作业1　　　　　　　　　　　　　　　　　　　　　　　　　　　　史文纬

1、 证：对于多维正态随机变量 X,Y， X Y不相关 $\Longleftrightarrow$ X Y独立

(1) $\Longleftarrow$

若 X Y 独立，则有 $E(XY)=E(X)E(Y)$ 协方差 $Cov(XY)=E(XY)-E(X)E(Y)=0$

相关系数 $\rho=\dfrac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=0$. 所以不相关

(2) $\Longrightarrow$

正态随机矢量的概率密度表达式

$$f(\vec{x})=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})'\Sigma^{-1}(\vec{x}-\vec{\mu})\right]$$

其中 $\vec{\mu}=E(\vec{x})$ 　 $\Sigma=\begin{pmatrix}\sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{n1}^2\\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{n2}^2\\ \vdots & \vdots & & \vdots\\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nn}^2\end{pmatrix}$

二维正态分布的概率密度函数

$$p(x_1,x_2)=\frac{1}{2\pi|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})'\Sigma^{-1}(\vec{x}-\vec{\mu})\right)$$

其中，$\Sigma=\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix} \Longrightarrow |\Sigma|=\sigma_1^2\sigma_2^2-\sigma_{12}^2=\sigma_1^2\sigma_2^2\left(1-\boxed{\dfrac{\sigma_{12}^2}{\sigma_1^2\sigma_2^2}}\right)=\sigma_1^2\sigma_2^2(1-\rho^2)$

$$(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})=\frac{(x_1-\mu_1)^2\sigma_2^2-2(x_1-\mu_1)(x_2-\mu_1)\sigma_{12}+(x_2-\mu_2)^2\sigma_1^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}$$

$$\text{故 } p(X,Y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\,\frac{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_2-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right)+\left(\frac{Y-\mu_2}{\sigma_2}\right)^2}{1-\rho^2}\right]$$

而X Y的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$f(y)=\frac{1}{\sqrt{2\pi}\sigma_2^2}e^{-\frac{(\cancel{x}-\mu_2)^2}{2\sigma_2^2}}$$

若 $\rho=0$，则 $f(x,y)=f(x)\cdot f(y)$，即不相关得独立.

综合(1)(2)，原命题得证.

2.

设 $X\sim N_p(\mu,\Sigma)$，B为 $s\times p$ 常数阵，d为 $s\times 1$ 常向量，令 $Z=BX+d$

则 $Z\sim N_s(B\mu+d,B\Sigma B')$，即多元正态随机变量线性变换仍为正态

证明：因为 $\Sigma\geq 0$，所以存在 $p\times q$ 矩阵A，使得 $\Sigma=AA'$

已知 $X\sim N_p(\mu,\Sigma)$，由定义可知：$X=AU+\mu$

其中 $U=(U_1,\cdots U_q)'$，且 $U_1\cdots U_q$ 相互独立同 $N(0,1)$ 分布

$$Z=BX+d\overset{d}{=}B(AU+\mu)+d$$

$$=(BA)\cancel{A}U+(B\mu+d)$$

$\Rightarrow\quad Z\sim N_s(B\mu+d,(BA)(BA)')$

即 $Z\sim N_s(B\mu+d,B\Sigma B')$

原命题得证。

作业1 史文纬 2015910018

3.证明：

设X(有p个分量)为多元正态随机矢量，即 $X \sim N(\mu, \Sigma)$
证明其分量线性组合为正态随机变量 即证明 对于非奇异阵C

① $Y = CX$ 的分布是 $N(C\mu, C\Sigma C')$

把x的概率密度 $n(x|\mu,\Sigma)$ 作变换

$$x = C^{-1}y \qquad ②$$

再乘以变换②的雅可比行列式

$$\mathrm{mod}|C^{-1}| = \frac{1}{\mathrm{mod}|C|} = \sqrt{\frac{1}{|C|^2}} = \sqrt{\frac{|\Sigma|}{|C|\cdot|\Sigma|\cdot|C'|}}$$

