2.5 证明 MSE 方法在选两类样本的 b 值分别为 N/N_1 和 N/N_2 时得到的解等价于 Fisher 线性判别的解且 $w_0 = -\hat{m}$，其中 N, N_1, N_2 分别为样本总数和两类样本的数目。

证明：这里延续课件的表达，有 MSE 的解如下形式

$$\boldsymbol{a}^* = (\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T\boldsymbol{b}$$

样本分为两类分别为 X_1, X_2，于是 $Y = \begin{bmatrix} I_1 & X_1 \\ -I_2 & X_2 \end{bmatrix}$，$a = \begin{bmatrix} w_0 \\ w \end{bmatrix}$

根据题目要求有 $b = \begin{bmatrix} \frac{N}{N_1} & I_1 \\ \frac{N}{N_2} & I_2 \end{bmatrix}$

由 $Y^TYa = Y^Tb$ 得 $\begin{bmatrix} I_1^T & -I_2^T \\ X_1^T & -X_2^T \end{bmatrix}\begin{bmatrix} I_1 & X_1 \\ -I_2 & -X_2 \end{bmatrix}\begin{bmatrix} w_0 \\ w \end{bmatrix} = \begin{bmatrix} I_1^T & -I_2^T \\ X_1^T & -X_2^T \end{bmatrix}\begin{bmatrix} \frac{N}{N_1} & I_1 \\ \frac{N}{N_2} & I_2 \end{bmatrix}$

$m_1 = \frac{1}{N_1}\sum_1 x_i$，$m_2 = \frac{1}{N_2}\sum_2 x_i$

$$\begin{aligned} S_w &= \sum\nolimits_1 (x_i - m_1)(x_i - m_1)^T + \sum\nolimits_2 (x_i - m_2)(x_i - m_2)^T \\ &= X_1^TX_1 - N_1m_1m_1^T + X_2^TX_2 - N_2m_2m_2^T \end{aligned}$$

进而有

$$\begin{bmatrix} N & (N_1m_1 + N_2m_2)^T \\ N_1m_1 + N_2m_2 & S_w + N_1m_1m_1^T + N_2m_2m_2^T \end{bmatrix}\begin{bmatrix} w_0 \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ N(m_1 - m_2) \end{bmatrix}$$

于是

$$Nw_0 + N_1m_1^Tw + N_2m_2^Tw = 0$$
$$Nw_0 + Nm^Tw = 0$$

得

$$w_0 = -m^Tw = -\hat{m}$$

又

$$(N_1m_1 + N_2m_2)w_0 + (S_w + N_1m_1m_1^T + N_2m_2m_2^T)w = N(m_1 - m_2)$$
$$-(N_1m_1 + N_2m_2)m^Tw + (S_w + N_1m_1m_1^T + N_2m_2m_2^T)w = N(m_1 - m_2)$$
$$-\frac{(N_1m_1 + N_2m_2)(N_1m_1^T + N_2m_2^T)}{N}w + (N_1m_1m_1^T + N_2m_2m_2^T)w + S_ww = N(m_1 - m_2)$$
$$\frac{N_1N_2}{N}(m_1 - m_2)(m_1 - m_2)^Tw + S_ww = N(m_1 - m_2)$$

$\frac{N_1N_2}{N}(m_1 - m_2)(m_1 - m_2)^Tw$ 的方向与 $(m_1 - m_2)$ 相同，于是 S_ww 的方向也必与 $(m_1 - m_2)$ 相同，因此 $S_ww = k(m_1 - m_2)$，得 $w = kS_w^{-1}(m_1 - m_2)$，与 Fisher 线性判别的解等价。

最优判别的渐近逼近

$B=I_n$时，MSE的解等同于以最小均方误差逼近Bayes 判别函数：

$$g_0(\mathbf{x}) = P(\omega_1 \mid \mathbf{x}) - P(\omega_2 \mid \mathbf{x})$$

即MSE的分类面
接近贝叶斯分类面

证明：按照概率定律

$$p(\mathbf{x}) = p(\mathbf{x} \mid \omega_1)P(\omega_1) + p(\mathbf{x} \mid \omega_2)P(\omega_2)$$

独立同分布抽取样本，得到

$$g(\mathbf{x}) = \mathbf{a}^t\mathbf{y}, \quad \mathbf{y} = \mathbf{y}(\mathbf{x})$$

定义均方逼近误差为

$$\varepsilon^2 = \int \left[\mathbf{a}^t\mathbf{y} - g_0(\mathbf{x})\right]^2 p(\mathbf{x})d\mathbf{x}$$

最优判别的渐近逼近

当$\mathbf{b}=\mathbf{1}_n$时，最小均方误差准则函数为：

$$J_s(\mathbf{a}) = \sum_{\mathbf{y}\in Y_1}\left(\mathbf{a}^t\mathbf{y}-1\right)^2 + \sum_{\mathbf{y}\in Y_2}\left(\mathbf{a}^t\mathbf{y}+1\right)^2$$

?

