模式识别 U4 Fisher线性判别

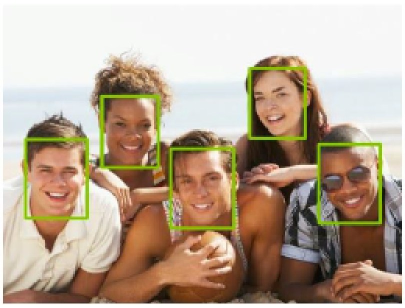
Fisher Discriminant

- 4.1 Fisher线性判别动机
- 4.2 Fisher线性判别分析
- 4.3 Fisher线性判别算法

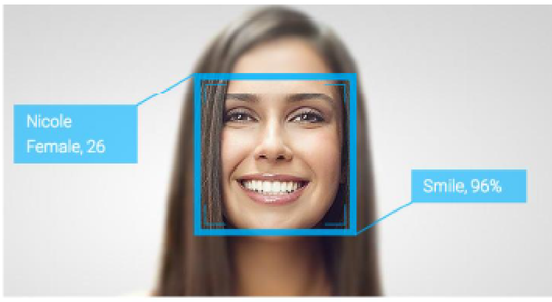
课程内容

Fisher线性判别 动机

应用示例:



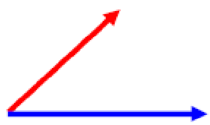
人脸检测
(Detection finds the faces in images)



人脸识别
(Recognition recognizes WHO the person is)