$$= \frac{|\Sigma|^{\frac{1}{2}}}{|C\Sigma C'|^{\frac{1}{2}}} \qquad ③$$

得到Y的概率密度. $n(x|\mu,\Sigma)$ 的指数中的二次型是

$$Q = (x-\mu)'\Sigma^{-1}(x-\mu) \qquad ④$$

将②代入④得

$$Q = (C^{-1}y-\mu)'\Sigma^{-1}(C^{-1}y-\mu)$$

$$= (C^{-1}y - C^{-1}C\mu)'\Sigma^{-1}(C^{-1}y - C^{-1}C\mu)$$

$$= [C^{-1}(y-C\mu)]'\Sigma^{-1}[C^{-1}(y-C\mu)]$$

$$= (y-C\mu)'(C^{-1})'\Sigma^{-1}C^{-1}(y-C\mu)$$

$$= (y-C\mu)'(C\Sigma C')^{-1}(y-C\mu) \qquad ⑤$$

因此Y的密度为

$$n(C^{-1}y|\mu,\Sigma)\,\mathrm{mod}|C|^{-1} \qquad ⑥$$

$$= (2\pi)^{-\frac{1}{2}p}|C\Sigma C'|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(y-C\mu)'(C\Sigma C')^{-1}(y-C\mu)\right]$$

$$= n(y|C\mu, C\Sigma C')$$

得证。

作业3

史文纬 2015910001

1、 w_1 。w_2 的判决界面方程为

$$\ln P(w_1) - \ln P(w_2) + (\vec{\mu_1} - \vec{\mu_2})' \Sigma^{-1} \vec{x} - \frac{1}{2} \vec{\mu_1} \Sigma^{-1} \vec{\mu_1} + \frac{1}{2} \vec{\mu_2} \Sigma^{-1} \vec{\mu_2} = 0$$

$$\vec{\mu_1} = (1, 1)^T \quad \vec{\mu_2} = (5, 5)^T \quad P(w_1) = P(w_2) = \frac{1}{2}$$

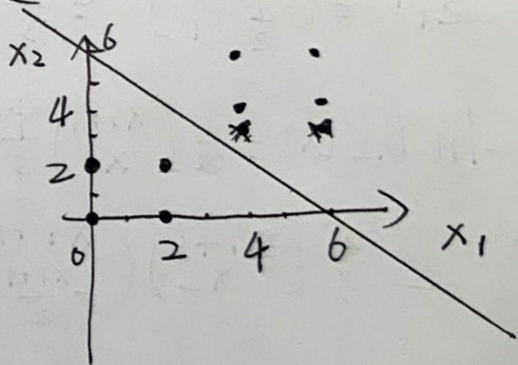
由 $\Sigma_i = E[(x - \mu_i)(x - \mu_i)^T]$

代入得 $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

将分步所得结果汇总

$$\ln(\tfrac{1}{2}) - \ln(\tfrac{1}{2}) + (-4, -4) \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \vec{x} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 5 \\ 5 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 5 \end{bmatrix} = 0$$



化简得： $(-4, -4)\vec{x} - 1 + 25 = 0$

即判别面方程为 $x_1 + x_2 = 6$

20159100078
史文纬

2\. (a) 由已知

$$p(\vec{x}|w_1) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}-1\\0\end{bmatrix}\right)' I \left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}-1\\0\end{bmatrix}\right)\right]$$

$$= \frac{1}{2\pi}\exp\left[-\frac{(x_1+1)^2+x_2^2}{2}\right]$$

$$p(\vec{x}|w_2) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}1\\0\end{bmatrix}\right)' I \left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}1\\0\end{bmatrix}\right)\right]$$

$$= \frac{1}{2\pi}\exp\left[-\frac{(x_1-1)^2+x_2^2}{2}\right]$$

$$l_{12} = \frac{p(\vec{x}|w_1)}{p(\vec{x}|w_2)}$$

$$h(x) = -\ln l_{12} = -\ln p(\vec{x}|w_1) + \ln p(\vec{x}|w_2)$$

$$= (x_1+1)^2-(x_1-1)^2 = 4x_1$$

$$\ln\frac{P(w_2)}{P(w_1)} = \ln 1 = 0$$

故若 $x_1<0$ 则 $x\in w_1$，若 $x_1>0$ 则 $x\in w_2$

(b) $|\Sigma_1| = \begin{vmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{vmatrix} = \frac{3}{4}$ $|\Sigma_2| = \begin{vmatrix}1 & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{vmatrix} = \frac{3}{4}$

$$\Sigma_1^{-1} = \begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{pmatrix}^{-1} = \frac{4}{3}\begin{pmatrix}1 & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{pmatrix} \quad \Sigma_2^{-1} = \begin{pmatrix}1 & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{pmatrix}^{-1} = \frac{4}{3}\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{pmatrix}$$

