2}} \\
&\therefore P_e = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{\mu^2}{2}} d\mu \leq \frac{1}{\sqrt{2\pi} a} e^{-\frac{a^2}{2}} \\
&\because \lim_{a\to\infty} \frac{1}{a\sqrt{2\pi}} e^{-\frac{1}{2}a^2} = 0 \\
&\therefore P_e = \frac{1}{\sqrt{2\pi}} \int_{a}^{\infty} e^{-\frac{1}{2}a^2} da \leq 0 \\
&\text{此时积分无限趋近于}0\text{，则下界}\alpha\text{逼近于}\infty \\
&\text{当}P_e\text{为}0\text{时，}\left|\frac{\mu 1+\mu 2}{\sigma}\right|\text{趋于无穷大即为所证}
\end{aligned}$$

# Question 3:

To classify a feature vector $x \in \mathcal{R}^d$ in a task of $c$ classes, we assume that for each class, **the prior is same** and the **class conditional probability density is a Gaussian distribution.** (a) Write the mathematical form of the conditional probability density function

(b) Write the discriminant function of minimum error rate in the following two cases: (a) class covariance matrices are unequal;(b) class covariance matrices are same.

(c) For the quadratic discriminant function based on Gaussian probability density, it becomes incalculable when the covariance matrix is singular. Name two ways to overcome the singularity.

## *Answer 3:*

(a) 因为类条件概率密度服从高斯分布，所以条件概率密度函数的数学表达形式如下：

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

其中$x$是一个$d$维的列向量，$\mu$是$d$维的均值向量，$\Sigma$是$d\times d$的协方差矩阵，$|\Sigma|$和$\Sigma^{-1}$分别是其行列式的值和逆，$(x-\mu)^t$是$(x-\mu)$的转置。

(b) 写出最小错误率的判别函数两种情况：（1）类协方差矩阵不相等；（2）类协方差矩阵相同。

最小误差概率判别函数可以简化为如下公式：

$$g_i(x)=\ln p\left(x\mid\omega_i\right)+\ln p\left(\omega_i\right)$$

$\because p(x|w_i)$是多元正态分布，则：

$$g_i(x)=-\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_i|+\ln P\left(\omega_i\right)$$

**情况一：类协方差矩阵不相等**

$\frac{d}{2}\ln 2\pi$与$i$无关，是不关紧要的附加常量，可以被省略。因此可以将判别函数简化为如下形式:

$$g_i(x)=-\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)-\frac{1}{2}\ln|\Sigma_i|+\ln P\left(\omega_i\right)$$

令：

$$\begin{aligned}
&\mathbf{W}_i=-\tfrac{1}{2}\mathbf{\Sigma}_i^{-1}\\
&\mathbf{w}_i=\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i\\
&w_{i0}=-\tfrac{1}{2}\boldsymbol{\mu}_i'\mathbf{\Sigma}_i^{-1}\boldsymbol{\mu}_i-\tfrac{1}{2}\ln|\mathbf{\Sigma}_i|+\ln P\left(\omega_i\right)
\end{aligned}$$

则：

$$g_i(x)=x^tW_ix+w_i^tx+w_{i0}$$

**情况二：类协方差相同**

$|\Sigma_i|$和$(d/2)\ln 2\pi$两项与$i$无关，是不关紧要的附加常量，可以被省略。且根据题目可知所有 c 类别的先验概率 $P\left(\omega_i\right)$ 都相同,那么 $\ln P\left(\omega_i\right)$ 项也可被省略。则判别函数可以被简化为如下形式：

$$g_i(\mathbf{x})=-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)'\mathbf{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)$$

令：

$$\mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$$

则：

$$g_i(\mathbf{x}) = \mathbf{w}_i^t\mathbf{x} + w_{i0}$$

(c) 基于高斯概率的二次判别函数：当协方差矩阵为奇异时，它变得无法计算，说出两种克服奇异性的方法。

克服奇异性的方法:

1. 降低维度，剔除特征为0得数据
2. 求矩阵得伪逆矩
3. 奇异值分解

# Question 4:

Suppose we have two normal distributions with **the same covariance** but **different means**: $\mathcal{N}(\mu_1,\Sigma)$ and $\mathcal{N}(\mu_2,\Sigma)$. In terms of their prior probabilities $P(w_1)$ and $P(w_2)$, state the condition that Bayes decision boundary does not pass between the two means.

