4.1

拉格朗日算子为

$$L(a,\lambda)=a^T\mathbf{B}a-\lambda(a^T\mathbf{W}a-1).$$

对a求偏导，结果为

$$\frac{\partial L(a,\lambda)}{\partial a}=2\mathbf{B}a+\lambda(2\mathbf{W}a)=0,$$

即

$$\mathbf{W}^{-1}\mathbf{B}a=\lambda a.$$

4.2

(1)

$$\begin{aligned}\log\frac{\Pr(G=1\mid X=x)}{\Pr(G=2\mid X=x)}&=\log\frac{f_1(x)}{f_2(x)}+\log\frac{\pi_1}{\pi_2}\\&=\log\frac{N_1}{N_2}-\frac{1}{2}(\mu_1+\mu_2)\Sigma^{-1}\left(\mu_1-\mu_2\right)\\&+x^T\Sigma^{-1}\left(\mu_1-\mu_2\right)\end{aligned}$$

若满足$x^T\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)>\frac{1}{2}(\hat{\mu}_2+\hat{\mu}_1)^T\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)-\log(N_2/N_1)$，有$\log\frac{\Pr(G=1|X=x)}{\Pr(G=2|X=x)}<0$，所以应该被分类至2，否则被分类至1.

（2）

$$\begin{aligned}&l\left(\beta_0,\beta\right)=\sum_{i=1}^N\left(y_i-\beta_0-x_i^T\beta\right)^2=\left(y-\mathbf{1}\beta_0-X\beta\right)^T\left(y-\mathbf{1}\beta_0-X\beta\right)\\&\tfrac{\partial l(\beta_0,\beta)}{\partial\beta}=-2X^T\left(y-\mathbf{1}\beta_0-X\beta\right)=0\\&\tfrac{\partial l(\beta_0,\beta)}{\partial\beta_0}=-2\mathbf{1}^T\left(y-\mathbf{1}\beta_0-X\beta\right)=0\end{aligned}$$

所以

$$\hat{\beta}_0=\left(\mathbf{1}^Ty-\mathbf{1}^TX\hat{\beta}\right)/N=-\bar{X}\hat{\beta}$$

$$(X^TX-N\bar{X}^T\bar{X})\beta=X^Ty=N(\mu_2-\mu_1)$$

另一方面

$$\begin{aligned}(N-2)\hat{\Sigma}&=\sum_{i:g_i=1}(x_i-\hat{\mu}_1)(x_i-\hat{\mu}_1)^T+\sum_{i:g_i=2}(x_i-\hat{\mu}_2)(x_i-\hat{\mu}_2)^T\\&=X^TX-N_1\mu_1\mu_1^T-N_2\mu_2\mu_2^T\\X^TX&=(N-2)\hat{\Sigma}+N_1\mu_1\mu_1^T+N_2\mu_2\mu_2^T\\\bar{X}^T\bar{X}&=\frac{1}{N^2}(N_1\mu_1+N_2\mu_2)(N_1\mu_1+N_2\mu_2)^T\end{aligned}$$

$$X^TX-N\bar{X}^T\bar{X}=(N-2)\hat{\Sigma}+N\hat{\Sigma}_B$$

所以

$$\left[(N-2)\hat{\Sigma}+N\hat{\Sigma}_B\right]\beta=N(\hat{\mu}_2-\hat{\mu}_1)$$

（3）

$$\begin{aligned}\hat{\Sigma}_B\beta&=\frac{N_1N_2}{N^2}(\hat{\mu}_2-\hat{\mu}_1)(\hat{\mu}_2-\hat{\mu}_1)^T\beta\\&=\left(\frac{N_1N_2}{N^2}(\hat{\mu}_2-\hat{\mu}_1)^T\beta\right)(\hat{\mu}_2-\hat{\mu}_1)\end{aligned}$$

其中$\frac{N_1N_2}{N^2}(\hat{\mu}_2-\hat{\mu}_1)^T\beta$是一个标量，所以$\hat{\Sigma}_B\beta$与$\hat{\mu}_2-\hat{\mu}_1$同向。根据上一问的结论，有

$$\hat{\beta}=\frac{1}{N-2}\left[N-\frac{N_1N_2}{N^2}(\hat{\mu}_2-\hat{\mu}_1)^T\beta\right]\Sigma^{-1}(\hat{\mu}_2-\hat{\mu}_1)$$