$$-\ln l_{12}(x) = \frac{1}{2}\begin{pmatrix}x_1+1\\x_2\end{pmatrix}\frac{4}{3}\begin{pmatrix}1 & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{pmatrix}\begin{pmatrix}x_1+1\\x_2\end{pmatrix} - \frac{1}{2}\begin{pmatrix}x_1-1\\x_2\end{pmatrix}\frac{4}{3}\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{pmatrix}\begin{pmatrix}x_1-1\\x_2\end{pmatrix}$$

$$= \frac{2}{3}\begin{pmatrix}x_1+1\\x_2\end{pmatrix}\begin{pmatrix}x_1+1-\frac{x_2}{2}\\ -\frac{x_1+1}{2}+x_2\end{pmatrix} - \frac{2}{3}\begin{pmatrix}x_1-1\\x_2\end{pmatrix}\begin{pmatrix}x_1-1+\frac{x_2}{2}\\ \frac{x_1-1}{2}+x_2\end{pmatrix}$$

$$= \frac{4}{3}(2x_1 - x_1x_2)$$

所以若 $2x_1-x_1x_2<0$，则 $\vec{x}\in w_1$，否则 $\vec{x}\in w_2$

即若($x_1<0$ 且 $x_2<2$) 或 ($x_1>0$ 且 $x_2>2$)，$\vec{x}\in w_1$ 否则 $\vec{x}\in w_2$

作业4

（一共M类）

一、1、用线性判别函数将属于ω_i类的模式与不属于ω_i类的模式分开

判别函数

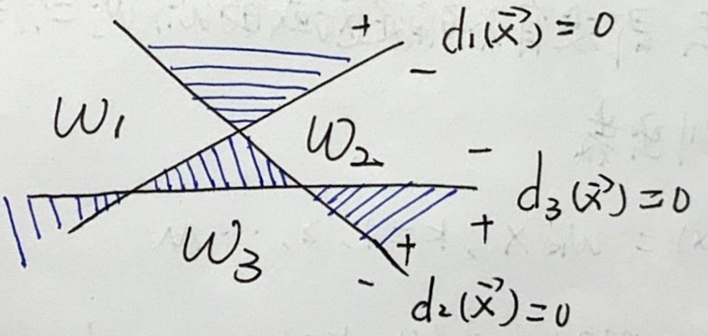
$$d_i(x) = w_i^T x = \begin{cases} >0 & x \in \omega_i \\ \leq 0 & x \notin \omega_i \end{cases} \quad i=1,2,\cdots M$$

$\omega_i / \overline{\omega_i}$ 二分法（1 vs 多），把M类多分类问题分成M个二分类问题

因此共M个判别函数，即M个判别面.

