HW2 Solution

18340013 陈琮昊

6.5

(a)若$N < D$，S_w不可逆。

(b)因为$S_B = \Sigma_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T$（其中$\mu_i$为为每类样本的均值，$\mu$为全部样本的均值，$N_i$为第$i$类样本所占个数）。由于$S_B$中的$rank(\mu_i - \mu) = 1$，根据秩的性质:

$rank(A) = rank(AA^T), rank(A + B) \le rank(A) + rank(B)$

可以知道$rank(S_B) \le C$。又因为存在线性组合使得$\Sigma_{i=1}^{C} N_i(\mu_i - \mu) = 0$，因此根据上述性质可以得到$rank(S_B) \le C - 1$。则广义特征向量的个数最多为$C - 1$。

(c)因为当$N > D(N >> D)$时，S_w通常是非奇异的，故很大可能上为可逆矩阵。

(d)

第三步里，不妨记对角化C得到的矩阵为D, 则$D = Q^T CQ(C = Q^{-T} DQ^{-1})$为对角矩阵。

$X^T S_B X = Q^T G^{-1} S_B G^{-T} Q = Q^T CQ = D$，因为$D$为对角矩阵，故$X^T S_B X$为对角矩阵。

$X^T S_W X = Q^T G^{-1} GG^T G^{-T} Q = Q^T Q = I$($Q$正交)。故$X^T S_W X$为单位阵。

由于$X^T S_B X(= D)$为对角矩阵，由对角矩阵性质可知该矩阵的特征值即为对角线上元素。

因此若证明广义特征值λ位于$X^T S_B X(= D)$的对角线上，只需证$Dt = \lambda t$.

(其中t为D的特征向量，该特征向量与本题的广义特征向量w不同)

将$C = G^{-1} S_B G^{-T}, S_W = GG^T$代入$S_B w = \lambda S_W w$得：

$GCG^T w = \lambda GG^T w$. 两边同时左乘G^{-1}：$CG^T w = \lambda G^T w$

将$D = Q^T CQ(C = Q^{-T} DQ^{-1})$代入得：$Q^{-T} DQ^{-1} G^T w = \lambda G^T w$

又因为$Q^T Q = I$(Q正交)，故上式化简得：$D(Q^{-1} G^T w) = \lambda(Q^{-1} G^T w)$

可以看到取$t = Q^{-1} G^T w$即可满足$Dt = \lambda t$形式，故证明了广义特征值λ位于$X^T S_B X$的对角线上。

根据前面的推导可以知道：$D(Q^{-1} G^T w) = \lambda(Q^{-1} G^T w)$, (即$Dt = \lambda t$)

由于D为对角矩阵(必对称)，那么必存在标准正交的特征向量系满足：

$$t_i^T t_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

回推可得$t_i^T t_j = (Q^{-1} G^T w_i)^T (Q^{-1} G^T w_j) = w_i^T S_W w_j$(根据$S_W = GG^T, QQ^T = I$)

即有：$w_i^T S_W w_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

所以在矩阵S_W下，广义特征向量w之间两两正交且具有单位范数。

最后再证明广义特征向量位于X的每一列上：

设广义特征向量组为$\{w_1, w_2, \ldots\ldots, w_n\}$。其中$w_i$为列向量，对应的特征值为$\lambda_i$.

显然有$S_B w_i = \lambda S_W w_i$. 若$X$的每一列为特征向量，不妨记$X = [w_1, w_2, \ldots\ldots, w_n]$.

而$w_i^T S_B w_j = \lambda w_i^T S_W w_j \begin{cases} 0 & i \neq j \\ \lambda_i & i = j \end{cases}$

有$X^T S_B X = diag(\lambda_1, \lambda_2, \ldots\ldots, \lambda_n) = D$符合前面所证。

即便X的列与列之间发生了顺序上的交换，得到的D也是对角阵，同样对角线上为广义特征值。

故广义特征向量位于X的每一列上。

7.4

(a)

$$\begin{aligned} &min \quad \frac{1}{2}w^Tw \\ &s.t. \quad y_iw^Tx_i - 1 \geq 0 \end{aligned}$$

(b)

首先得到拉格朗日函数：$L(w,\alpha) = \frac{1}{2}w^Tw - \Sigma_{i=1}^n\alpha_i(y_iw^Tx_i - 1)$

$s.t. \quad \alpha_i \geq 0, 1 \leq i \leq n.$

$\frac{\partial L}{\partial w} = w - \Sigma_{i=1}^n\alpha_iy_ix_i = 0$，可得$w = \Sigma_{i=1}^n\alpha_iy_ix_i$

故而$\frac{1}{2}w^Tw = \frac{1}{2}\Sigma_{i=1}^n\Sigma_{j=1}^n\alpha_i\alpha_jy_iy_jx_i^Tx_j$

将$w = \Sigma_{i=1}^n\alpha_iy_ix_i$代入得$L(w,\alpha) = -\frac{1}{2}w^Tw + \Sigma_{i=1}^n\alpha_i$

