Discrimination & Classification

一、Bayes 分析框架

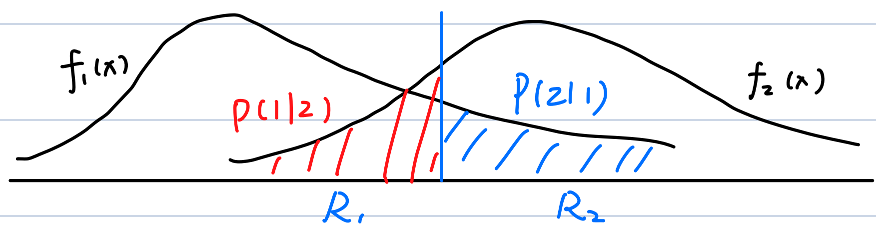
对于类别 π_1, π_2 的总体1，总体2，其概率密度为 $f_1(x), f_2(x)$

将样本集分2类 $\Omega = R_1 \cup R_2 \quad (R_1 \cap R_2 = \phi)$

误判概率：

$$P(2|1) = P(x \in R_2 | \pi_1) = \int_{R_2} f_1(x)\,dx$$

$$P(1|2) = P(x \in R_1 | \pi_2) = \int_{R_1} f_2(x)\,dx$$



(一). 先验概率

设属于 π_1, π_2 的先验概率 p_1, p_2 $(p_1+p_2=1)$ 则有后验概率：

$$P(\text{观测对象被错误判为}\ \pi_1) = P(1|2)\,p_2$$

$$P(\text{观测对象被错误判为}\ \pi_2) = P(2|1)\,p_1$$

(二). 错分代价

代价矩阵：

		分类为（样本）π_1	分类为（样本）π_2
真实总体	π_1	0	$c(2|1)$
真实总体	π_2	$c(1|2)$	0

平均(期望)错分代价：ECM

$$\min_{R_1, R_2} ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

$$= c(2|1)p_1 \int_{R_2} f_1(x)dx + c(1|2)p_2 \int_{R_1} f_2(x)dx$$

$$= c(2|1)p_1 [1 - \int_{R_1} f_1(x)dx] + c(1|2)p_2 \int_{R_1} f_2(x)dx$$

$$= c(2|1)p_1 + \int_{R_1} [c(1|2)p_2 f_2(x) - c(2|1)p_1 f_1(x)]dx$$

$$\Rightarrow \quad R_1 = \{ c(1|2)p_2 f_2(x) - c(2|1)p_1 f_1(x) \le 0 \}$$

$$= \left\{ \frac{f_1(x)}{f_2(x)} \ge \frac{c(1|2)}{c(2|1)} \cdot \frac{p_2}{p_1} \right\}$$

同理：

$$R_2 = \left\{ \frac{f_1(x)}{f_2(x)} \le \frac{c(1|2)}{c(2|1)} \cdot \frac{p_2}{p_1} \right\}$$

特殊 当 $c(1|2) = c(2|1)$ 时

判别规则： $R_1 : \frac{f_1(x)}{f_2(x)} \ge \frac{p_2}{p_1}$

$\min \ TPM = P(2|1)p_1 + P(1|2)p_2$ 总错分概率

后验概率判别规则：

$$P(\pi_1|x_0) = \frac{p_1 f_1(x_0)}{p_1 f_1(x_0) + p_2 f_2(x_0)}$$

$$P(\pi_2|x_0) = \frac{p_2 f_2(x_0)}{p_1 f_1(x_0) + p_2 f_2(x_0)}$$

$x_0 \in \pi \,|\, \max$ 后验

Tip: $c(1|2) = c(2|1)$ 时 后验与ECM等价

特殊：当 $c(1|2) = c(1|2)$ ， $p_1 = p_2$ 时

判别规则： $R_1 : \frac{f_1(x)}{f_2(x)} \ge 1$

此时即为 MLE

二、两正态总体的分类

(一). 同方差 $\Sigma_1=\Sigma_2=\Sigma$ (LDA 判别)

