作业1

史文纬

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

(1) $\Longleftarrow$

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

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

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

而 X Y 的概率密度为

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

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

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

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

$$= (BA)U + (B\mu + d)$$

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

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

$$x = C^{-1}y \quad ②$$

再乘以变换②的雅可比行列式

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

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

得到Y的概率密度．$n(x|\mu,\Sigma)$ 的指数中的二次型是

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

将②代入④得

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

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

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

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

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

因此Y的密度为

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

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

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

得证。

模式识别作业 2

Shi 20159100018

December 2022

本实验所有支撑材料均上传个人 GitHub 仓库, 超链接见脚注[1]

1 基本内容

用身高体重数据、或 IRIS 数据作为本次实验使用的样本集，利用 C 均值、FCM 和层次聚类方法对样本集进行聚类分析，对结果进行分析，从而加深对所学内容的理解和感性认识。

2 作业要求

1. 同时采用身高和体重数据作为特征，设类别数为 2，利用 C 均值聚类方法对数据进行聚类，并将聚类结果表示在二维平面上。尝试不同初始值对此数据集是否会造成不同的结果。
2. 对 1 中的数据利用 C 均值聚类方法分别进行两类、三类、四类、五类聚类，画出聚类指标与类别数之间的关系曲线，探讨是否可以确定出合理的类别数目。
3. 对 1 中的数据利用层次聚类方法进行聚类，分析聚类结果，体会层次聚类方法。
4. 对 1 中的数据进行 FCM 聚类，讨论 m 对聚类的影响，以及不同聚类个数的合理性。
5. 对于遗传学数据，分析是否采用父母数据对结果有什么影响。

[1] code

3 算法思想及程序实现

3.1 算法思想

3.1.1 C 均值聚类法

该方法取定 C 个类别和选取 C 个初始聚类中心，按最小距离原则将各模式分配到 C 类中的某一类，之后不断地计算类心和调整各模式的类别，最终使各模式到其判属类别中心的距离平方之和最小

3.1.2 层次聚类法

首先将 N 个模式视作各自成为一类，然后计算类与类之间的距离，选择距离最小的一对合并成一个新类，计算在新的类别分划下各类之间的距离，再将距离最近的两类合并，直至所有模式聚成两类为止。

3.1.3 FCM

根据模糊集合的概念，一个元素隶属于模糊集合不是硬性的，在聚类的问题中，可以把聚类生成的类看成模糊集合，因此，每个样本点隶属于某个类的隶属度就是 $[0,1]$ 区间里面的值。

假定条件及约定，设待分类的模式为 $\{\vec{x}_1, \vec{x}_2, \cdot\cdot\cdot \vec{x}_N,\}$, 类的数目 c 事先约定，那么对应的就有 c 个类中心为 V，每个样本 j 属于某一类 i 的隶属度为 u_{ij} 为求约束条件下目标函数的极值，得必要条件如下：

$$u_{ij} = \frac{1}{\sum\limits_{r=1}^{c} (\frac{d_{ji}}{d_{jr}})^{\frac{2}{m-1}}} \tag{1}$$

$$v_i = \frac{\sum\limits_{j=1}^{n} (u_{ij})^m x_j}{\sum\limits_{j=1}^{n} (u_{ij})^m} \tag{2}$$

用上式循环迭代得到满足要求的聚类中心和隶属度矩阵。

3.2 算法步骤

3.2.1 C 均值聚类法

(1) 任选 C 个模式特征矢量作为初始聚类中心:

$$\vec{z}_1^{(0)}, \vec{z}_2^{(0)}, \cdot\cdot\cdot, \vec{z}_c^{(0)}, \quad k=0$$

(2) 将待分类的模式特征矢量集 $\vec{x}_i$ 中的模式逐个按最小距离原则分划给 C 类中的某一类，如果 $d_{il}^{(k)} = min[d_{ij}^{(k)}] \quad (i=1,2,...)$, 则 $\vec{x}_i \in \omega_l^{(k+1)}$。式中 $d_{ij}^{(k)}$ 表示 $\vec{x}_i$ 和 $\omega_j^{(k)}$ 的中心 $\vec{z}_j^{(k)}$ 的距离，上角标表示迭代次数。于是产生新聚类 $\omega_j^{(k+1)}(j=1,2...c)$

(3) 计算重新分类后的各类心

$$\vec{z}_j^{(k+1)} = \frac{1}{n_j^{k+1}\sum\vec{x}_i} \quad (j=1,2...c)$$

式中 n_j^{k+1} 为类 $\omega_j^{(k+1)}$ 中所含模式的个数。

(4) 如果 $\vec{z}_j^{(k+1)} = \vec{z}_j^{(k)}, j=1,2,...,c$, 则结束，否则 $k=k+1$, 转至 (2)

3.2.2 层次聚类法

(1) 初始分类。令 k=0, 每个模式自成一类, 即 $G_i^{(0)} = \{\vec{x}_i\}(i=1,2,\cdot\cdot\cdot,N)$

(2) 计算各类间的距离 D_{ij}, 由此生成一个兑成距离矩阵 $D^{(k)} = (D_{ij})_{m\times m}$,$m$ 为类的个数 (初始时 $m=N$)

