模式识别 6　　陈斯杰　2016310721

Problem 1

1.1 Show that the optimal w_0 is the $w_0 = -w^T m$

$$\frac{\partial E}{\partial w_0} = \frac{1}{2}\sum_{n=1}^{N} 2(w^T x_n + w_0 - t_n)$$

$$= \left(\sum_{n=1}^{N} w^T x_n\right) + (N w_0) - \sum_{n=1}^{N} t_n$$

$$= \left(w^T \sum_{n=1}^{N} x_n\right) + N w_0 - \left(N_1 \cdot \frac{N}{N_1} + N_2 \cdot \frac{-N}{N_2}\right)$$

$$= \left(w^T \sum_{n=1}^{N} x_n\right) + N w_0$$

$$= 0 \Rightarrow w_0 = -\frac{w^T \sum_{n=1}^{N} x_n}{N} = -w^T m$$

1.2 把E对w求偏导:

$$\frac{\partial E}{\partial w} = \sum_{n=1}^{N} (w^T x_n + w_0 - t_n) x_n$$

$$= \sum_{n=1}^{N} (w^T x_n - w^T m - t_n) x_n$$

$$= \sum_{n=1}^{N} \left[(w^T x_n - w^T m) x_n - t_n x_n\right]$$

$$= \sum_{n=1}^{N} \left[(x_n x_n^T - x_n m^T) w - x_n t_n\right]$$

$$= \sum_{n\in C_1} \left[(x_n x_n^T - x_n m^T) w - x_n t_n\right] + \sum_{n\in C_2} \left[(x_n x_n^T - x_n m^T) w - x_n t_n\right]$$

$$= \left(\sum_{n\in C_1} x_n x_n^T - N_1 m_1 m^T\right) w - N_1 m_1 \cdot \frac{N}{N_1}$$

$$+ \left(\sum_{n\in C_2} x_n x_n^T - N_2 m_2 m^T\right) w + N_2 m_2 \cdot \frac{N}{N_2}$$

$$= \left(\sum_{n\in C_1} x_n x_n^T + \sum_{n\in C_2} x_n x_n^T - N_1 m_1 m^T - N_2 m_2 m^T\right) - N(m_1 - m_2)$$

$$= \left[\sum_{n\in C_1} x_n x_n^T + \sum_{n\in C_2} x_n x_n^T - (N_1 m_1 + N_2 m_2) m^T\right] - N(m_1 - m_2)$$

模式识别 6　陈斯杰　2016310721

对于任一类 C_K，有以下性质：

$$\sum_{i\in C_K}(x_i-m_K)(x_i-m_K)^T=\sum_{i\in C_K}(x_ix_i^T-x_im_K^T-m_Kx_i^T+m_Km_K^T)$$

$$=\left(\sum_{i\in C_K}x_ix_i^T\right)-0-0+\left(\sum_{i\in C_K}m_Km_K^T\right)$$

$$=\sum_{i\in C_K}x_ix_i^T-N_K\cdot m_Km_K^T$$

所以上式可以写成：

$$RHS=\left[\left(\sum_{n\in C_1}x_nx_n^T-N_1m_1m_1^T\right)+\left(\sum_{n\in C_2}x_nx_n^T-N_2m_2m^T\right)\right]w-N(m_1-m_2)$$

$$\cancel{=\left[\left(\sum_{n\in C_1}x_nx_n^T-N_1m_1m_1^T\right)+\left(\sum_{n\in C_2}x_nx_n^T-N_2m_2m_2^T\right)-(N_1m_1+N_2m_2)\frac{1}{N}(N_1m_1+N_2m_2)\right]w-N(m_1-m_2)}$$

$$=\left[S_W+N_1m_1m_1^T+N_2m_2m_2^T-\frac{(N_1m_1+N_2m_2)}{N}(N_1m_1+N_2m_2)^T\right]w-N(m_1-m_2)$$

$$=\left[S_W+\left(N_1-\frac{N_1^2}{N}\right)m_1m_1^T+\left(N_2-\frac{N_2^2}{N}\right)m_2m_2^T-\frac{N_1N_2}{N}(m_1m_2^T+m_2m_1^T)\right]w-N(m_1-m_2)$$

$$=\left[S_W+\frac{(N_1+N_2)N_1-N_1^2}{N}m_1m_1^T+\frac{(N_1+N_2)N_2-N_2^2}{N}m_2m_2^T-\frac{N_1N_2}{N}(m_1m_2^T+m_2m_1^T)\right]w-N(m_1-m_2)$$

$$=\left[S_W+\frac{N_1N_2}{N}(m_1m_1^T+m_2m_2^T-m_1m_2^T-m_2m_1^T)\right]w-N(m_1-m_2)$$

$$=\left(S_W+\frac{N_1N_2}{N}S_B\right)w-N(m_1-m_2)$$

$$=0$$

$$\Rightarrow\left(S_W+\frac{N_1N_2}{N}S_B\right)w=N(m_1-m_2)$$

1.3

由 $S_B=(m_2-m_1)(m_2-m_1)^T$，知 S_Bw 永远是与 (m_2-m_1) 同方向的

$\therefore\left(S_W+\frac{N_1N_2}{N}S_B\right)w=N(m_1-m_2)$ 可以写成：

$$S_Ww+C(m_1-m_2)=N(m_1-m_2)$$

$$\Rightarrow w=\frac{(N-C)(m_1-m_2)}{S_W}$$ 即 $w\propto S_W^{-1}(m_2-m_1)$

模式识别6 陈斯杰 2016310721

2.1 Decompose S_T

$$S_T = \sum_{k=1}^{K} \sum_{n \in C_k} (x_n - m)(x_n - m)^T$$

$$= \sum_{k=1}^{K} \sum_{n \in C_k} (x_n - m_k + m_k - m)(x_n^T - m_k^T + m_k^T - m^T)$$