不确定区域：若对某一模式区域，$d_i(x)>0$条件超过1个，或全部$d_i(x)\leq 0$

则分类失败，为不可判别区域



阴影部分
为不可判别区

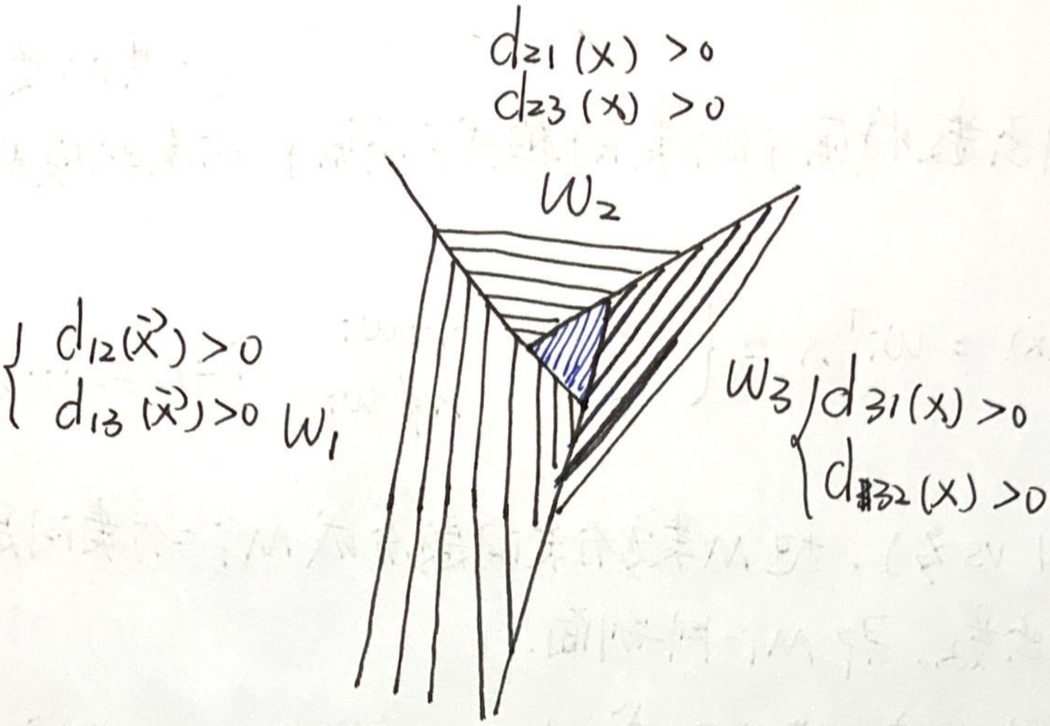
2、ω_i/ω_j两分法。此时一个判别界面只能分开两种类别

但不能把它与其他所有界面分开，判别函数为

$$d_{ij}(x) = w_{ij}^T x = \begin{cases} >0 & x \in \omega_i \\ \leq 0 & x \in \omega_j \end{cases} \quad \forall i \neq j$$

其中$d_{ij} = -d_{ji}$。共需$\frac{M(M-1)}{2}$个判别面

不确定区域：若所有$d_{ij}(x)$，找不到$\forall j \neq i$，$d_{ij}>0$的情况



蓝色阴影为不确定区域 不可判别

相比于上一种方法，该方法不可判别区域减少.

3. 最大判别准则法，即没有不确定区域的 w_i/w_j 二分法

首先构造M个判别函数

$$d_k(x)=w_k^T x,\ k=1,2,\cdots M$$

此时若存在 $d_i(x)=\max\{d_k(x),k=1,2,\cdots M\}$ 则 $x\in w_i$

或 $d_i(x)>d_j(x)\quad \forall j\neq i$ 则 $x\in w_i$

该方法需M个判别面. 无~~不确定~~不可判别区域

作业4

二、解：

$$S_w = \frac{1}{2}(S_1+S_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

二阶矩阵 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的逆为 $A^{-1} = \frac{1}{A} A^*$

解得：$S_w^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$d(\vec{x}) = (\vec{\mu_1} - \vec{\mu_2})' S_w^{-1} \left(\vec{x} - \frac{\vec{\mu_1}+\vec{\mu_2}}{2}\right)$$

$$= (0,-2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}(x_1-2, x_2-1)' = -2x_2+2 = 0$$

故决策面方程为 $-2x_2+2=0$

即 $x_2 = 1$

三. 期望风险（实际误差）：Risk 对于所有可能出现的样本的分类风险

经验风险（训练误差）：Empirical risk，线性可分情况下，可达到零

在给定训练集上对期望风险的估计 R_{emp}

在有限样本下，期望风险有可能大于训练风险

$$R \le R_{emp} + \varphi\left(\frac{h}{n}\right)$$ —— 置信区间

这是期望风险的上界，表明在训练误差相同的情况下，VC维越低两者差别越小。要使分类器具有好的推广能力，应使得置信区间达到最小，即在经验风险都达到0的情况下，寻求期望风险上界的最小化

SVM基本思想：折衷考虑经验风险和推广的置信界限，取得结构风险的最小化。即根据有限样本信息在模型复杂性和学习能力之间寻求最佳折中

核心概念：VC维和结构风险最小化

SVM的实际风险由经验风险值和置信范围值两部分组成。以训练误差作为优化问题的约束条件，以置信范围值最小化作为优化目标。即SVM是一种$\boxed{\text{基于结构风险最小化准则}}$的学习方法

四. 解：本题用MATLAB编程实现.

不同初始值	结果
$(0,0,0)^T$	$(-1,3,1)^T$
$(0,0,1)^T$	$(0,2,1)^T$
$(0,1,1)^T$	$(-1,2,0)^T$
$(1,1,1)^T$	$(0,3,2)^T$
⋮	⋮

```
x=[1,2,1;......];
y=[1;1;-1;-1];
N=10;
pho=1;
W=[1;1;1]
for k=1:N
    for i=1:4
        tmp=W'*x(i,:)';
        if((y(i)>0&&tmp<=0)||
           (y(i)<0&&tmp>=
            W=W+y(i)*x(i,:)
        end
    end
end
```