(3) 找出前一步求得的矩阵 $D^{(k)}$ 中最小的元素，设它是 $G_i^{(k)}$ 与 $G_j^{(k)}$ 间的距离，将 $G_i^{(k)}$ 与 $G_j^{(k)}$ 两类合并称一类，于是产生新的聚类 $G_1^{(k+1)}$, $G_2^{(k+1)}\cdot\cdot\cdot$，令 $k=k+1, m=m-1$

(4) 检查类的个数。如果类数 m 大于 2，转至 (2); 否则停止。

3.2.3 FCM

(1) 设定聚类数目 c 和参数 m（模糊因子）

(2) 给出初始隶属度矩阵 $U^{(0)}(U^{(0)}$ 各行元素之和应为 1，在实验程序给 $U^{(0)}$ 每一列随机赋予 $c-1$ 个 0, 1 个 1)

(3) 利用式 (2) 计算新的聚类中心 V_i

(4) 利用式 (1) 计算新的隶属度矩阵 (若分母存在中 $d_{ji} = 0$, 则 $u_{ij} = 1$, 且对 $j \neq i, u_{ij} = 0$)

(5) 用一个矩阵范数比较两次迭代之间隶属度矩阵，如果

$$||U^{k+1} - U^k|| \leq e$$

则停止迭代，否则转 (3)。

3.3 数据说明

根据实验材料所给，word 文档的表中共 52 位同学的身高体重数据。即本次聚类分析实验样本数量为 52。将实验样本绘制成二维图如下：

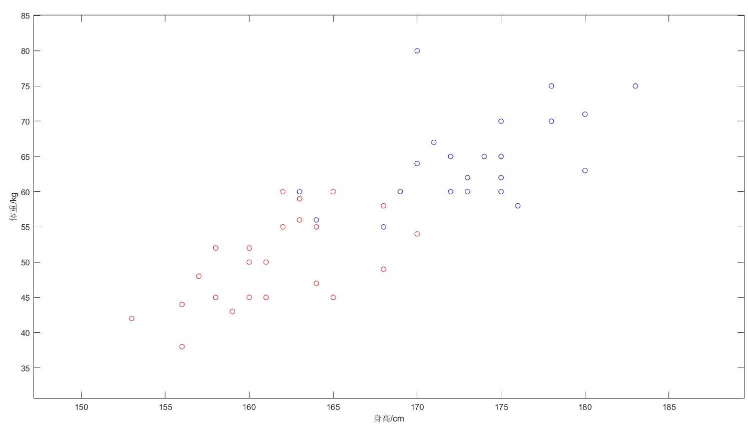


图 1: 实验数据，红色为女生，蓝色为男生，各 26 个

3.4 程序实现

程序实现主要依赖 MATLAB 中所提供的函数，其中 C-means 聚类主要使用`kmeans`函数; 层次聚类法主要使用`pdist`、`linkage`、`dendrogram`、`cluster`; FCM 主要使用`fcm`。具体的完整代码与注释见附录，.m 文件见 GitHub 仓库： code

4 实验结果分析

4.1 要求 1

利用 C-means 对数据进行聚类，选择类别数为 2。第一次选择第一个男生和第一个女生的身高体重数据，初始聚类中心为:(163,60)(153,42)。得到聚类结果如下。聚类中心为 (160.773，48.5455)(172，63.4),
$J_e = 2.1565e + 03$。

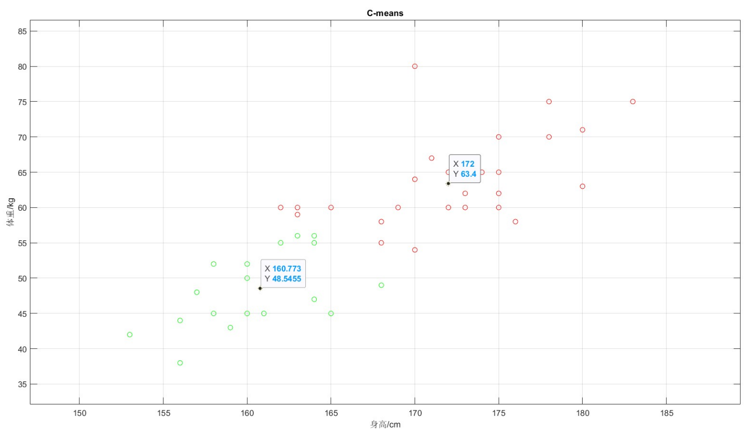


图 2: C=2 的第一次聚类结果

第二次选择最后一个男生和最后一个女生,初始聚类中心为:(183,75)(170,54)。得到聚类结果如下。聚类中心为 (162.29,51.2258)(174.571,65.8095),
$J_e = 1.9466e + 03$。

