石家莊鐵道大學
SHIJIAZHUANG TIEDAO UNIVERSITY

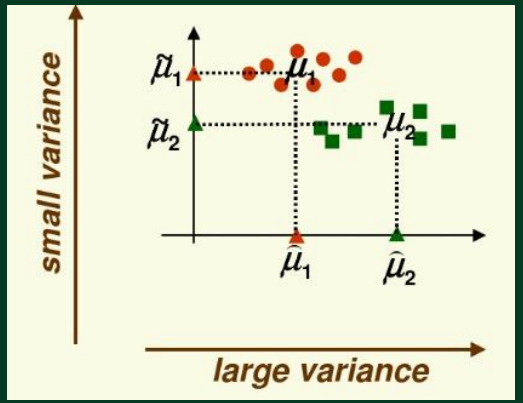
在线开放课程

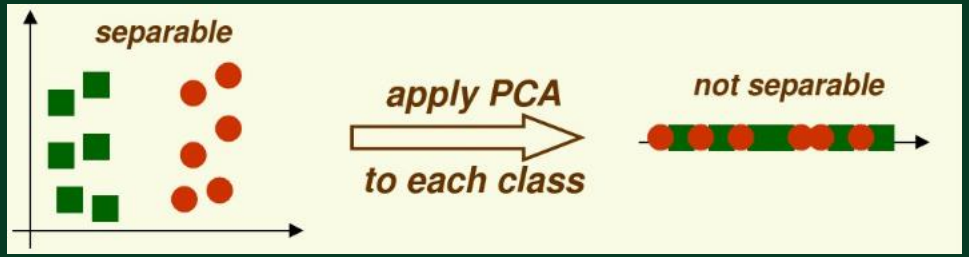
MATLAB在科学研究中的应用

Fisher线性判别

主讲 ： 卞建鹏

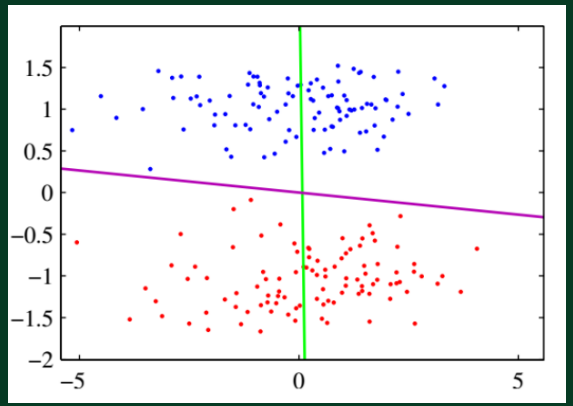
1、PCA存在的问题

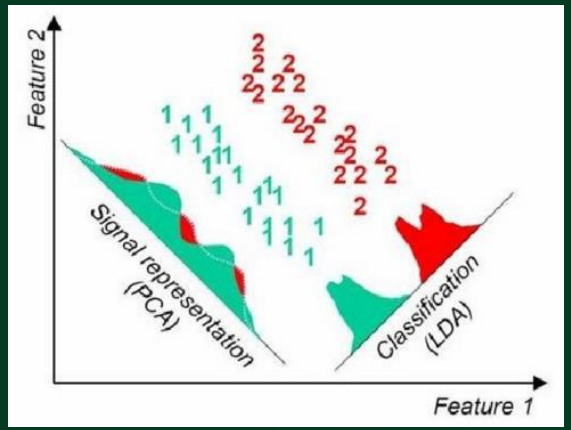

small variance
large variance


separable
apply PCA
to each class
not separable

2、Fisher线性判别法

基本思想：通过将多维数据投影到某一方向上，使得投影后类与类之间尽可能的分开，然后再选择合适的判别准则。





2、Fisher线性判别法

最大化

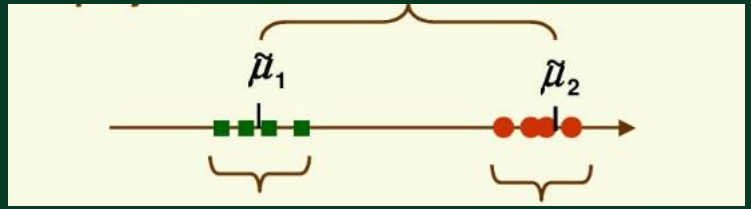
$$J(v)=\frac{(\tilde{\mu}_1-\tilde{\mu}_2)^2}{\tilde{s}_1^2+\tilde{s}_2^2}$$