# *Answer 4:*

因为两个分布均为高斯分布，根据贝叶斯决策该问题的边界是线性的，决策面的位置为:

$$\mathbf{w}^t = \Sigma^{-1}(\mu_1-\mu 2)$$

其中：

$$\begin{aligned}\mathbf{w} &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\ \mathbf{x}_o &= \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)-\frac{\ln[P(\omega_1)/P(\omega_2)]}{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\end{aligned}$$

贝叶斯决策边界不经过均值$\mu_1,\mu_2$，可以理解为决策面同向。即$\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o)$和$\mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o)$同方向:

$$\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o)>0 \quad \text{and} \quad \mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o)>0$$

或者:

$$\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o)<0 \quad \text{and} \quad \mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o)<0$$

将$\mathbf{w}$和$\mathbf{x}_0$带入,则可以将条件转变为如下:

$$\begin{aligned}\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o) &= (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1-\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\right)-\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]\\ &= \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)-\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]\\ \mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_0) &= (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1-\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\right)-\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]\\ &= -\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)-\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]\end{aligned}$$

当属于$\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_0)>0$ and $\mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o)>0$的情况时:

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)>2\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]$$

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)<-2\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]$$

当属于$\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o)<0$ and $\mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o)<0$的情况时:

$$\begin{aligned}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)<2\ln\left[\tfrac{P(\omega_1)}{P(\omega_2)}\right] \text{ and}\\ (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)>-2\ln\left[\tfrac{P(\omega_1)}{P(\omega_2)}\right]\end{aligned}$$

根据上述两种情况，要使Bayes判定边界不在两个平均值之间通过的条件可以表述如下: **情况1:**

$P(\omega_1)\leq P(\omega_2)$时:

当 $(\mu_1-\mu_2)^t\Sigma^{-1}(\mu_1-\mu_2)<2\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]$， $\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o)>0$ 且$\mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o)>0$时成立。

**情况2:**

$P(\omega_1) > P(\omega_2)$.时：

当$(\mu_1-\mu_2)^t\Sigma^{-1}(\mu_1-\mu_2) < 2\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]$， $\mathbf{w}^t(\boldsymbol{\mu}_1-\mathbf{x}_o) < 0$ 且 $\mathbf{w}^t(\boldsymbol{\mu}_2-\mathbf{x}_o) < 0$时成立。

# Question 5:

Maximum likelihood methods apply to estimate of prior probability as well. Let samples be drawn by successive, independent selections of a state of nature $w_i$ with unknown probability $P(w_i)$. Let $z_{ik}=1$ if the state of nature for the $k$ -th sample is $w_i$ and $z_{ik}=0$ otherwise. (a) Show that

$$P(z_{i1},\ldots,z_{in}\mid P(w_i))=\prod_{k=1}^{n}P(w_i)^{z_{ik}}(1-P(w_i))^{1-z_{ik}}$$

(b) Show that the maximum likelihood estimate for $P(w_i)$ is

$$\hat{P}(w_i)=\frac{1}{n}\sum_{k=1}^{n}z_{ik}$$

# *Answer 5*:

(a)证：

因为假设样本是连续独立地从自然状态$w_i$中抽取的，每一个自然状态的概率为$P(w_i)$，则：

$$z_{ik}=\begin{cases}1 & \text{第 } k^{th} \text{个样本的自然状态为 } \omega_i\\ 0 & \text{否则}\end{cases}$$

通过以概率$P(\omega_i)$连续选择自然状态$\omega_i$来绘制样本,则

$$\Pr[z_{ik}=1\mid P(\omega_i)]=P(\omega_i)$$

$$\Pr[z_{ik}=0\mid P(\omega_i)]=1-P(\omega_i)$$

这两个方程可以统一为：

$$P(z_{ik}\mid P(\omega_i))=[P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}$$

根据最大似然估计的基本原理得：

$$\begin{aligned}P(z_{i1},\cdots,z_{in}\mid P(\omega_i))&=\prod_{k=1}^{n}P(z_{ik}\mid P(\omega_i))\\&=\prod_{k=1}^{n}[P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}\end{aligned}$$

(b)证：

$P(\omega_i)$ 的对数似然函数为：

$$\begin{aligned}l(P(\omega_i))&=\ln P(z_{i1},\cdots,z_{in}\mid P(\omega_i))\\&=\ln\left[\prod_{k=1}^{n}[P(\omega_i)]^{z_{ik}}[1-P(\omega_i)]^{1-z_{ik}}\right]\\&=\sum_{k=1}^{n}[z_{ik}\ln P(\omega_i)+(1-z_{ik})\ln(1-P(\omega_i))]\end{aligned}$$

对上面式子关于$P(w_i)$求导，得到了一组求解最大似然估计值$P(w_i)$的必要条件：

$$\nabla_{P(\omega_i)}l(P(\omega_i))=\frac{1}{P(\omega_i)}\sum_{k=1}^{n}z_{ik}-\frac{1}{1-P(\omega_i)}\sum_{k=1}^{n}(1-z_{ik})=0$$

求解上式：

$$\left(1-\hat{P}\left(\omega_{i}\right)\right) \sum_{k=1}^{n} z_{ik}=\hat{P}\left(\omega_{i}\right) \sum_{k=1}^{n}\left(1-z_{ik}\right)$$

$$\sum_{k=1}^{n} z_{ik}=\hat{P}\left(\omega_{i}\right) \sum_{k=1}^{n} z_{ik}+n \hat{P}\left(\omega_{i}\right)-\hat{P}\left(\omega_{i}\right) \sum_{k=1}^{n} z_{ik}$$
$$\sum_{k=1}^{n} z_{ik}=n \hat{P}\left(\omega_{i}\right)$$

$$\hat{P}\left(\omega_{i}\right)=\frac{1}{n} \sum_{k=1}^{n} z_{ik}$$

即为所证。

# Question 6:

Let the sample mean $\hat{\mu}_n$ and the sample covariance matrix $C_n$ for a set of $nd$ -dimensional samples $x_1, \ldots, x_n$ be defined by