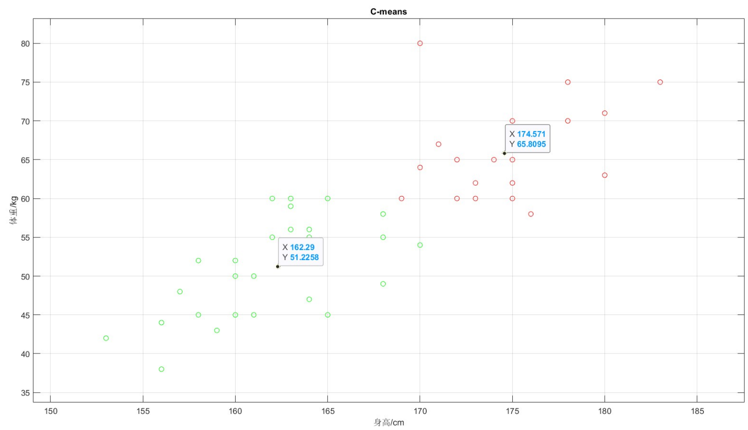


图 3: C=2 的第二次聚类结果

第三次选择最后一个男生和第一个女生,初始聚类中心为:(183,75)(153,42). 得到聚类结果如下。聚类中心为 (162.29,51.2258)(174.571,65.8095), $J_e = 1.9466e + 03$。

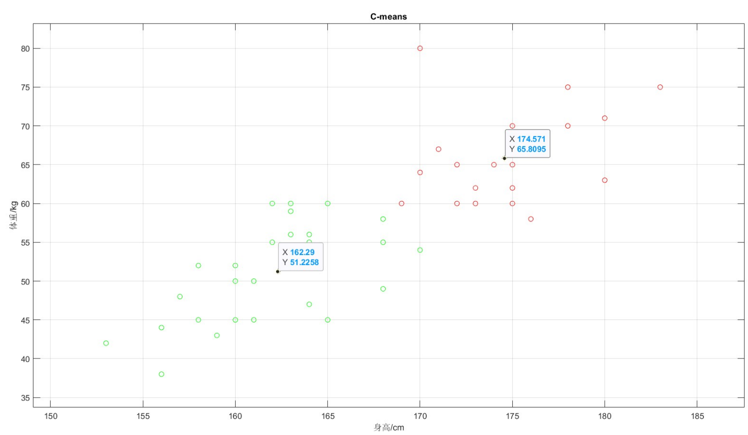


图 4: C=2 的第三次聚类结果

通过选择不同的初始聚类中心，可以得到不同的结果，但不同的聚类中心也有可能得到相同的聚类结果。

4.2 要求 2

选择类别数为 3, 初始聚类中心为 (183,75)(170,54)(153,42)。得到聚类结果如下。聚类中心为: (160.059,46.4706)(167.727,58.2727)(175.846,69.0769). $J_e = 903.2275$

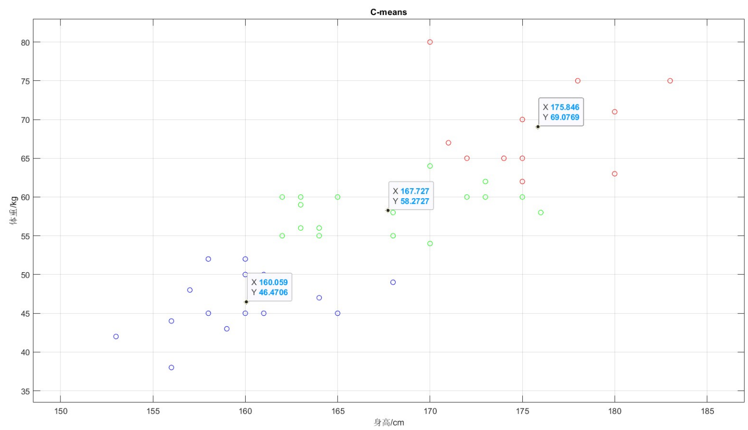


图 5: C=3 的聚类结果

选择类别数为 4, 初始聚类中心为 (183,75)(170,54)(163,60)(153,42)。得到聚类结果如下。聚类中心为: (160.059,46.4706)(167.727,58.2727)(175.846,69.0769).$J_e = 562.9040$

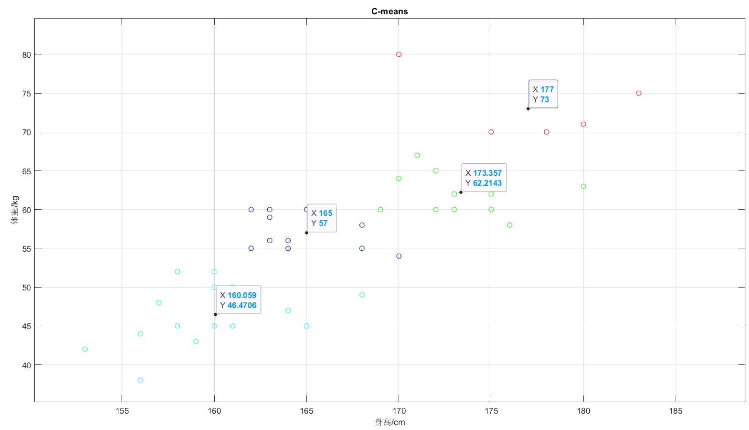


图 6: C=4 的聚类结果

选择类别数为 5，初始聚类中心为 (183,75)(170,54)(163,60)(153,42)(172,60)。得到聚类结果如下。聚类中心为：(160.059,46.4706)(167.727,58.2727)(175.846,69.0769). $J_e = 370.8981$