$\begin{cases} ① 假设1: 正态 f_1, f_2 \\ ② 假设2: \Sigma_1=\Sigma_2 \end{cases}$

记 $X=(X_1, X_2, \cdots, X_p)^T \in R^p$ 的概率密度为

$$f_i(x) = (2\pi)^{-\frac{p}{2}} \cdot |\Sigma|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)\}$$

其中 $i=1,2$ 代表 2 分类

分类规则: $R_1: \quad f_1(x)/f_2(x) \geq \frac{C(1|2)}{C(2|1)} \cdot P_2/P_1$

代入取对数

$$R_1: \quad -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) \geq \log\left(\frac{C(1|2)P_2}{C(2|1)P_1}\right)$$

其中 $LHS = (\mu_1-\mu_2)^T\Sigma^{-1}x - \frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1 - \mu_2^T\Sigma^{-1}\mu_2)$

$= (\mu_1-\mu_2)^T\Sigma^{-1}x - \frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)$

$= (\mu_1-\mu_2)^T\Sigma^{-1}(x-\frac{\mu_1+\mu_2}{2})$

样本估计

分类规则: 最小 ECM 法则

$$R_1: (\bar{x}_1-\bar{x}_2)^T S_p^{-1} x - \frac{1}{2}(\bar{x}_1-\bar{x}_2)^T S_p^{-1}(\bar{x}_1+\bar{x}_2) \geq \log\left(\frac{C(1|2)P_2}{C(2|1)P_1}\right)$$

或

$$R_1: (\bar{x}_1-\bar{x}_2)^T S_p^{-1}\left(x-\frac{\bar{x}_1+\bar{x}_2}{2}\right) \geq \log\left(\frac{C(1|2)P_2}{C(2|1)P_1}\right)$$

其中: $\bar{x}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1j} \qquad \bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}$

$S_1 = \frac{1}{n_1-1}\cdot\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)(x_{1j}-\bar{x}_1)^T$

$S_2 = \frac{1}{n_2-1}\cdot\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)(x_{2j}-\bar{x}_2)^T$

$S_p = \frac{n_1-1}{n_1+n_2-2}S_1 + \frac{n_2-1}{n_1+n_2-2}S_2$

特殊 当 $C(1|2)=C(2|1)$ $P_1=P_2$ 时

$$R_1: \quad (\bar{x}_1-\bar{x}_2)^T S_p^{-1} x \geq (\bar{x}_1-\bar{x}_2)^T S_p^{-1} \cdot \frac{\bar{x}_1+\bar{x}_2}{2}$$

记 $\hat{y} = \hat{a}^T x$ 其中 $\hat{a}^T = (\bar{x}_1-\bar{x}_2)^T S_p^{-1}$

记 $\hat{m} = \frac{1}{2}(\bar{y}_1+\bar{y}_2)$ 其中 $\bar{y}_1+\bar{y}_2 = (\bar{x}_1-\bar{x}_2) S_p^{-1}(\bar{x}_1+\bar{x}_2)$

$$= \hat{a}^T\bar{x}_1 + \hat{a}^T\bar{x}_2$$

即 $\bar{y}_i = \hat{a}^T\bar{x}_i$

判别

$$R_1: \quad \hat{a}^T x = \hat{y} \geq \hat{a}^T \cdot \frac{\bar{x}_1+\bar{x}_2}{2} = \hat{m}$$

Fisher 线性判别函数： $\hat{w} = \hat{y} - \hat{m}$

e.g. 对 $\forall$ 新值 $x_0 \Rightarrow \hat{y}_0 = \hat{a}^T x_0$，若 $\hat{y}_0 \geq \hat{m}$ 则属于 π_1