$$= \sum_{k=1}^{K} \sum_{n \in C_k} \left[(x_n - m_k)(x_n - m_k)^T + (m_k - m)(m_k - m)^T \right]$$

$$= S_W + \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T$$

$$= S_W + S_B$$

2.2 在 projected D'-dimensional space 上，

$$S_W' = \sum_{k=1}^{K} \sum_{n \in C_k} (y_n - m_k')(y_n - m_k')^T$$

$$= \sum_{k=1}^{K} \sum_{n \in C_k} (W^T x_n - W^T m_k)(W^T x_n - W^T m_k)^T$$

$$= \sum_{k=1}^{K} \sum_{n \in C_k} W^T (x_n - m_k)(x_n - m_k)^T W$$

$$= W^T S_W W$$

$$S_B = \sum_{k=1}^{K} N_k (m_k' - m')(m_k' - m')^T$$

$$= W^T S_B W$$

2.3: $W = [w^{(1)}, w^{(2)}, \cdots, w^{(D')}]$

$$W^T A W = \begin{bmatrix} w^{(1)T} A w^{(1)} & & w^{(1)T} A w^{(D')} \\ & & \\ w^{(D')T} A w^{(1)} & & w^{(D')T} A w^{(D')} \end{bmatrix}$$

$$\prod \operatorname{diag} W^T A W = \prod_{i=1}^{D'} \left[w^{(i)T} A w^{(i)} \right]$$

$$J(W) = \frac{\prod_{i=1}^{D'} w^{(i)T} S_B w^{(i)}}{\prod_{i=1}^{D'} w^{(i)} S_W w^{(i)}}$$

模式识别 6　陈斯杰　201631072[1]

2.4
$$J(W)=\frac{\prod_{i=1}^{D'} w_i^T S_B w_i}{\prod_{i=1}^{D'} w_i^T S_w w_i}$$

优化 $J(W)$ 可以转化为：

$$\max \prod w_i^T S_B w_i$$
$$s.t. \prod w_i^T S_w w_i = c$$

引入 Lagrange 乘子转化为无约束问题：

$$L(w,\lambda)=\prod_{i=1}^{D'} w_i^T S_B w_i - \lambda\left(\prod_{i=1}^{D'} w_i^T S_w w_i - c\right)$$

$$\frac{\partial L}{\partial w}=0 \Rightarrow 2S_B w_i \frac{\prod_{i=1}^{D'} w_i^T S_B w_i}{w_i^T S_B w_i} = \lambda \cdot 2S_w w_i \frac{\prod_{i=1}^{D'} w_i^T S_w w_i}{w_i^T S_w w_i}$$

可以设：

$$S_B w_i = \lambda' S_w w_i$$

$$S_w^{-1} S_B w_i = \lambda' S_w^{-1} S_w w_i$$

$$(S_w^{-1} S_B) w_i = \lambda' w_i$$

∴ w_i 对应 $(S_w^{-1} S_B)$ 的特征向量

2.5　这是一个 Data Reduction 问题

$1 < D' \leq D$，　至多维持维数不变，仍为 D 维

模式识别 6　陈斯杰　2016310721

Problem 3

3.1　$P(e) = 1 - P(c) = 1 - \sum_{k=1}^{K} \int_{R_k} p(x, C_k)\, dx$

3.2　只看 Feature x_1 时，由于是两类问题，故有 $p(w_1) = p(w_2) = 0.5$

$e(x_1) = \frac{1}{2}(1 + \alpha_1 - \beta_1)$，同理，只看 x_2, x_3 时

$e(x_2) = \frac{1}{2}(1 + \alpha_2 - \beta_2)$

$e(x_3) = \frac{1}{2}(1 + \alpha_3 - \beta_3)$

又有 $\alpha_1 - \beta_1 < \alpha_2 - \beta_2 < \alpha_3 - \beta_3$　$\therefore e(x_1) < e(x_2) < e(x_3)$

3.3　代入得：

$e(x_1) = 0.1$　$e(x_2) = 0.125$　$e(x_3) = 0.155$

$e(x_1, x_2) = P(x_1=0, x_2=0 \mid w_2) + P(x_1=1, x_2=0 \mid w_1) + P(x_1=0, x_2=1 \mid w_1) + P(x_1=1, x_2=1 \mid w_1)$

$= (0.1 \times 0.2 + 0.1 \cdot 0.95 + 0.05 \cdot 0.9 + 0.1 \cdot 0.05) \times 0.5$

$= 0.0825$

$e(x_1, x_3) = (0.1 \times 0.3 + 0.1 \cdot 0.99 + 0.01 \times 0.9 + 0.1 \cdot 0.01) \times 0.5 = 0.0695$

$e(x_2, x_3) = (0.2 \cdot 0.3 + 0.5 \cdot 0.99 + 0.01 \cdot 0.95 + 0.05 \cdot 0.01) \times 0.5 = 0.05975$

选 x_2, x_3 错误率最低