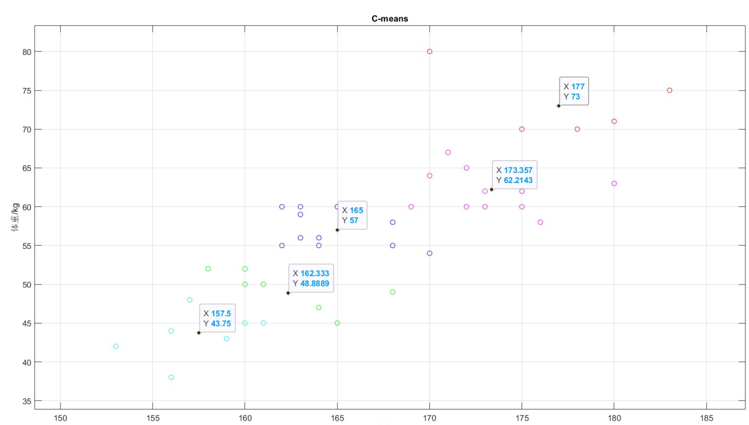


图 7: C=5 的聚类结果

画出不同聚类类别数 C 与 J_e 的函数曲线如下，可以看出，拐点离 2 较近，所以此样本集聚为二类最佳。

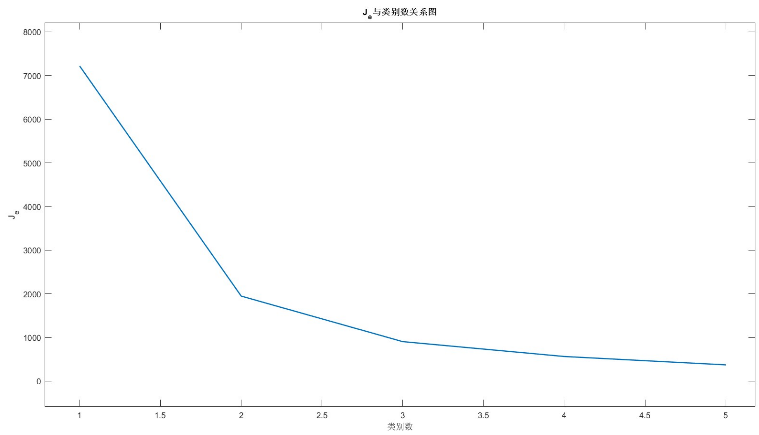


图 8: J_e 随类别数的变化

4.3 要求 3

层次聚类法所得系统树图如下。(为什么聚类结果很差？)

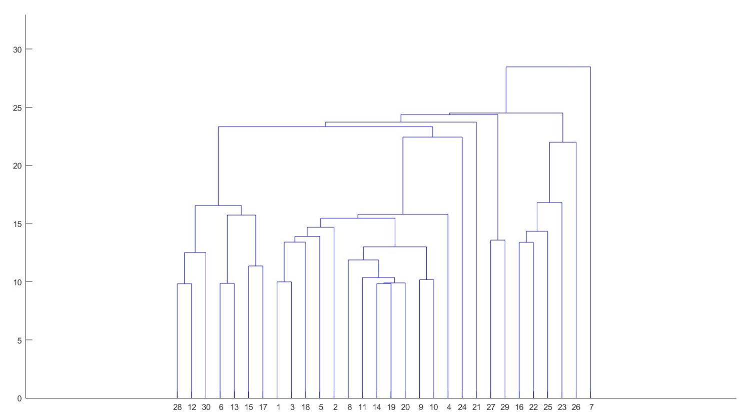


图 9: 层次聚类聚类树

4.4 要求 4

首先选择 m=1.5,c=2 进行聚类，得到结果如下:

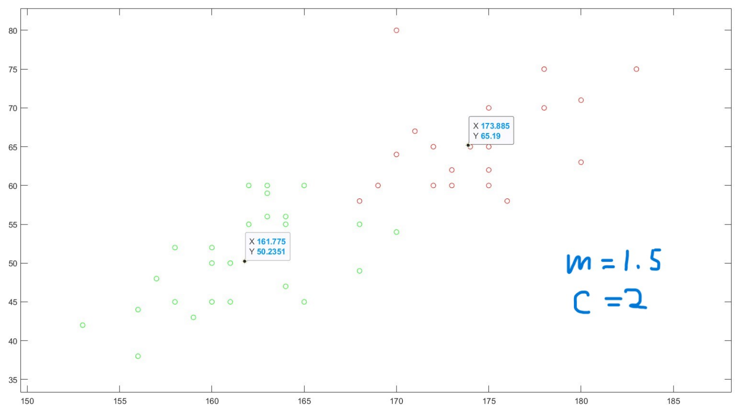


图 10: m=1.5,c=2 聚类结果

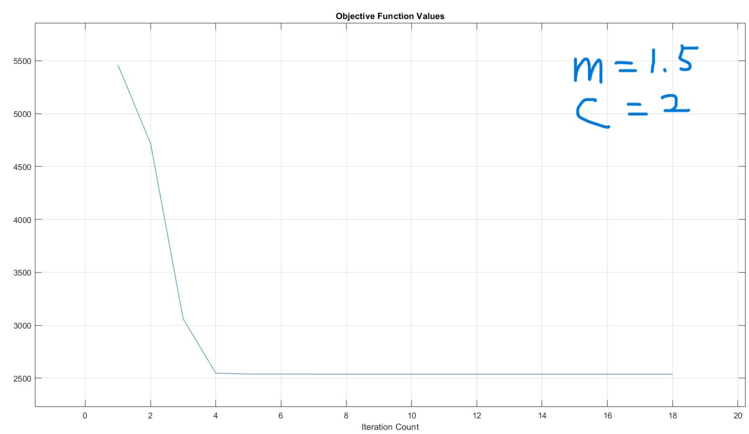


图 11: m=1.5,c=2 目标函数 obj_fcn 变化

选择 m=2,c=2 进行聚类，得到结果如下:

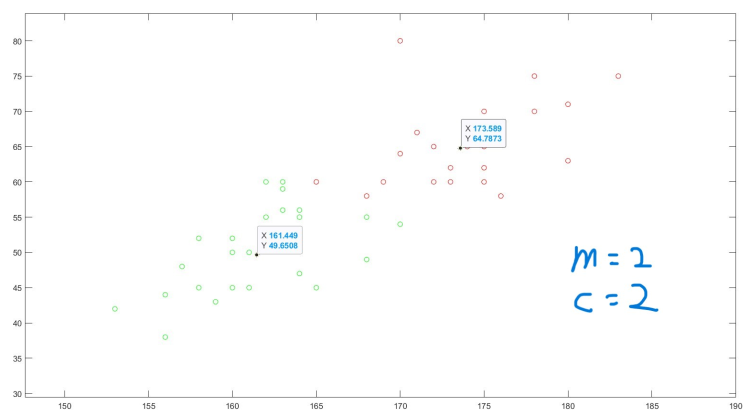


图 12: m=2,c=2 聚类结果

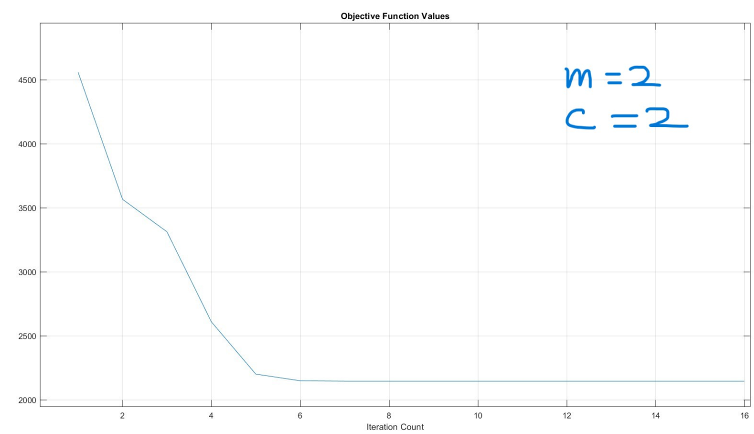


图 13: m=2,c=2 目标函数 obj_fcn 变化

选择 m=3,c=2 进行聚类，得到结果如下:

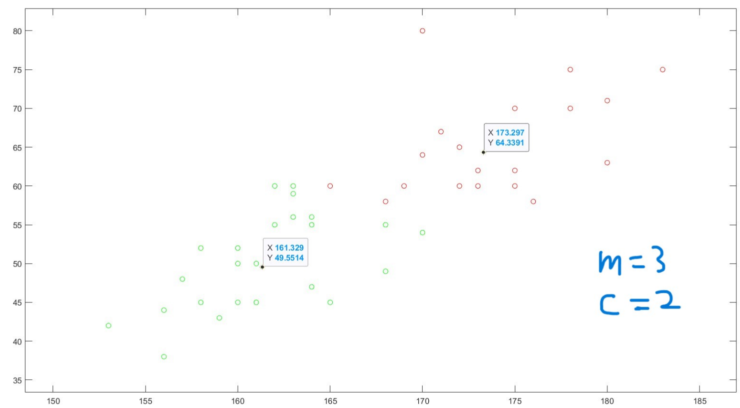


图 14: m=3,c=2 聚类结果

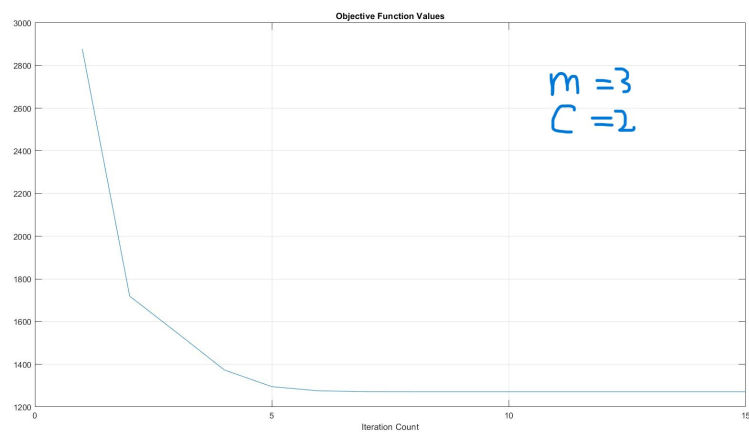


图 15: m=3,c=2 目标函数 obj_fcn 变化

选择 m=1.5,c=3 进行聚类，得到结果如下:

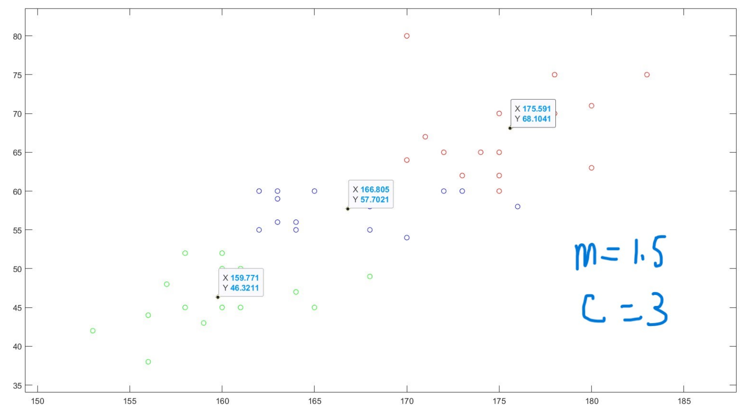


图 16: m=1.5,c=3 聚类结果

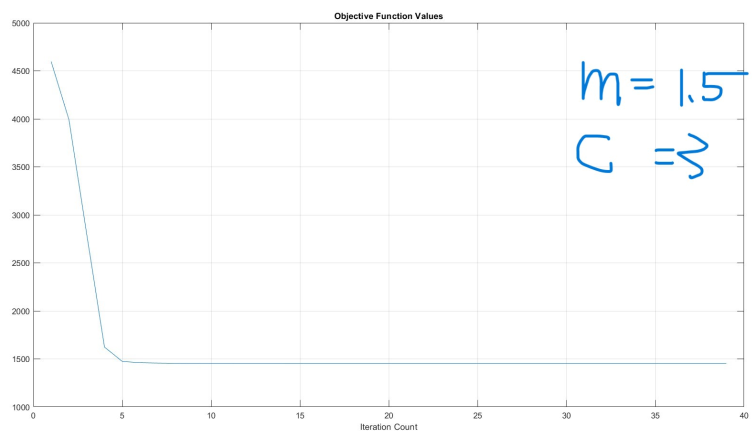


图 17: m=1.5,c=3 目标函数 obj_fcn 变化

当 m 分别取 2，1.5，2.5，C 为 2 时, 可以看出分类结果大致相同，但目标函数 obj_fcn 曲线在 m 的不同取值下有明显变化。

根据查阅的相关资料,在实际大多数应用中,m 最佳取值范围是 $(1.5, 2.5)$

4.5 要求 5

通过观察遗传学数据注意到，父母与子女的身高体重是一种拟合较好的线形关系，但仍然存在例外现象：矮个的人的儿子比其父要高，身材较高的父母所生子女的身高将回降到人的平均身高。当父母身高走向极端（或者非常高，或者非常矮）的人的子女，子女的身高不会象父母身高那样极端化，其身高要比父母们的身高更接近平均身高。高尔顿选用“回归”一词，把这一现象叫做“向平均数方向的回归”。但也存在外界环境对子女身高体重的影响，这部分对于普遍数据偏差不大。所以对聚类分析来说，父母的数据对分类结果的趋势影响不大。

5 心得体会

通过本次上机实验，我加深了对 C-means 聚类算法、层次聚类法和 FCM 聚类法的理解。并编程实现了三种算法。同时，对 MATLAB 编程更加熟悉。

在本次上机过程中，遇到了很多问题，通过查询资料和与同学讨论，逐一解决或避开问题，提高了解决问题的能力。同时更加深刻的认识到，官方文档才是解决问题的‘真理’，CSDN 与百度内容质量参差不齐，很容易在其上浪费大量时间。但是官方文档大部分为英文，很少有翻译为中文的文档，所以，学好英语才能更好地学习技术。