$$= n\left[\frac{n_1}{n}\frac{1}{n_1}\sum_{\mathbf{y}\in Y_1}\left(\mathbf{a}^t\mathbf{y}-1\right)^2 + \frac{n_2}{n}\frac{1}{n_2}\sum_{\mathbf{y}\in Y_2}\left(\mathbf{a}^t\mathbf{y}+1\right)^2\right]$$

利用大数定理，n趋向无穷大时

$$J(\mathbf{a}) = \frac{1}{n}J_s(\mathbf{a}) = P(\omega_1)E_1\left[\left(\mathbf{a}^t\mathbf{y}-1\right)^2\right] + P(\omega_2)E_2\left[\left(\mathbf{a}^t\mathbf{y}+1\right)^2\right]$$

其中，$E_1\left[\left(\mathbf{a}^t\mathbf{y}-1\right)^2\right] = \int\left(\mathbf{a}^t\mathbf{y}-1\right)^2 p(\mathbf{x} \mid \omega_1)d\mathbf{x}$

$$E_2\left[\left(\mathbf{a}^t\mathbf{y}+1\right)^2\right] = \int\left(\mathbf{a}^t\mathbf{y}+1\right)^2 p(\mathbf{x} \mid \omega_2)d\mathbf{x}$$

最优判别的渐近逼近

$$g_0(\mathbf{x}) = \frac{p(\mathbf{x}, \omega_1) - p(\mathbf{x}, \omega_2)}{p(\mathbf{x})}$$

$$\bar{J}(\mathbf{a}) = \int (\mathbf{a}^t\mathbf{y} - 1)^2 p(\mathbf{x}, \omega_1) d\mathbf{x} + \int (\mathbf{a}^t\mathbf{y} + 1)^2 p(\mathbf{x}, \omega_2) d\mathbf{x}$$

$$= \int (\mathbf{a}^t\mathbf{y})^2 p(\mathbf{x}) d\mathbf{x} - 2\int \mathbf{a}^t\mathbf{y} g_0(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} + 1$$

$$= \int [\mathbf{a}^t\mathbf{y} - g_0(\mathbf{x})]^2 p(\mathbf{x}) d\mathbf{x} + \left[1 - \int g_0^2(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}\right]$$

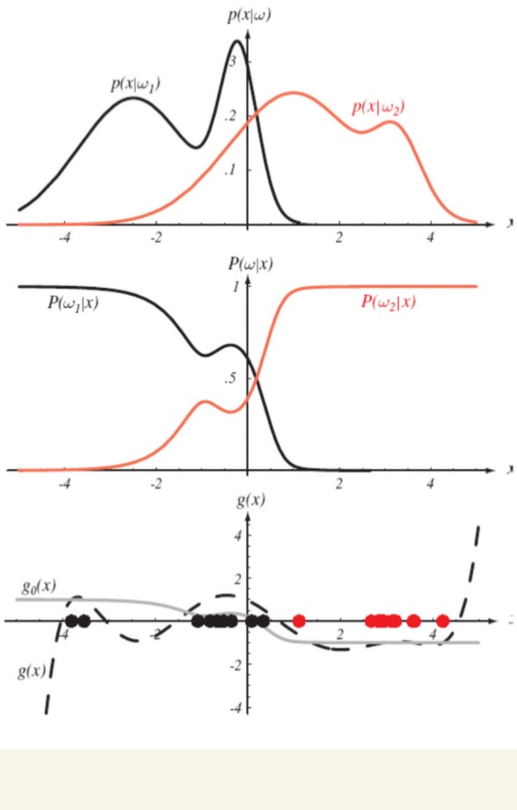
$$= \varepsilon^2 + \text{独立于}\mathbf{a}\text{的成分}$$

于是，MSE的解等同于以最小均方

误差逼近Bayes 判别函数：

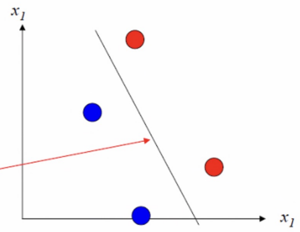
$$g_0(\mathbf{x}) = P(\omega_1 \mid \mathbf{x}) - P(\omega_2 \mid \mathbf{x})$$

MSE 解和 Bayes 判别函数



例解

- ω_1: $(1,2)^t$ and $(2,0)^t$
- ω_2: $(3,1)^t$ and $(2,3)^t$



$$a^t\begin{pmatrix}1\\x_1\\x_2\end{pmatrix}=0$$

Sample Matrix (d = 1+2, n = 4)

$$Y=\begin{bmatrix}1 & 1 & 2\\ 1 & 2 & 0\\ -1 & -3 & -1\\ -1 & -2 & -3\end{bmatrix}$$