一般情况

$$R_1: (\bar{x}_1-\bar{x}_2)^T S_p^{-1} x - \frac{1}{2}(\bar{x}_1-\bar{x}_2)^T S_p^{-1}(\bar{x}_1+\bar{x}_2) \geq \log\left(\frac{C(1|2)P_2}{C(2|1)P_1}\right)$$

或

$$R_1: (\bar{x}_1-\bar{x}_2)^T S_p^{-1}\left(x-\frac{\bar{x}_1+\bar{x}_2}{2}\right) \geq \log\left(\frac{C(1|2)P_2}{C(2|1)P_1}\right)$$

（Tip：可以标准化后比较，$\hat{a}^* = \frac{\hat{a}}{\sqrt{\hat{a}^T\hat{a}}}$ x_i 均标准化 ）

（二）异方差 $\Sigma_1 \neq \Sigma_2$ （QDA 判别）

类似地，计算密度函数之比：

$$\frac{f_1(x)}{f_2(x)} = \exp\Big\{ -\frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) \Big\} \cdot \left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)^{-\frac{1}{2}}$$

$$= \exp\Big\{ -\frac{1}{2}x^T(\Sigma_1^{-1}-\Sigma_2^{-1})x + (\mu_1^T\Sigma_1^{-1}-\mu_2^T\Sigma_2^{-1})x - \frac{1}{2}(\mu_1^T\Sigma_1^{-1}\mu_1 + \mu_2^T\Sigma_2^{-1}\mu_2) - \frac{1}{2}\log\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) \Big\}$$

记 $k = -\frac{1}{2}(\mu_1^T\Sigma_1^{-1}\mu_1 + \mu_2^T\Sigma_2^{-1}\mu_2) - \frac{1}{2}\log\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)$

判别：

$$R_1: -\frac{1}{2}x'(\Sigma_1^{-1}-\Sigma_2^{-1})x + (\mu_1'\Sigma_1^{-1}-\mu_2'\Sigma_2^{-1})x - k \geqslant \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

$$R_2: -\frac{1}{2}x'(\Sigma_1^{-1}-\Sigma_2^{-1})x + (\mu_1'\Sigma_1^{-1}-\mu_2'\Sigma_2^{-1})x - k < \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (11\text{-}27)$$

三、评估分类函数

由之前的推导，总错分概率：

$$TPM = p_1\int_{R_2} f_1(x)\,dx + p_2\int_{R_1} f_2(x)\,dx$$

选择 R_1^*, R_2^* 使 min TPM 称为最优失误率 OER

$$OER = p_1\int_{R_2^*} f_1(x)\,dx + p_2\int_{R_1^*} f_2(x)\,dx$$

Rm：R_1^* R_2^* 是对总体的考虑，

(一). 正态总体假设（例）

由之前的推导：$(p_1=p_2 \quad c(1|2)=c(2|1) \quad \Sigma_1=\Sigma_2=\Sigma)$

$$y=(\mu_1-\mu_2)^T\Sigma^{-1}x := a^Tx$$

有：

$$R_1(y):\quad y \geqslant \frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)$$

$$R_2(y):\quad y < \frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)$$

由 $X \sim N(\mu,\Sigma)$ 故 $Y \sim f_1(\mu_{1Y},\sigma_Y^2)$ 或 $f_2(\mu_{2Y},\sigma_Y^2)$

$$\mu_{iY}=a^T\mu_i=(\mu_1-\mu_2)^T\Sigma^{-1}\mu_i \qquad i=1.2$$

$$\sigma_Y^2=a^T\Sigma a=(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)\triangleq\Delta^2$$