附录

C-means

```
clear
clc
%载入数据，绘图。
load('x.mat');
load('male.mat');
load('female.mat');
figure;
plot(male(:,1),male(:,2),'color','b','LineStyle','none','Marker','o');
hold on;
plot(female(:,1),female(:,2),'color','r','LineStyle','none','Marker','o');
hold on;
xlabel('身高/cm');
ylabel('体重/kg');
k = 1;%类别数选择

%初始质心坐标选择
%C=1
matric=[170,54];

%C = 2;
%matric=[163,60;153,42];%01
%matric=[183,75;170,54];%02
%matric=[183,75;153,42];%03

%C = 3;
%matric=[183,75;170,54;153,42];

%C = 4;
%matric=[183,75;170,54;163,60;153,42];

%C = 5;
%matric=[183,75;170,54;163,60;153,42;172,60];

%真正的代码就这一句，如果不能用封装函数，需要重写。
[idx,cen,sumD,D] = kmeans(x,k,'Start',matric);
```

```
%silhouette(x,idx); %轮廓值/轮廓系数---衡量聚类分析好坏的一个指标

%把图像画出来
color=['r','g','b','c','m','y'];


figure,
for i=1:k
    plot(x(idx==i,1),x(idx==i,2),'color',color(i),'LineStyle','none','Marker','o')
    hold on
end
xlabel('身高/cm');
ylabel('体重/kg');
title('C-means');

% %计算J_e 1
% J_e = 0;
% for j=1:k
%     J_tmp = [x(idx == i,1)-cen(i,1) x(idx == i,2)-cen(i,2)];
%     J_e = J_e + sum(sum(J_tmp.^2));
% end

%计算J_e 2
J_e = norm(sumD);


%显示质心坐标
plot(cen(:,1),cen(:,2),'Color','k','LineStyle','none','Marker','*')
hold off
grid on


%%绘制J_e与类别数关系图
Je = [7.2171e+03,1.9466e+03,903.2275,562.9040,370.8981];
c=[1,2,3,4,5];
figure;
plot(c,Je,'LineWidth',1.5);
xlabel('类别数')
```

```
ylabel('J_e')
title('J_e与类别数关系图')
```

层次聚类

```
%参考代码:
%https://ww2.mathworks.cn/help/stats/cluster.html
% 载入数据
clear;
clc;
load('x.mat');
k = 2;
% 计算前边点与后边点距离
disVector = pdist(x); %
    pdist之后的Y是一个行向量，15个元素分别代表X点之间的距离

% 转换成方阵
disMatrix = squareform(disVector);

% 确定层次聚类树
treeCluster = linkage(disMatrix);

% 可视化聚类树
dendrogram(treeCluster);

% 聚类下标
idx = cluster(treeCluster,'maxclust',k);

% 画图
color=['r','g','b','c','m','y'];
figure,
for i=1:k
    plot(x(idx==i,1),x(idx==i,2),'color',color(i),'LineStyle','none','Marker','o')
    hold on
end
title('Hierarchical');
```

FCM

```
clear;
load('x.mat');
k=3;%分类类别数
m=1.5;%m值的选择
[center,U,obj_fcn] = fcm(x,k,m);

maxU = max(U);

%画图
color=['r','g','b','c','m','y'];
figure;
for i=1:k
    plot(x(U(i,:) == maxU,1),x(U(i,:) ==
        maxU,2),'marker','o','LineStyle','none','color',color(i));
    hold on;
end

plot(center(:,1),center(:,2),'Color','k','LineStyle','none','Marker','*')

%随迭代次数变化
figure
plot(obj_fcn)
title('Objective Function Values')
xlabel('Iteration Count')

hold off;
grid on;
```

作业3　　　　　　　　　　　　　　史文纬 20159100018

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

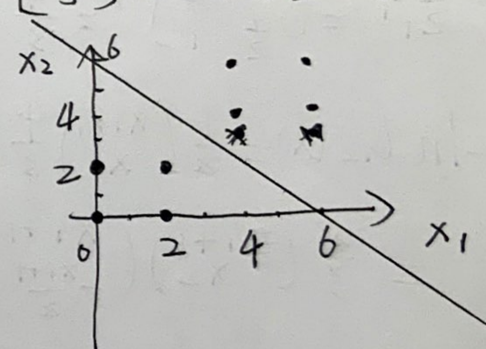
$$\Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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 $x_1 + x_2 = 6$



2015910007?

史文纬

2. (a) 由已知

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

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

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

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

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

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

$$= (x_1+1)^2-(x_1-1)^2 = 4x_1 \qquad = \ln 1 = 0$$

故~~两~~若 $x_1<0$ 则 $x\in w_1$，若 $x_1>0$ 则 $x\in w_2$

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

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

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

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

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

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w_i类的模式与不属于w_i类的模式分开

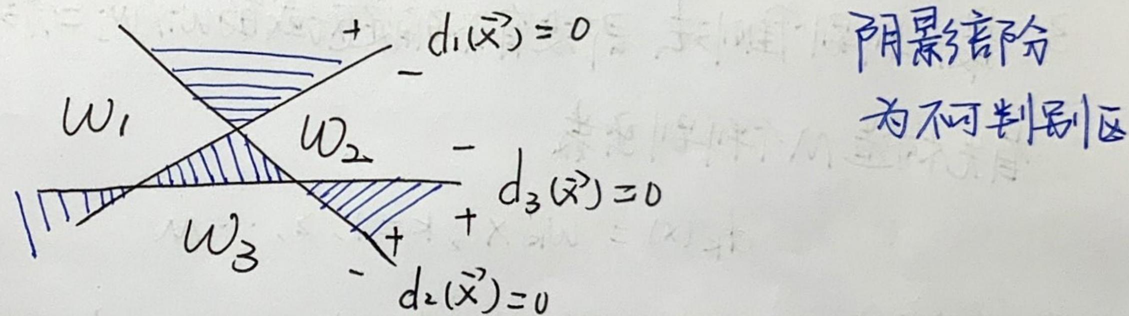
判别函数 $d_i(x) = w_i^T x = \begin{cases} >0 & x \in w_i \\ \leq 0 & x \notin w_i \end{cases} \quad i=1,2,\cdots M$