类内离散度矩阵

$$S_1=\sum_{x_i\in Class\ 1}(x_i-\mu_1)(x_i-\mu_1)^t$$

$$S_2=\sum_{x_i\in Class\ 2}(x_i-\mu_2)(x_i-\mu_2)^t$$

类间散度



类内散度　　类内散度

2、Fisher线性判别法

最大化

$$J(v)=\frac{(\tilde{\mu}_1-\tilde{\mu}_2)^2}{\tilde{s}_1^2+\tilde{s}_2^2}=\frac{v^t S_B v}{v^t S_W v}$$

$$S_1=\sum_{x_i\in Class\ 1}(x_i-\mu_1)(x_i-\mu_1)^t$$

$$S_2=\sum_{x_i\in Class\ 2}(x_i-\mu_2)(x_i-\mu_2)^t$$

$$S_B=(\mu_1-\mu_2)(\mu_1-\mu_2)^t$$

$$S_W=S_1+S_2$$

求导，可得最优投影向量

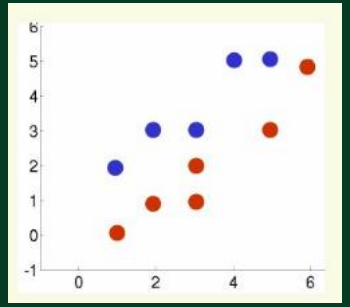
$$v=S_W^{-1}(\mu_1-\mu_2)$$

2、Fisher线性判别法

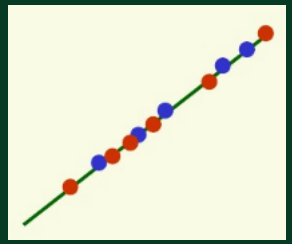
Data

- Class 1 has 5 samples c_1=[(1,2),(2,3),(3,3),(4,5),(5,5)]
- Class 2 has 6 samples c_2=[(1,0),(2,1),(3,1),(3,2),(5,3),(6,5)]

$$c_1 = \begin{bmatrix} 1 & 2 \\ \vdots & \vdots \\ 5 & 5 \end{bmatrix} \qquad c_2 = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 6 & 5 \end{bmatrix}$$



如果应用PCA可得



2、Fisher线性判别法

计算过程

（1）均值 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$c_1 = \begin{bmatrix} 1 & 2 \\ \vdots & \vdots \\ 5 & 5 \end{bmatrix} \qquad c_2 = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 6 & 5 \end{bmatrix}$$

$$\mu_1 = mean(c_1) = [3 \quad 3.6] \qquad \mu_2 = mean(c_2) = [3.3 \quad 2]$$

（2）类内离散度矩阵

$$S_1 = \sum_{x_i \in Class\ 1} (x_i - \mu_1)(x_i - \mu_1)^t$$
$$S_2 = \sum_{x_i \in Class\ 2} (x_i - \mu_2)(x_i - \mu_2)^t$$

$$S_1 = 4 * cov(c_1) = \begin{bmatrix} 10 & 8.0 \\ 8.0 & 7.2 \end{bmatrix} \qquad S_2 = 5 * cov(c_2) = \begin{bmatrix} 17.3 & 16 \\ 16 & 16 \end{bmatrix}$$

（3）总类内离散度矩阵 $S_W = S_1 + S_2 = \begin{bmatrix} 27.3 & 24 \\ 24 & 23.2 \end{bmatrix}$