计算误分率：

$$TPM=p_1\cdot P(2|1)+p_2\cdot P(1|2) \qquad (p_1=p_2)$$

而 $P(2|1)=P(Y<\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)\mid Y\sim f_1)$

$$=P\left(\frac{Y-\mu_{1Y}}{\sigma_Y}<-\frac{1}{2}\Delta\right) \qquad (\because \text{实际为}1)$$

$$=P(Z<-\tfrac{1}{2}\Delta)$$

$$=\Phi(-\tfrac{1}{2}\Delta)$$

同理 $P(1|2)=P(Y\geqslant\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)\mid Y\sim f_2)$

$$=P\left(\frac{Y-\mu_{2Y}}{\sigma_Y}\geqslant\frac{1}{2}\Delta\right)$$

$$=P(Z\geqslant\tfrac{1}{2}\Delta)$$

$$=\Phi(-\tfrac{1}{2}\Delta)$$

$\Delta^2=\sigma_Y^2$

若 $p_1=p_2=\frac{1}{2}$ 有

$$OER=\frac{1}{2}\Phi(-\tfrac{1}{2}\Delta)+\frac{1}{2}\Phi(-\tfrac{1}{2}\Delta)=\Phi(-\tfrac{1}{2}\Delta)$$

Tip: 意义不大，因为我们不知总体信息

(二). 真实失误率与表观失误率

真实失误率 AER

$$AER = p_1 \int_{\hat{R}_2} f_1(x)\,dx + p_2 \int_{\hat{R}_1} f_2(x)\,dx$$

其中 $\hat{R}_1$ 与 $\hat{R}_2$ 为分类模型对样本训练后的分类：

$\hat{R}_1$: $(\bar{x}_1 - \bar{x}_2)^T S_p^{-1} x - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T S_p^{-1} (\bar{x}_1 + \bar{x}_2) \geq \log\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$

$\hat{R}_2$: $(\bar{x}_1 - \bar{x}_2)^T S_p^{-1} x - \frac{1}{2} (\bar{x}_1 - \bar{x}_2)^T S_p^{-1} (\bar{x}_1 + \bar{x}_2) < \log\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$

表观失误率 APER

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$$

$$\text{正确率} = \frac{n_{1C} + n_{2C}}{n_1 + n_2}$$

真阳性率，灵敏度，命中率，查出率 $\frac{n_{1C}}{n_1}$

假阳性率 $FPR = \frac{n_{2M}}{n_2}$

特异度 $1 - FPR = 1 - \frac{n_{2M}}{n_2}$

准确度 $\frac{n_{1C}}{n_{1C} + n_{2M}}$

样本预测

		π_1	π_2	
实际总体	π_1	n_{1C}	n_{1M}	n_1
	π_2	n_{2M}	n_{2C}	n_2

四. Fisher 线性判别

(一) 两总体　　　　(二) $p_1=p_2$　$c(2|1)=c(1|2)$ 下的最小ECM准则

$x_1, x_2, \cdots, x_n \in \mathbb{R}^p$ 来自2个类 π_1 和 π_2，各 n_1 n_2 个，即 $n_1+n_2=n$

记 $x_{(ij)} \in \{x_1, x_2, \cdots, x_n\}$ 为来自第 i 类的第 j 个个体，即：

$$\pi_1: \quad x_{(11)} \quad x_{(12)} \quad \cdots \quad x_{(1n_1)}$$

$$\pi_2: \quad x_{(21)} \quad x_{(22)} \quad \cdots \quad x_{(2n_2)}$$

总体均值、方差为 $\mu_1, \mu_2 \in \mathbb{R}^p$　$\Sigma_1, \Sigma_2 \in \mathbb{R}^{p\times p}$；类似地样本参数为 $\bar{x}_1, \bar{x}_2 \in \mathbb{R}^p$　$S_1, S_2 \in \mathbb{R}^{p\times p}$

Fisher 假设：　$\Sigma_1 = \Sigma_2$

分离度：　$\dfrac{|\bar{y}_1 - \bar{y}_2|}{S_y}$　　$D^2=(\bar{x}_1-\bar{x}_2)^T S_p^{-1}(\bar{x}_1-\bar{x}_2)$

其中 $S_y^2 = \dfrac{1}{n_1+n_2-2}\left(\sum_{j=1}^{n_1}(y_{(1j)}-\bar{y}_1)^2 + \sum_{j=1}^{n_2}(y_{(2j)}-\bar{y}_2)^2\right) = \hat{a}^T S_p \hat{a}$