五、 (1) N个已知样本，分别属于c个类别 $w_1, w_2, \cdots w_c$. 每类有样本 $N_i\ (i=1,2,\cdots c)$ 个

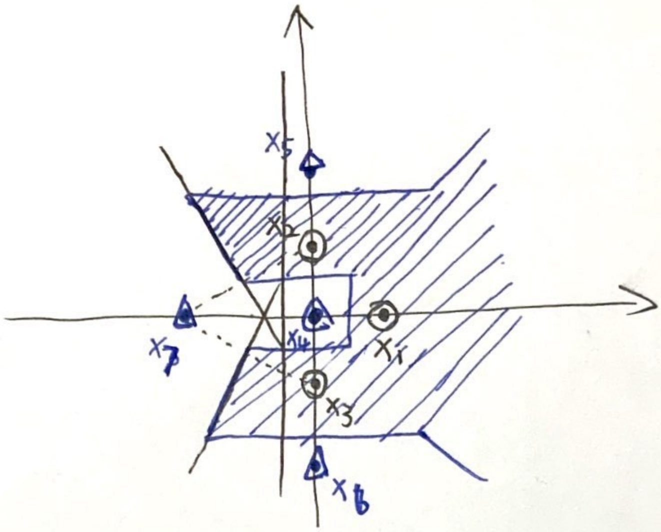
(2) 对于任意一个未知类别待识别模式$\vec{x}$，分别计算它与N个已知类别的样本$\vec{x}_j^{(i)}$的~~距理~~距离

(3) 取k个最近邻样本，这k个样本中哪一类最多就判x属于哪一类.

```
for k=1:N
    for i=1:4
        tmp=W'*x(i,:)';
        if((y(i)>0&&tmp<=0)||(y(i)<0&&tmp>=0))
            W=W+y(i)*x(i,:)';
        end
    end
end
```

六.

(1)

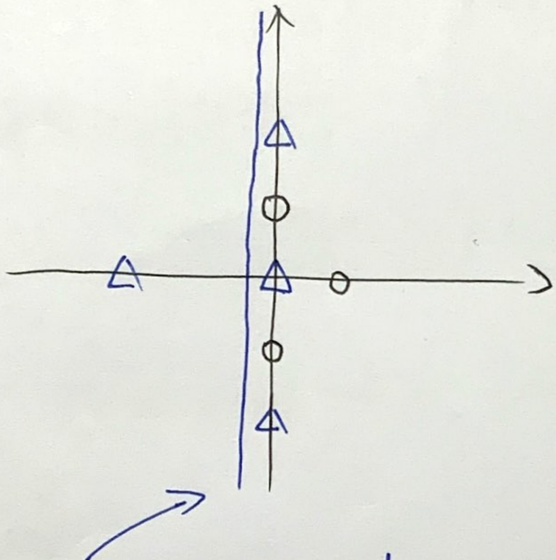


(仅画出 x_7 与第一类)

一共有 3×4=12 个分界面

蓝色阴影为 w_1 决策区域

(2)



按离均值距离的决策面 $-\frac{1}{12}$

三、(改) SVM思想概括为两点：

(1) 它基于结构风险最小化理论之上在特征空间中建构最优分割超平面，使得学习器得到全局优化，并且在整个样本空间的期望风险以某个概率满足一定上界

(2) 它针对线性可分情况进行分析，对于线性不可分的情况，通过使用非线性映射算法将低维输入空间线性不可分的样本转化为高维特征空间使其线性可分从而使得高维特征空间采用线性算法对样本的非线性特征进行线性分析成为可能

作业5

史文纬
2015910001[illegible]

1、(1) 协方差矩阵(Covariance matrix)

$$C=(C_{ij})_{n\times n}=\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}$$

其中 $C_{ij}=Cov(x_i,x_j)$

所以其对角元素是各分量的方差，非对角元素是各分量之间协方差

(2) 求协方差矩阵特征值

$$\begin{pmatrix} \lambda-1 & -\frac{1}{2} \\ -\frac{1}{2} & \lambda-1 \end{pmatrix}=0 \Rightarrow (\lambda-1)^2=\frac{1}{4} \Rightarrow \lambda=\frac{1}{2}\text{或}\frac{3}{2}$$

$\lambda=\frac{1}{2}$ 时，对应特征向量 $\begin{pmatrix}1\\-1\end{pmatrix}$ $\lambda=\frac{3}{2}$ 时，对应特征向量 $\begin{pmatrix}1\\1\end{pmatrix}$