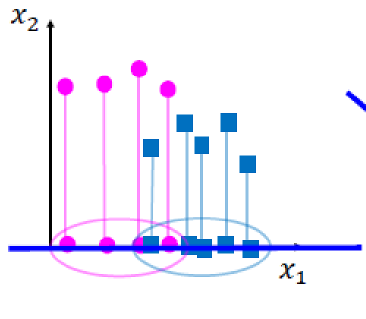
Fisher判别的核心思想是：在两个类别之间找到最好的区分，进行特征降维

投影表示：

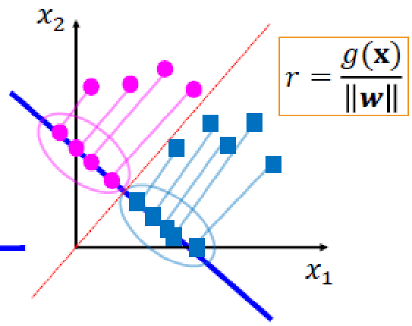


$w^T\mathbf{x} = \|\mathbf{x}\|\|w\|cos\theta$

将 x 向 w 投影
（Project x to w）

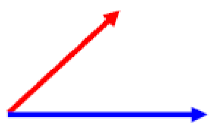


不是好的投影
（Poor Projection）



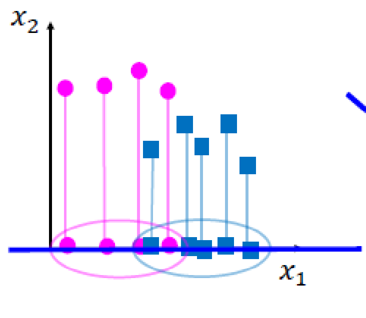
好的投影
（Good Projection）

投影表示：

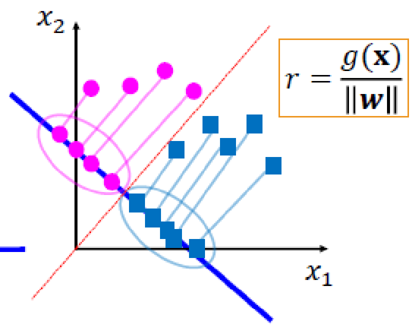


$w^T\mathbf{x} = \|\mathbf{x}\|\|w\|cos\theta$

将 x 向 w 投影
（Project x to w）

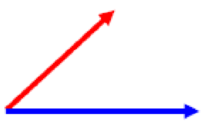


不是好的投影
（Poor Projection）



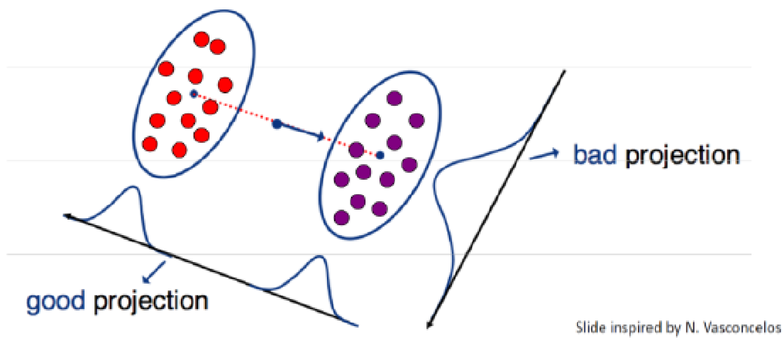
好的投影
（Good Projection）

投影表示：



$w^T\mathbf{x} = \|\mathbf{x}\|\|w\|cos\theta$

将 x 向 w 投影
（Project x to w）

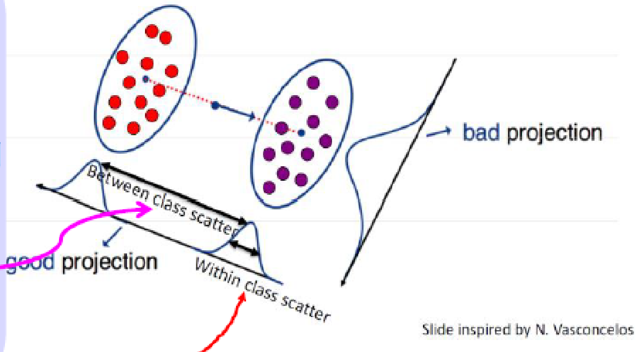


Fisher线性判别的目的：在两个类别之间找到最好的区分
（*find the best separation between two classes*）

Fisher判别 目的

Fisher线性判别的目的：

- 在尽可能保留类别可区分性的前提下实现维数减少
  （*Perform dimensionality reduction "while preserving as much of the class discriminatory information as possible".*）
- 找到让类别最好区分的投影方向
  （*Seeks to find directions along which the classes are best separated.*）
- 同时考虑类内散布和类间散布
  （*Takes into consideration the scatter within-classes but also the scatter between-classes.*）



Fisher线性判别 分析

线性回归目的：找到**误差最小的拟合模型**

二分类问题的Fisher线性判别：**学习最佳投影，它能将所有样本投影到w的方向**

二分类问题的Fisher线性判别：

学习最佳投影$\boldsymbol{w}^*$，它能将所有样本投影到$\boldsymbol{w}^*$的方向

假设 $s = \mathbf{w}^T\mathbf{x} \qquad \mathbf{x} \in \mathcal{R}^d, \qquad s \in \mathcal{R}^1$

类别集合：$\mathcal{C} = \{c|(1,-1)\}$

第 c 个类别的均值为：$\boldsymbol{\mu}_c = E[\mathbf{x}|y=c] = \frac{1}{N_c}\sum_{n=1}^{N_c}[\mathbf{x}_n|y=c]$

第 c 个类别的协方差为：$\boldsymbol{\Sigma}_c = E[(\mathbf{x}-\mu_c)(\mathbf{x}-\mu_c)^T|y=c] = \sum_{n=1}^{N_c}[(\mathbf{x}_n-\mu_c)(\mathbf{x}_n-\mu_c)^T|y=c]$

目标函数

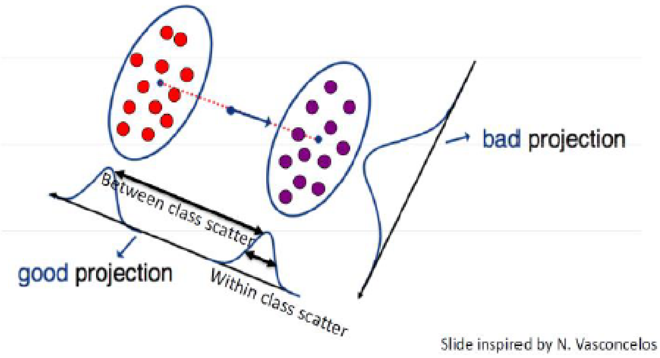
二分类问题的Fisher线性判别:

学习最佳投影$\boldsymbol{w}^*$的目标函数:

$$J(\boldsymbol{w}) = \frac{between\ class\ scatter}{within\ class\ scatter}$$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{argmax}\, J(\boldsymbol{w})$$

$$J(\boldsymbol{w}) = \frac{(E[s|y=1]-E[s|y=-1])^2}{var[s|y=1]+var[s|y=-1]}$$



代数推演过程

目标函数

$$\begin{aligned} J(\mathbf{w}) &= \frac{\text{类间差异}}{\text{类内差异}} \\ \mathbf{w}^* &= \arg\max_{\mathbf{w}} J(\mathbf{w}) \end{aligned}$$

在上述二分类问题中，则有

$$J(\mathbf{w}) = \frac{(\mathbb{E}[s|y=1]-\mathbb{E}[s|y=-1])^2}{var[s|y=1]+var[s|y=-1]}$$

对分子:

$$\begin{aligned} &(\mathbb{E}[s|y=1]-[s|y=-1])^2 \\ =&(\mathbb{E}[\mathbf{w}^T\mathbf{x}|y=1]-\mathbb{E}[\mathbf{w}^T\mathbf{x}|y=-1])^2 \\ =&\Big(\mathbf{w}^T(\mathbb{E}[\mathbf{x}|y=1]-\mathbb{E}[\mathbf{x}|y=-1])\Big)^2 \end{aligned}$$

根据概率论知识,

$\mathbb{E}[\mathbf{x}|y=c] = \frac{1}{N}\sum_{i=1}^{N_c}[x_i|y=c] = \mu_c$

因而我们可以改写上式：

$$
\begin{aligned}
&(\mathbb{E}[s|y=1]-[s|y=-1])^2\\
=&\Big(\mathbf{w}^T(\mathbb{E}[\mathbf{x}|y=1]-\mathbb{E}[\mathbf{x}|y=-1])\Big)^2\\
=&\Big(\mathbf{w}^T(\mu_1-\mu_{-1})\Big)^2\\
=&\ \mathbf{w}^T(\mu_1-\mu_{-1})(\mu_1-\mu_{-1})^T\mathbf{w}
\end{aligned}
$$

对分母：

根据协方差计算方法：
$var[s|y=c]=\mathbb{E}[(s-\mathbb{E}[s|y=c])^2]$

则有：

$$
\begin{aligned}
var[s|y=c] &= \mathbb{E}[(s-\mathbb{E}[s|y=c])^2]\\
&= \mathbb{E}[(\mathbf{w}^T\mathbf{x}-\mathbb{E}[\mathbf{w}^T\mathbf{x}|y=c])^2]\\
&= \mathbb{E}[\Big(\mathbf{w}^T(\mathbf{x}-\mathbb{E}[\mathbf{x}|y=c])\Big)^2]\\
&= \mathbb{E}[\Big(\mathbf{w}^T(\mathbf{x}-\mu_c)\Big)^2]\\
&= \mathbb{E}[\mathbf{w}^T(\mathbf{x}-\mu_c)(\mathbf{x}-\mu_c)^T\mathbf{w}]\\
&= \mathbf{w}^T\mathbb{E}[(\mathbf{x}-\mu_c)(\mathbf{x}-\mu_c)^T]\mathbf{w}\\
&= \mathbf{w}^T\Sigma_c\mathbf{w}
\end{aligned}
$$

因此:

$$var[s|y=c] = \mathbf{w}^T \Sigma_c \mathbf{w}$$

$$\Sigma_c = \frac{1}{N_C}\sum_{n=1}^{Nc}[(\mathbf{x}_n - \mu_c)(\mathbf{x}_n - \mu_c)^T|y=c]$$

综上:

二分类问题的Fisher线性判别:

学习最佳投影$\boldsymbol{w}^*$的目标函数:

$$J(\boldsymbol{w}) = \frac{between\ class\ scatter}{within\ class\ scatter}$$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{argmax}\, J(\boldsymbol{w})$$

$$J(\boldsymbol{w}) = \frac{(E[s|y=1]-E[s|y=-1])^2}{var[s|y=1]+var[s|y=-1]}$$

$$(E[s|y=1]-E[s|y=-1])^2$$
$$= \boldsymbol{w}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})^T\boldsymbol{w}$$

$$var[s|y=c] = E[(s-E[s|y=c])^2]$$
$$= \boldsymbol{w}^T\boldsymbol{\Sigma}_c\boldsymbol{w}$$

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})^T\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{\Sigma}_1\boldsymbol{w}+\boldsymbol{w}^T\boldsymbol{\Sigma}_{-1}\boldsymbol{w}}$$

二分类问题的Fisher线性判别:

学习最佳投影$\boldsymbol{w}^*$的目标函数:

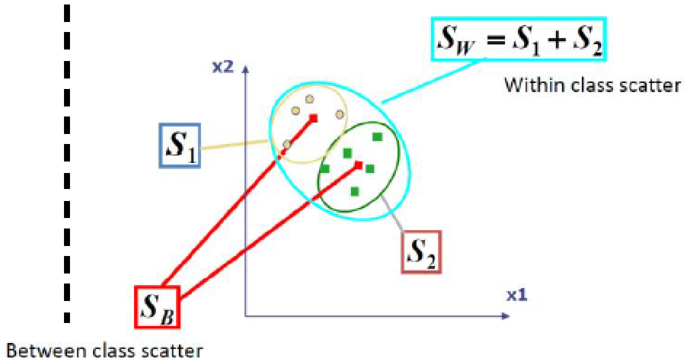
$$J(\boldsymbol{w}) = \frac{between\ class\ scatter}{within\ class\ scatter}$$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{argmax}\, J(\boldsymbol{w})$$

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T\boldsymbol{S}_B\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{S}_W\boldsymbol{w}}$$

$$\boldsymbol{S}_B = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_{-1})^T$$

$$\boldsymbol{S}_W = \boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_{-1} = \boldsymbol{S}_1+\boldsymbol{S}_2$$



优化问题：线性规划+拉格朗日乘数法

$$
\begin{aligned}
J(\mathbf{w}) &= \frac{(\mathbb{E}[s|y=1]-\mathbb{E}[s|y=-1])^2}{var[s|y=1]+var[s|y=-1]} \\
&= \frac{\mathbf{w}^T S_{B(between)}\mathbf{w}}{\mathbf{w}^T S_{W(within)}\mathbf{w}}
\end{aligned}
$$

$$
\mathbf{w}^* = \arg\max_{\mathbf{w}} J(\mathbf{w})
$$

分式的最优化不好处理，我们利用拉格朗日乘数法将其转化为易处理的形式

我们假定分母一定，此时取得分子的最大值，即可最大化目标函数

用数学语言表示为↓

$$
\arg\max_{\mathbf{w}} \ (\mathbf{w}^T S_B \mathbf{w}) \ \ Subject\ to \ (\mathbf{w}^T S_W \mathbf{w} = K)
$$

$$
Lagrange\ Multipliers:
$$

$$
\begin{aligned}
L(\mathbf{w},\lambda) &= \mathbf{w}^T S_B \mathbf{w} + \lambda(K - \mathbf{w}^T S_W \mathbf{w}) \\
&= \mathbf{w}^T (S_B - \lambda S_W)\mathbf{w} + \lambda K
\end{aligned}
$$

$$
\begin{aligned}
\text{令：} \nabla L_{\mathbf{w}}(\mathbf{w},\lambda) &= \frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}} = 2(S_B - \lambda S_W)\mathbf{w} = \mathbf{0}^T \\
\text{则：} S_B\mathbf{w} &= \lambda S_W \mathbf{w}
\end{aligned}
$$

二分类问题的Fisher线性判别：

学习最佳投影 $\boldsymbol{w}^*$ 的目标函数：

$$J(\boldsymbol{w}) = \frac{between\ class\ scatter}{within\ class\ scatter}$$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{argmax}\, J(\boldsymbol{w})$$