而 y 为 x 的线性组合：← 与之前 Bayes 推出结果相同

$$\hat{y} = \hat{a}^T x = (\bar{x}_1 - \bar{x}_2)^T S_p^{-1} x$$

分离度：　$(\bar{y}_1-\bar{y}_2)^T(\bar{y}_1-\bar{y}_2)/\hat{a}^T S_p \hat{a} = (\hat{a}^T d)^T(\hat{a}^T d)/\hat{a}^T S_p \hat{a}$

Proof：证明 $\hat{y}$ 可使分离度 $\dfrac{|\hat{\bar{y}}_1 - \hat{\bar{y}}_2|}{S_y}$ 最大　$= d^T S_p^{-1} d$

$$\max_{\hat{a}} \frac{(\bar{y}_1-\bar{y}_2)^2}{S_y^2} = \frac{(\hat{a}^T\bar{x}_1 - \hat{a}^T\bar{x}_2)^2}{\hat{a}^T S_p \hat{a}} = \frac{(\hat{a}^T d)^2}{\hat{a}^T S_p \hat{a}}$$

其中 $d = \bar{x}_1 - \bar{x}_2$

分离度 $(\bar{x}_1-\bar{x}_2)^T S_p^{-1} (\bar{x}_1-\bar{x}_2)$

由之前的引理：$\boxed{\max = d^T S_p^{-1} d}$　当 $\hat{a}^* = c\cdot S_p^{-1}(\bar{x}_1-\bar{x}_2)$

五、多总体判别

设有 g 个总体/类，其概率密度为 $f_i(x)$ $i=1,2,\cdots,g$ 记先验概率为 p_i $i=1,2,\cdots,g$，记 $c(k|i)$ 为将实际为 π_i 的个体划分到 π_k 的代价。

Bayes

设划分结果为 R_k $k=1,2,\cdots,g$ 有

$$P(k|i) = \int_{R_k} f_i(x)\,dx$$

单个误判代价为

$$ECM(i) = \sum_{k\neq i} P(k|i)\cdot c(k|i) \qquad i=1,2,\cdots,g$$

期望误判代价为

$$ECM = \sum_{i=1}^{g} ECM(i)\cdot p_i$$

目标：

$$\begin{aligned}\min_{R_1 R_2\cdots R_g} ECM &= \sum_{i=1}^{g} ECM(i)\cdot p_i \\ &= \sum_{i=1}^{g} p_i \sum_{k\neq i} P(k|i)\cdot c(k|i) \\ &= \sum_{i=1}^{g} p_i \sum_{k\neq i}\left(\int_{R_k} f_i(x)\,dx\right)\cdot c(k|i)\end{aligned}$$

结论：最小化 ECM 可化为

$$\min_{k} \sum_{\substack{i=1 \\ i\neq k}}^{g} p_i f_i(x)\cdot c(k|i)$$

将 x 分配给 π_k $(k=1,2,\cdots,g)$ （证明略）

简化：设 $c(k|i) = 1$

错分代价相同时的最小 ECM 分类法则

若

$$p_k f_k(x) > p_i f_i(x), \text{任意 } i \neq k \tag{11-40}$$

则将 x 分配到 π_k，或等价的，若

$$\ln p_k f_k(x) > \ln p_i f_i(x), \text{任意 } i \neq k \tag{11-41}$$

则将 x 分配到 π_k。

上式 (11-40) 等价于："后验"概率 $P(\pi_k | x) = P$(给定观测值 x，x 来自 π_k) 最大化的法则，其中：

$$P(\pi_k | x) = \frac{p_k f_k(x)}{\sum_{i=1}^{g} p_i f_i(x)} \quad (= \text{(先验} \times \text{密度)归一化})$$