$w_i/\overline{w_i}$ 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



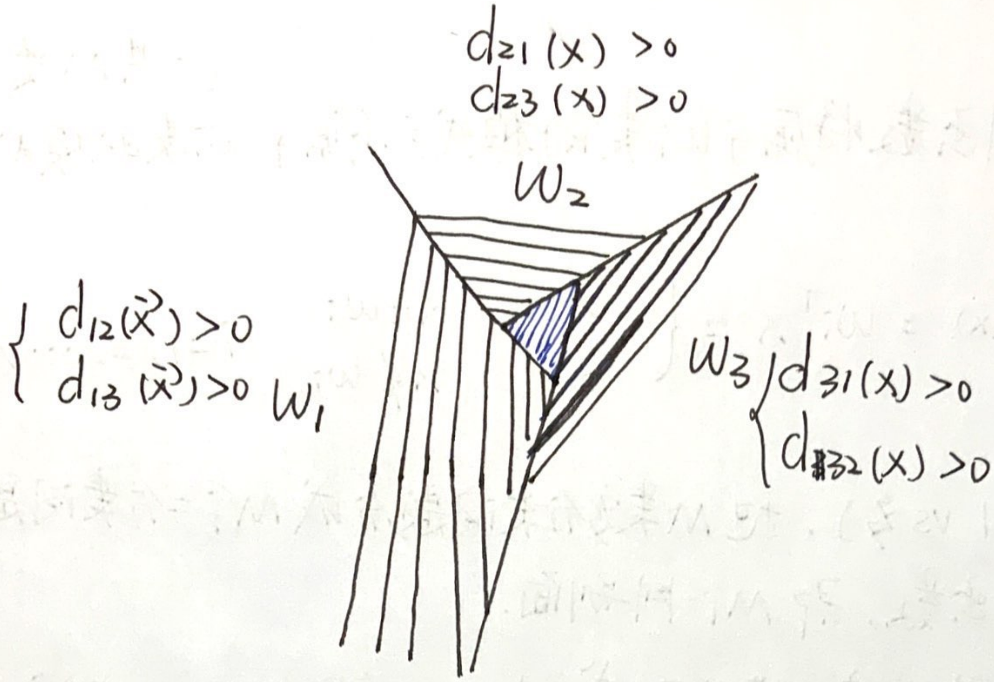
2. w_i/w_j 两分法。此时一个判别界面只能分开两种类别

但不能把它与其他所有界面分开，判别函数为

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

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　　Remp

在有限样本下，期望风险有可能大于训练风险

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

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

SUM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

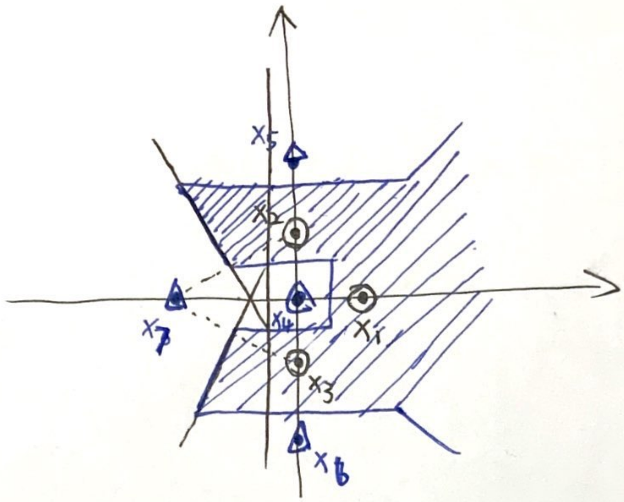
(2) 对于任意一个未知类别待识别模式$\vec{x}$，分别计算它与N个已知类别的样本$\vec{x}^{(i)}$的~~距理~~距离

(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)

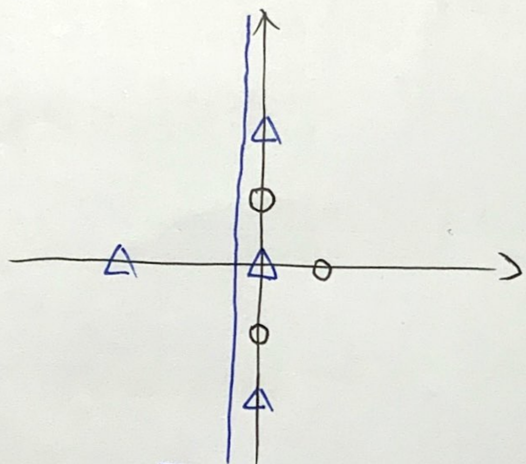


(仅画出x_7与第一类)

一共有$3\times4=12$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

史文纬
20159100018

作业5

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 $c_{ij}=\mathrm{Cov}(x_i, x_j)$

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

2015910018

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

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

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

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~~中~~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

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

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

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