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_W \boldsymbol{w}}$$

$$\max_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} \quad Subject\ to\ \ \boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w} = \boldsymbol{K}$$

$$L(\boldsymbol{w},\lambda) = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} + \lambda(\boldsymbol{K} - \boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w})$$

$$\nabla L_w(\boldsymbol{w},\lambda) = \boldsymbol{0}$$

$$\boldsymbol{S}_B \boldsymbol{w} = \lambda \boldsymbol{S}_W \boldsymbol{w}$$

如果$\boldsymbol{S}_W^{-1} = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_{-1})^{-1}$ 存在, 则有：

$$\boldsymbol{S}_W^{-1} \boldsymbol{S}_B \boldsymbol{w} = \lambda \boldsymbol{w}$$

$$\boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})\,(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})^T \boldsymbol{w} = \lambda \boldsymbol{w}$$

$$\boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})\, a = \lambda \boldsymbol{w} \quad \boxed{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})^T \boldsymbol{w} = a\text{标量}}$$

$$\boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1}) = \frac{\lambda}{a} \boldsymbol{w}$$

只关注投影向量的方向 ：

$$\boldsymbol{w}^* = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})$$

二分类问题的Fisher线性判别：

学习最佳投影 $\boldsymbol{w}^*$ 的目标函数：

$$J(\boldsymbol{w}) = \frac{between\ class\ scatter}{within\ class\ scatter}$$

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{argmax}\, J(\boldsymbol{w})$$

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_W \boldsymbol{w}}$$

$$\max_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} \quad Subject\ to\ \ \boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w} = \boldsymbol{K}$$

$$L(\boldsymbol{w},\lambda) = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w} - \lambda(\boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w} - \boldsymbol{K})$$

$$\nabla L_w(\boldsymbol{w},\lambda) = \boldsymbol{0} \Rightarrow \boxed{\boldsymbol{w}^* = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})}$$

找到投影向量后，对任一测试样本 $\mathbf{x}$ ：

$$s = \boldsymbol{w}^{*T}\mathbf{x} = (\boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1}))^T \mathbf{x}$$

假设类别的判别门限设为 s' ：

$$\boxed{s' = \frac{\boldsymbol{w}^{*T}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_{-1})}{2}}$$

对任一测试样本 $\mathbf{x}$ 所属类别的判断：

$$\begin{cases} y = 1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} > s' \\ y = -1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} < s' \end{cases}$$

Fisher线性判别 算法

二分类问题的Fisher线性判别算法：

① 获取具有标签的两类样本

② 依据下式得到$\boldsymbol{\mu}_1$ 和$\boldsymbol{\mu}_{-1}$：

$$\boldsymbol{\mu}_c = \frac{1}{N_c}\sum_{n=1}^{N_c}[\mathbf{x}_n | y = c]$$

③ 依据下式得到$\boldsymbol{\Sigma}_1$ 和$\boldsymbol{\Sigma}_{-1}$ ：

$$\boldsymbol{\Sigma}_c = \sum_{n=1}^{N_c}[(\mathbf{x}_n - \mu_c)(\mathbf{x}_n - \mu_c)^T | y = c]$$

④ 计算**类内总离差阵**：$\boldsymbol{S}_W = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_{-1}$

⑤ 计算类内总离差阵的逆：$\boldsymbol{S}_W^{-1}$

⑥ 计算最佳投影：$\boldsymbol{w}^* = \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_{-1})$

⑦ 计算判别门限s'：$s' = \frac{\boldsymbol{w}^{*T}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_{-1})}{2}$

⑧ 对任一测试样本 $\mathbf{x}$ ：

$$\begin{cases} y = 1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} > s' \\ y = -1 & if\ s = \boldsymbol{w}^{*T}\mathbf{x} < s' \end{cases}$$

小结

4.1 Fisher线性判别动机

在尽可能保留类别可区分性的前提下实现维数减少

4.2 Fisher线性判别分析

找到让类别最好区分的投影方向

4.3 Fisher线性判别算法

通过计算类内散布和类间散布，找到最佳 $\boldsymbol{w}^*$ 和判别门限 s'

作业

纸质作业

1，已知两类样本的数据如下：

$\omega_1 : \{(5,37),(7,30),(10,35),(11.5,40),(14,38),(12,31)\}$

$\omega_2 : \{(35,21.5),(39,21.7),(34,16),(37,17)\}$

试用 Fisher 判别函数法，求出最佳投影方向 W，及分类阈值 y_0

T1. 解: $w_1: \{(5,37)^T, (7,30)^T, (10,35)^T, (11.5,40)^T, (14,38)^T, (12,31)^T\}$.