正态假设：

计算 $\log(p_k f_k(x))$ $\forall x \in \mathbb{R}^p$

$$= \log p_k - \frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)$$

记

$$d_k^Q(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) + \log p_k$$

样本估计

$$\hat{d}_k^Q(x) = -\frac{1}{2}\log|S_k| - \frac{1}{2}(x-\bar{x}_k)^T S_k^{-1}(x-\bar{x}_k) + \log p_k$$

$$k = 1, 2, \cdots, g$$

Σ_k 相同假设：

x相同时全相同，可省略

$$d_k^Q(x) = \underline{-\frac{1}{2}\log|\Sigma| - \frac{1}{2}x^T\Sigma^{-1}x} + \mu_k^T\Sigma^{-1}x - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log p_k$$

$$d_k(x) = \mu_k^T\Sigma^{-1}x - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log p_k$$

$\leftarrow d_k(x)$ 线性判别得分

Rule: $x_0 \in \pi_k \mid$ 若 $d_k^Q(x)$ 最大

样本估计：

$$S_p = \frac{1}{(n_1+n_2+\cdots+n_g)-g}\sum_{i=1}^{g}(n_i-1)S_i$$

故

$$\hat{d}_k(x) = \bar{x}_k^T S_p^{-1} x - \frac{1}{2}\bar{x}_k^T S_p^{-1}\bar{x}_k + \log p_k \quad \left(-\frac{1}{2}x^T S_p^{-1} x\right)$$

补

$$= -\frac{1}{2}(x-\bar{x}_k)^T S_p^{-1}(x-\bar{x}_k) + \log p_k$$

$$= -\frac{1}{2}D_k^2(x) + \log p_k$$

其中 $D_k^2(x) = (x-\bar{x}_k)^T S_p^{-1}(x-\bar{x}_k) \qquad k=1,2,\cdots,g$

简化： $p_1 = p_2 = \cdots = p_g \Rightarrow \log p_k$ 相同

可记 $\hat{d}_k(x) = -\frac{1}{2}D_k^2(x)$ （距离判别）

$$= -\frac{1}{2}(x-\bar{x}_k)^T S_p^{-1}(x-\bar{x}_k) \qquad k=1,2,\cdots,g$$

六、多总体 Fisher 判别

(一) 总体

Fisher

与上面问题相同，但假设只为

$$\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$$

记 $\bar{\mu}$ 为联合总体均值向量

$$\bar{\mu} = \frac{1}{g}\sum_{i=1}^{g}\mu_i \in R^p$$

记

$$B_\mu = \sum_{i=1}^{g}(\mu_i-\bar{\mu})(\mu_i-\bar{\mu})^T$$

对个体 $X \in \mathbb{R}^p$ 作线性变换：

$$Y = a^T X \qquad a \in \mathbb{R}^p$$

于是有：

$$E(Y) = a^T \mu_i \qquad \text{对每个总体 } \pi_i \quad i=1,2,\cdots,g$$

$$Var(Y) = a^T \Sigma a \qquad \text{对 } \forall \pi_i \text{ 均相同}$$

记

$$\mu_Y = \frac{1}{g}\sum_{i=1}^{g} \mu_{iY} = a^T \bar{\mu} \quad \in \mathbb{R}$$

其中 $\mu_{iY} = a^T \mu_i \in \mathbb{R}$

组间差距：

$$\frac{\sum_{i=1}^{g}(\mu_{iY}-\mu_Y)^2}{Var(Y)} = \frac{\sum_{i=1}^{g}(a^T\mu_i - a^T\bar{\mu})^2}{a^T\Sigma a}$$

$$= \frac{a^T\left(\sum_{i=1}^{g}(\mu_i-\bar{\mu})(\mu_i-\bar{\mu})^T\right)a}{a^T\Sigma a}$$

$$= \frac{a^T B_\mu a}{a^T \Sigma a}$$

目标：最大化组间差距

$$\max_a \frac{a^T B_\mu a}{a^T \Sigma a}$$

[引理1] 设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$ 为 $\Sigma^{-1}B_\mu$ 的 $s \leq \min(g-1, p)$ 个正特征值，$e_1, e_2, \cdots, e_s$ 为相应特征向量，且使得