Assuming $b=\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$

Pseudo-inverse

$$Y^*=(Y^tY)^{-1}Y^t=\begin{bmatrix}5/4 & 13/12 & 3/4 & 7/12\\ -1/2 & -1/6 & -1/2 & -1/6\\ 0 & -1/3 & 0 & -1/3\end{bmatrix}$$

our solution is $a=Y^tb=\begin{bmatrix}11/3\\-4/3\\-2/3\end{bmatrix}$

MSE解与 Bayes 判别函数的关系

- 当 $N\to\infty$， $\mathbf{b}=\mathbf{1}_n$ 时，MSE的解以最小均方误差逼近 Bayes 判别函数：$g_0(\mathbf{x})=p(\omega_1|\mathbf{x})-p(\omega_2|\mathbf{x})$;
  - 均方误差 $e^2=\int\left[\mathbf{a}^T\mathbf{y}-g_0(\mathbf{x})\right]^2 p(\mathbf{x})d\mathbf{x}$;
  - 即有 $\mathbf{a}^*\equiv Y^+\mathbf{b}=\arg\min e^2$.

$$J_s(\mathbf{a})=\sum_{\mathbf{x}_i\in\omega_1}(\mathbf{a}^T\mathbf{y}_i-1)^2+\sum_{\mathbf{x}_i\in\omega_2}(\mathbf{a}^T\mathbf{y}_i+1)^2$$

$$=N\left(\frac{N_1}{N}\frac{1}{N_1}\sum_{\mathbf{x}_i\in\omega_1}(\mathbf{a}^T\mathbf{y}_i-1)^2+\frac{N_2}{N}\frac{1}{N_2}\sum_{\mathbf{x}_i\in\omega_2}(\mathbf{a}^T\mathbf{y}_i+1)^2\right);$$

MSE解与 Bayes 判别函数的关系

$$\frac{J_s(\mathbf{a})}{N}$$

$$\doteq P(\omega_1)\int(\mathbf{a}^T\mathbf{y}-1)^2p(\mathbf{x}|\omega_1)d\mathbf{x}+P(\omega_2)\int(\mathbf{a}^T\mathbf{y}+1)^2p(\mathbf{x}|\omega_2)d\mathbf{x}$$

$$=\int(\mathbf{a}^T\mathbf{y}-1)^2p(\mathbf{x},\omega_1)d\mathbf{x}+\int(\mathbf{a}^T\mathbf{y}-1)^2p(\mathbf{x},\omega_2)d\mathbf{x}$$

$$=\int\left\{\left(\mathbf{a}^T\mathbf{y}\right)^2\left[p(\mathbf{x},\omega_1)+p(\mathbf{x},\omega_2)\right]-2\mathbf{a}^T\mathbf{y}\frac{p(\mathbf{x},\omega_1)-p(\mathbf{x},\omega_2)}{p(\mathbf{x})}\times p(\mathbf{x})+\left[p(\mathbf{x},\omega_1)+p(\mathbf{x},\omega_2)\right]\right\}d\mathbf{x}$$

$$=\int\left(\mathbf{a}^T\mathbf{y}\right)^2p(\mathbf{x})d\mathbf{x}-2\int\left(\mathbf{a}^T\mathbf{y}\right)g_0(\mathbf{x})p(\mathbf{x})d\mathbf{x}+1$$

$$=\int\left[\mathbf{a}^T\mathbf{y}-g_0(\mathbf{x})\right]^2p(\mathbf{x})d\mathbf{x}+\left[1-\int g_0^2(\mathbf{x})p(\mathbf{x})d\mathbf{x}\right]$$

$$=e^2+\left[1-\int g_0^2(\mathbf{x})p(\mathbf{x})d\mathbf{x}\right];$$

求解MSE解

- 计算 $\mathbf{a}^*=Y^\dagger\mathbf{b}$，其中 $Y^\dagger = (Y^\mathrm{T}Y)^{-1}Y^\mathrm{T}$；
- 梯度下降法

$$\nabla J_s(\mathbf{a}) = \sum_{i=1}^{N} 2(\mathbf{a}^T\mathbf{y}_i - b_i)\mathbf{y}_i = 2Y^T(Y\mathbf{a}-\mathbf{b})$$

$$\begin{cases} \mathbf{a}(1), \text{任意初始化} \\ \mathbf{a}(k+1) = \mathbf{a}(k) - \eta(k)Y^T(Y\mathbf{a}-\mathbf{b}) \end{cases}$$

until $\nabla J_s(\mathbf{a}) \le \theta$ 或者 $\|\mathbf{a}(k+1)-\mathbf{a}(k)\| \le \theta$;

可选 $\eta(k) = \eta(1)/k \ (\eta(1) > 0)$.

求解MSE解

- 单样本修正法（Widrow-Hoff 算法，LMS）

$$\begin{cases} \mathbf{a}(1), \text{任意初始化} \\ \mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k)(b_k - \mathbf{a}(k)^T\mathbf{y}^k)\mathbf{y}^k \end{cases}$$

其中 $\mathbf{y}^k$ 是使得 $\mathbf{a}(k)^T\mathbf{y}^k \ne b_k$ 的样本。

Algorithm 10 (LMS)

```
begin initialize a, b, criterion θ, η(·), k = 0
    do k ← k + 1
        a ← a + η(k)(b_k − a^t y^k) y^k
    until η(k)(b_k − a^t y^k) y^k < θ
    return a
end
```