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad C_n = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{\mu}_n)(x_k - \hat{\mu}_n)^T$$

(a) Show that alternative, recursive techniques for calculating $\hat{\mu}_n$ and $C_n$ based on the successive addition of new samples $x_{n+1}$ can be derived using the recursion relations

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)$$

and

$$C_{n+1} = \frac{n-1}{n}C_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)(x_{n+1} - \hat{\mu}_n)^T$$

(b) Discuss the computational complexity of finding $\hat{\mu}_n$ and $C_n$ by the recursive methods.

# *Answer 6:*

(a):

1. 由题目可知:

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

则对于新样本$x_{n+1}$而言：

$$\begin{aligned}
\mu^{\wedge}_{n+1} &= \frac{1}{n+1}\sum_{k=1}^{n+1} x_k \\
&= \frac{1}{n+1}\left(\sum_{k=1}^{n} x_k + x_{n+1}\right) \\
&= \frac{1}{n+1}(n\mu_n + x_{n+1}) \\
&= \frac{n}{n+1}\mu_n + \frac{1}{n+1}x_{n+1} \\
&= \frac{(n+1)-1}{n+1}\mu_n + \frac{1}{n+1}x_{n+1} \\
&= \mu_n + \frac{1}{n+1}(x_{n+1} - \mu_n)
\end{aligned}$$

即可以基于新样本 $x_{n+1}$ 连续相加， 递归计算出 $\widehat{\mu}_{n+1}$。

2.根据题意可知：

$$C_n = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{\mu}_n)(x_k - \hat{\mu}_n)^T$$

则对于新样本$x_{n+1}$而言：

$$
\begin{aligned}
C_{n+1} &= \frac{1}{n}\sum_{k=1}^{n+1}\left(x_k - u_{n+1}^{\wedge}\right)\left(x_k - u_{n+1}^{\wedge}\right)^T \\
&= \frac{1}{n}\sum_{k=1}^{n+1}\left[x_k - \hat{u}_n - \frac{1}{n+1}\left(x_{n+1} - \hat{u}_n\right)\right]\left[x_k - \hat{u}_n - \frac{1}{n+1}\left(x_{n+1} - u_n\right)\right]^T \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(x_k - \hat{u}_n\right)\left(x_k - \hat{u}_n\right)^T - \frac{1}{n(n+1)}\sum_{k=1}^{n}\left(x_{n+1} - \hat{u}_n\right)^T\left(x_k - \hat{u}_n\right) - \frac{1}{n(n+1)}\sum_{k=1}^{n}\left(x_{n+1} - \hat{u}_n\right)\left(x_k - \hat{u}_n\right)^T \\
&\quad + \frac{1}{n(n+1)^2}\sum_{k=1}^{n}\left(x_{n+1} - \hat{u}_n\right)\left(x_{n+1} - \hat{u}_n\right)^T + \frac{n}{(n+1)^2}\left(x_{n+1} - \hat{u}_n\right)\left(x_{n+1} - \hat{u}_n\right)^T \\
&= \frac{n-1}{n}C_n - \frac{1}{n(n+1)}\left[\sum_{k=1}^{n}\left(x_{n+1} - \hat{u}_n\right)^T\left(x_k - \hat{u}_n\right) - \sum_{k=1}^{n}\left(x_{n+1} - \hat{u}_n\right)\left(x_k - \hat{u}_n\right)^T\right] + \\
&\quad \frac{1}{n+1}\left(x_{n+1} - \hat{u}_n\right)\left(x_{n+1} - \hat{u}_n\right)^T \\
&= \frac{n-1}{n}C_n - \frac{1}{n(n+1)}\left[\left(x_{n+1} - \hat{u}_n\right)^T\left(nu_n - n\hat{u}_n\right) - \sum_{k=1}^{n}\left(x_{n+1} - u_n\right)\left(nu_n - n\hat{u}_n\right)^T\right] + \\
&\quad \frac{1}{n+1}\left(x_{n+1} - \hat{u}_n\right)\left(x_{n+1} - u_n\right)^T \\
&= \frac{n-1}{n}C_n + \frac{1}{n+1}\left(x_{n+1} - \hat{u}_n\right)\left(x_{n+1} - \hat{u}_n\right)^T
\end{aligned}
$$

即可以基于新样本 $x_{n+1}$ 连续相加， 递归计算出 $C_{n+1}$。

(b) 因为由已知条件可知：

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

所以 $\hat{\mu}_n$ 的计算复杂度为 $O(nd)$ 因为由已知条件可知：

$$C_n = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k - \hat{u}_n\right)\left(x_k - \hat{u}_n\right)^T$$

所以 $\hat{\mu}_n$ 的计算复杂度为 $O\left(nd^2\right)$