$$e_i^T \Sigma e_i = 1$$

于是，能使 $\frac{a^TB_\mu a}{a^T\Sigma a}$ 最大化的 $a_1 = e_1$ 在满足约束

$Cov(a_1^T X, a_2^T X) = 0$ 条件下，使上述比值最大

以此类推： $a_k = e_k$ 使 $Cov(a_k^T X, a_i^T X) = 0$

当 $i<k$ 时，使上式比值最大

故找到对应的 $a_k = e_k$ $k = 1, 2, \cdots, g$

称 $a_k^T X$ 为第 k 判别量。（满足 $Var(a_k^T X) = 0$）

[引理 2] 记 $y_j = a_j^T X$ 其中 a_j 为 $\Sigma^{-1} B_\mu$ 的特征向量

就有：

$$\sum_{j=1}^{p} (y_j - \mu_{iy_j})^2 = (x-\mu_i)^T \Sigma^{-1} (x-\mu_i)$$

$$= -2 d_i(x) + x^T \Sigma^{-1} x + 2\log p_i$$

若 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0 = \lambda_{s+1} = \cdots = \lambda_p$ 则

$\sum_{j=s+1}^{p} (y_j - \mu_{iy_j})^2$ 对所有 π_i $(i = 1, 2, \cdots, g)$ 均为常值，即

只有前 s 个判别量 y_j 对分类有贡献（不妨设 $\forall p_i = \frac{1}{g}$）

判别规则 若只用前 r $(r \leq s)$ 个判别量：

$$\sum_{j=1}^{r} (y_j - \mu_{k y_j})^2 = \sum_{j=1}^{r} (a_j^T (x - \mu_k))^2$$

$$\leq \sum_{j=1}^{r} (a_j^T (x - \mu_i)) \quad \forall i \neq k$$

对 $\forall i \neq k$ 成立，则判断 $x \in \pi_k$

$\Sigma^{-1} B_\mu$

Tip：离 μ_k 更近，故为 π_k

[引理 3] 定义分离变量：

$$\Delta_s^2 = \sum_{i=1}^{g} (\mu_i - \bar{\mu})^T \Sigma^{-1} (\mu_i - \bar{\mu})$$

也可表示为

$$\Delta_s^2 = \lambda_1 + \lambda_2 + \cdots + \lambda_s$$

其中 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$ 为 $\Sigma^{-1} B_\mu$ 的正特征值

$$B_\mu = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T$$

(二) 样本 （下面所有的 $x \in \mathbb{R}^p$）

$$B = \sum_{i=1}^{g} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \qquad \bar{x} = \frac{1}{g} \sum_{i=1}^{g} \bar{x}_i$$

$$W = \sum_{i=1}^{g} (n_i - 1) S_i \qquad S_i = \sum_{l=1}^{n_i} (x_{il} - \bar{x}_i)(x_{il} - \bar{x}_i)^T$$

$$S_p = \frac{W}{n_1 + n_2 + \cdots + n_g - g}$$

此时目标

$$\max_{\hat{a}} \frac{\hat{a} B \hat{a}}{\hat{a} W \hat{a}}$$

判别规则 $W^{-1}B$

计算 $W^{-1}B$ 的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$

和对应的 $e_1, e_2, \cdots, e_s$ 有 将 $\mu_i \Rightarrow \bar{x}_i$

$$y_j = e_j^T x \in \mathbb{R} \qquad (\mu_{iY_j} = e_j^T \mu_i)$$

$\sum_{j=1}^{r \leq s} (y_j - \mu_{kY_j})^2 \leq \sum_{j=1}^{r \leq s} (y_j - \mu_{iY_j})^2 \ \forall i \neq k$ 则 $y_j = e_j^T x \in \pi_k$