求$\underset{w}{min}L(w,\alpha)$对$\alpha$的极大，即得到对偶问题：

$\underset{\alpha}{max} \quad -\frac{1}{2}w^Tw + \Sigma_{i=1}^n\alpha_i = -\frac{1}{2}\Sigma_{i=1}^n\Sigma_{j=1}^n\alpha_i\alpha_jy_iy_jx_i^Tx_j + \Sigma_{i=1}^n\alpha_i$

$s.t. \quad 0 \leq \alpha_i \leq C, i = 1,2,3\ldots\ldots,n.$

加负号将求解极大转为求解极小，便得到了题目要求的对偶形式：

$\underset{\alpha}{min} \quad \frac{1}{2}\Sigma_{i=1}^n\Sigma_{j=1}^n\alpha_i\alpha_jy_iy_jx_i^Tx_j - \Sigma_{i=1}^n\alpha_i$

$s.t. \quad 0 \leq \alpha_i \leq C, i = 1,2,3\ldots\ldots,n.$

如果得到关于α的解为α^*，可以通过$w = \Sigma_{i=1}^n\alpha_iy_ix_i$来得到最优的$w^*$.

(c)

偏置项b对应$d+1$维数据的情况下，w的$d+1$维.

因为$w_d^* = \Sigma_{i=1}^n\alpha_i^*y_ix_i$，且在第$i+1$维的情况下$x_i = 1$,

所以$b = w_{d+1}^* = \Sigma_{i=1}^n\alpha_i^*y_i$

8.6

根据式(8.38)，可知$y^* = \underset{1\leq i\leq m}{argmax} \quad p(y=i|x;\theta)$

而$p(y=i|x;\theta) = \frac{p(x|y=i;\theta)p(y=i)}{p(x;\theta)}$

因为$p(x;\theta)$与y无关，而题目里又给定$p(x|y=i;\theta)$满足高斯分布，$p(y=1)=p(y=2)=0.5$

所以优化问题转化为：$y^* = \underset{1\leq i\leq m}{argmax} \quad p(x|y=i;\theta)$

故原题等价于$y^* = \begin{cases} 1 & p(x|y=1;\theta) > p(x|y=2;\theta) \\ 2 & p(x|y=1;\theta) \leq p(x|y=2;\theta) \end{cases}$

$p(x|y=i;\theta) = N(\mu, \Sigma)$,记$D$为$x$的维度，则$PDF$为：

$p(x|y=i) = \frac{1}{(2\pi)^{\frac{D}{2}}(\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)}$

因为Σ相同，所以只需比较 $-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$与 $-(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)$的大小关系：

$-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - [-(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)]$

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

因为$x^T\Sigma^{-1}(\mu_1 - \mu_2) = (\mu_1^T - \mu_2^T)(\Sigma^{-1})^T x$

故代入上式得：$-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - [-(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)]$

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

令$w^T = (\mu_1^T - \mu_2^T)(\Sigma^{-1} + (\Sigma^{-1})^T), b = \mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1$

故当$w^Tx + b > 0$时, $-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) > -(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)$, 即$p(x|y=1;\theta) > p(x|y=2;\theta)$;

故当$w^Tx + b \leq 0$时, $-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) \leq -(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)$, 即$p(x|y=1;\theta) \leq p(x|y=2;\theta)$.

即原题可以表示成如下形式：$y^* = \begin{cases} 1 & w^Tx + b > 0 \\ 2 & w^Tx + b \leq 0 \end{cases}$

其中$w = [(\mu_1^T - \mu_2^T)(\Sigma^{-1} + (\Sigma^{-1})^T)]^T = (\Sigma^{-1} + (\Sigma^{-1})^T)(\mu_1 - \mu_2), b = \mu_2^T\Sigma^{-1}\mu_2 - \mu_1^T\Sigma^{-1}\mu_1$

9.1

$||y_1 - y_2||_2^2 = (E_d^T(x_1 - \overline{x}) - E_d^T(x_2 - \overline{x}))^2 = (E_d^T(x_1 - x_2))^2 = (E_d^T(x_1 - x_2))^T E_d^T(x_1 - x_2)$

$= (x_1 - x_2)^T E_d E_d^T (x_1 - x_2)$

而由(9.15)式可知$d_A(x,y) = (x-y)^T A (x-y)$

即$d_A(x_1, x_2) = (x_1 - x_2)^T A (x_1 - x_2)$.

通过对比可以知道当$A = E_d E_d^T$时，

有$d_A(x_1, x_2) = ||y_1 - y_2||_2^2$成立，得证。

10.3

首先有该式成立：$CE(p,q) = h(p) + KL(p||q)$

而$KL(p||q) \geq 0$, 故$CE(p,q) \geq h(p)$成立

等号成立时当且仅当$KL(p||q) = 0$,

即对于任意的x，有$p(x) = q(x)$.