2、Fisher线性判别法

（4）计算总类内离散度矩阵的逆矩阵

$$S_w^{-1} = inv(S_w) = \begin{bmatrix} 0.39 & -0.41 \\ -0.41 & 0.47 \end{bmatrix}$$

（5）可得最优映射向量、一维向量

$$v = S_w^{-1}(\mu_1 - \mu_2) = \begin{bmatrix} -0.79 \\ 0.89 \end{bmatrix}$$

$$Y_1 = v^t c_1^t = [-0.79 \quad 0.89] \begin{bmatrix} 1 \cdots 5 \\ 2 \cdots 5 \end{bmatrix} = [0.99 \quad 1.09 \quad 0.30 \quad 1.29 \quad 0.50]$$

$$Y_2 = v^t c_2^t = [-0.79 \quad 0.89] \begin{bmatrix} 1 \cdots 6 \\ 0 \cdots 5 \end{bmatrix} = [-0.79 \quad -0.69 \quad -1.48 \quad -0.59 \quad -1.28 \quad -0.29]$$

3、Fisher线性判别应用


```
w12=xlsread('E:\data.xls','C2:F16');
w1=w12(1:5,:);   w2=w12(6:12,:);      % training data
 sample=w12(13:15,:);                 % test data
r1=size(w1,1);
r2=size(w2,1);
r3=size(sample,1);
m1=mean(w1);
m2=mean(w2);
```

	A	B	C	D	E	F	G
1	胃病类型		铜蓝蛋白	蓝色反应	乙酸	硫化物	归类
2	胃病		228	134	20	11	1
3			245	134	10	40	1
4			200	167	12	27	1
5			170	150	7	8	1
6			100	167	20	14	1
7	非胃病		150	117	7	6	2
8			120	133	10	26	2
9			160	100	5	10	2
10			185	115	5	19	2
11			170	125	6	4	2
12			165	142	5	3	2
13			185	108	2	12	2
14	未知样		225	125	7	14	
15			100	117	7	2	
16			130	100	6	12	

```
%类内散度矩阵
 s1=cov(w1)*(r1-1);
s2=cov(w2)*(r2-1);
sw=s1+s2;

w=inv(sw)*(m1-m2)';
 %最优投影向量
y1=w'*m1';
y2=w'*m2';
w0=-1/2*(y1+y2);      % midpoint  threshold
```

	A	B	C	D	E	F	G
1	胃病类型		铜蓝蛋白	蓝色反应	乙酸	硫化物	归类
2	胃病		228	134	20	11	1
3			245	134	10	40	1
4			200	167	12	27	1
5			170	150	7	8	1
6			100	167	20	14	1
7	非胃病		150	117	7	6	2
8			120	133	10	26	2
9			160	100	5	10	2
10			185	115	5	19	2
11			170	125	6	4	2
12			165	142	5	3	2
13			185	108	2	12	2
14	未知样		225	125	7	14	
15			100	117	7	2	
16			130	100	6	12	

```
%分类
for i=1:r3
y(i)=sample(i,:)*w+w0;
  if   y(i)*(w'*w1(1,:)'+w0)>0
      y(i)=1;
else
    y(i)=2;
end
end
```

	A	B	C	D	E	F	G
1	胃病类型		铜蓝蛋白	蓝色反应	乙酸	硫化物	归类
2	胃病	w1	228	134	20	11	1
3			245	134	10	40	1
4			200	167	12	27	1
5			170	150	7	8	1
6			100	167	20	14	1
7	非胃病	w2	150	117	7	6	2
8			120	133	10	26	2
9			160	100	5	10	2
10			185	115	5	19	2
11			170	125	6	4	2
12			165	142	5	3	2
13			185	108	2	12	2
14	未知样	sample	225	125	7	14	
15			100	117	7	2	
16			130	100	6	12	

小结

1. PCA存在的问题
2. Fisher线性判别法
3. 实例分析