即为主分量

(3) 对一组数据进行按一组正交基分解，在只取相同数量分量的条件下，以均方误差计算截尾误差最小

(4) 在经主分量分解后，协方差矩阵成为对角矩阵，因而各主分量之间相关性消除

2. 解：

(1) ① 求样本总体均值向量 $\mu=(-6,-6)^T+(6,6)^T=(0,0)^T$

② 求协方差矩阵 $R=[(-6,-6)^T(-6,-6)+(6,6)^T(6,6)]/2=\begin{bmatrix}36&36\\36&36\end{bmatrix}$

③ 求特征根，令 $\begin{vmatrix}36-\lambda & 36\\ 36 & 36-\lambda\end{vmatrix}=0 \Rightarrow \lambda_1=72 \quad \lambda_2=0$

由 $R\varphi_i=\lambda_i\varphi_i$，得特征向量 $\varphi_1=\begin{bmatrix}1\\1\end{bmatrix}/\sqrt{2}$，$\varphi_2=\begin{bmatrix}1\\-1\end{bmatrix}/\sqrt{2}$

PCA为 $[\varphi_1,\varphi_2]^T\begin{bmatrix}-6\\-6\end{bmatrix}=\begin{bmatrix}-6\sqrt{2}\\0\end{bmatrix}$，$[\varphi_1,\varphi_2]^T\begin{bmatrix}6\\6\end{bmatrix}=\begin{bmatrix}6\sqrt{2}\\0\end{bmatrix}$

(2) 一维数据压缩，即向最大特征根对应的特征向量做投影，

得 $-6\sqrt{2}$，$6\sqrt{2}$.

3.

(1) 优点：

① 变换在均方误差最小的意义下使新样本集 $\{X^*\}$ 逼近原样本集 $\{X\}$ 的分布，既压缩了维数又保留了类别鉴别信息；

② 变换后的新模式向量各分量相对总体均值的方差等于原样本集总体自相关矩阵的大特征值，表明变换突出了模式类之间的差异性；

③ 变换后样本各分量互不相关，便于进一步进行特征的选择

(2) 缺点：

① 对两类问题容易得到较满意的结果，类别越多，效果越差；

② 需要通过足够多的样本估计样本集的协方差矩阵或其它类型的散布矩阵，当样本数不足时，矩阵的估计会变得十分粗略，变换的优越性也就不能充分地显示出来；

③ 不能反映原始数据的结构信息。

作业5

史文纬
2015910001 8

4. 解：

$$\vec{m} = \frac{1}{5}\sum_{i=1}^{5}\vec{x}_i^{(1)} + \frac{1}{5}\sum_{i=1}^{5}\vec{x}_i^{(2)} = 0$$

因为 $\hat{P}(w_1) = \hat{P}(w_2) = \frac{5}{10} = \frac{1}{2}$

所以 $R = E[\vec{x}\vec{x}'] = \sum_{i=1}^{2}\hat{P}(w_i)E[\vec{x}^{(i)}\vec{x}^{(i)'}]$

$$= \frac{1}{2}\left[\frac{1}{5}\sum_{i=1}^{5}\vec{x}_i^{(1)}\vec{x}_i^{(1)'}\right] + \frac{1}{2}\left[\frac{1}{5}\sum_{i=1}^{5}\vec{x}_i^{(2)}\vec{x}_i^{(2)'}\right]$$

$$= \begin{pmatrix} 25.4 & 25 \\ 25 & 25.4 \end{pmatrix}$$

求R的特征值、特征矢量

$$|R - \lambda I| = (25.4 - \lambda)^2 - 25^2 = 0 \Rightarrow \lambda_1 = 50.4 \quad \lambda_2 = 0.4$$

由 $R\vec{t}_j = \lambda_j\vec{t}_j$，得 $\vec{t}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$，$\vec{t}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$

选 λ_1 对应的 $\vec{t}_1$ 作为变换矩阵 $T = \vec{t}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$

由 $y = T'\vec{x}$ 得变换后的一维模式特征为

$$y_i^{(1)} = T'\vec{x}_i^{(1)} \qquad i = 1, 2 \ldots$$

$$w_1 = \left\{-\frac{10}{\sqrt{2}}, -\frac{9}{\sqrt{2}}, -\frac{9}{\sqrt{2}}, -\frac{11}{\sqrt{2}}, -\frac{11}{\sqrt{2}}\right\}$$

$$w_2 = \left\{\frac{10}{\sqrt{2}}, \frac{11}{\sqrt{2}}, \frac{11}{\sqrt{2}}, \frac{9}{\sqrt{2}}, \frac{9}{\sqrt{2}}\right\}$$