$w_2: \{(35,21.5)^T, (39,21.7)^T, (34,16)^T, (37,17)^T\}$.

$$\mu_1 = \frac{1}{6}\sum_{i=1}^{6} X_i^{(1)} = (9.92 \quad 35.17)^T \quad 2\times1$$

$$\mu_2 = \frac{1}{4}\sum_{i=1}^{4} X_i^{(2)} = (36.25 \quad 19.05)^T$$

计算类内离差阵:

$$\Sigma_1 = \sum_{i=1}^{6}(X_i^{(1)}-\mu_1)(X_i^{(1)}-\mu_1)^T \quad 2\times1 - 2\times1 \rightarrow 2\times2$$

$$= \begin{bmatrix} 56.21 & 16.58 \\ 16.58 & 78.83 \end{bmatrix}_{2\times2}$$

$$\Sigma_2 = \sum_{i=1}^{4}(X_i^{(2)}-\mu_2)(X_i^{(2)}-\mu_2)^T$$

$$= \begin{bmatrix} 14.75 & 9.55 \\ 9.55 & 26.53 \end{bmatrix}_{2\times2}$$

$$S_w = \Sigma_1 + \Sigma_2 = \begin{bmatrix} 70.96 & 26.13 \\ 26.13 & 105.36 \end{bmatrix}$$

$$S_w^{-1} = \frac{1}{|S_w|}\begin{bmatrix} 105.36 & -26.13 \\ -26.13 & 70.96 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0155 & -0.0038 \\ -0.0038 & 0.0104 \end{bmatrix}_{2\times2}$$

$$W^* = S_w^{-1}(\mu_1-\mu_2) = (-0.4704,\ 0.2696)^T$$

$$y_0 = W^{*T}\cdot\frac{\mu_1+\mu_2}{2} = -3.55$$

2，在 Fisher 判别中，用向量梯度的计算法则证明：$\boldsymbol{S_B w} = \lambda \boldsymbol{S_w w}$

T2. 解: $J(w)=\dfrac{w^T S_B w}{w^T S_W w}$

利用 Lagrange 乘数法:

$$\arg\max_{w}\ w^T S_B w \quad \text{Subject to} \quad w^T S_W w = K$$

$$\Rightarrow \quad L(w,\lambda) = w^T S_B w + \lambda(K - w^T S_W w)$$

$$= w^T(S_B - \lambda S_W)w + \lambda K$$

$$\frac{\partial L}{\partial w} = \frac{\partial(w^T(S_B-\lambda S_W)w)}{\partial w} + 0.$$

在矩阵计算中.

$$\frac{\partial u^T v}{\partial x} = u^T \frac{\partial v}{\partial x} + v^T \frac{\partial u}{\partial x} \qquad \frac{\partial Ax}{\partial x} = A \qquad \frac{\partial x}{\partial x} = I.$$

$$\therefore \frac{\partial L}{\partial w} = w^T(S_B-\lambda S_W)I + [(S_B-\lambda S_W)w]^T I$$

$$= w^T(S_B-\lambda S_W)I + w^T(S_B-\lambda S_W)^T I.$$

$\because S_B = (\mu_1-\mu_{-1})(\mu_1-\mu_{-1})^T \qquad S_W = \Sigma_1 + \Sigma_{-1}$

$\therefore S_B, S_W$ 均为对称矩阵 $\quad \therefore S_B - \lambda S_W = (S_B-\lambda S_W)^T$

$$\therefore \frac{\partial L}{\partial w} = 2w^T(S_B-\lambda S_W)I.$$

令 $\nabla L = 0 \Rightarrow \quad 2w^T(S_B-\lambda S_W)I = 0$

$$\Rightarrow 2(S_B-\lambda S_W)w = 0 \qquad S_B w = \lambda S_W w.$$