那么

$$\hat{\beta}\propto\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)$$

（4）不妨假设编码分别为t_1,t_2，于是$X^TX-N\bar{X}^T\bar{X}=X^Ty-\bar{X}^Ty=\frac{N_1N_2}{N}(t_2-t_1)(\hat{\mu}_2-\hat{\mu}_1)$，只要$t_1\neq t_2$上述结论依然成立。

（5）使用与题干相同的编码方式，即$-N/N_1,N/N_2$

$$\hat{\beta}_0=\left(\mathbf{1}^Ty-\mathbf{1}^TX\hat{\beta}\right)/N=-\bar{X}\hat{\beta}=-\frac{1}{N}(N_1\hat{\mu}_1^T+N_2\hat{\mu}_2^T)\hat{\beta}$$

所以

$$\hat{f}(x)=\hat{\beta}_0+x^T\hat{\beta}=\left[x^T-\frac{1}{N}(N_1\hat{\mu}_1^T+N_2\hat{\mu}_2^T)\right]\hat{\beta}$$

由第(3)问的结论，$\hat{\beta}\propto\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)$，那么

$$\begin{aligned}\hat{f}(x)&=\lambda\left[x^T-\frac{1}{N}(N_1\hat{\mu}_1^T+N_2\hat{\mu}_2^T)\right]\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)\\&=\lambda\left[x^T\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)-\frac{1}{N}(N_1\hat{\mu}_1^T+N_2\hat{\mu}_2^T)\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1)\right]\end{aligned}$$

显然，当且仅当$N_1=N_2$时，上式才与第(1)问中表达式等价。

4.3

该映射不改变每个分类的元素数量，故$\pi_k'=\pi_k$

由于该映射是线性变换，所以$\mu_k'=\hat{B}^T\mu_k$

由定义

$$\begin{aligned}\hat{\Sigma}'&=\sum_{k=1}^{K}\sum_{g_i=k}(\hat{B}^Tx_i-\hat{\mu}_k')(\hat{B}^Tx_i-\hat{\mu}_k')^T/(N-K)\\&=\sum_{k=1}^{K}\sum_{g_i=k}(\hat{B}^Tx_i-\hat{B}^T\hat{\mu}_k)(\hat{B}^Tx_i-\hat{B}^T\hat{\mu}_k)^T/(N-K)\\&=\hat{B}^T\left[\sum_{k=1}^{K}\sum_{g_i=k}(x_i-\hat{\mu}_k)(x_i-\hat{\mu}_k)^T/(N-K)\right]\hat{B}\\&=\hat{B}^T\hat{\Sigma}\hat{B}\end{aligned}$$

$$\begin{aligned}\delta_k'(x)&=(\hat{B}^Tx)^T(\hat{\Sigma}')^{-1}\hat{\mu}_k'-\frac{1}{2}(\hat{\mu}_k')^T(\hat{\Sigma}')^{-1}\hat{\mu}_k'+\log\pi_k'\\&=x^T\hat{B}\left(\hat{B}^T\hat{\Sigma}\hat{B}\right)^{-1}\hat{B}^T\hat{\mu}_k-\frac{1}{2}\hat{\mu}_k^T\hat{B}(\hat{B}^T\hat{\Sigma}\hat{B})^{-1}\hat{B}^T\hat{\mu}_k+\log\pi_k\\&=x^T\hat{\Sigma}^{-1}\hat{\mu}_k-\frac{1}{2}\hat{\mu}_k^T(\hat{\Sigma})^{-1}\hat{\mu}_k+\log\pi_k\end{aligned}$$

4.6

（1）由于有可分性，

$$\begin{aligned}\beta^T x_i^* &> 0 \text{ for } \ y_i = 1\\ \beta^T x_i^* &< 0 \text{ for } \ y_i = -1\end{aligned}$$

上式可以表示为$y_i\beta^T x_i^* > 0, \forall i$。由于$z_i = x_i^*/\|x_i^*\|$，$y_i\beta^T z_i > 0, \forall i$。于是，$\exists\alpha = \min_i y_i\beta^T z_i$，满足$y_i\beta^T z_i \geq \alpha$。

令$\beta_{sep} = \frac{1}{\alpha}\beta$，有$y_i\beta_{\text{sep}}^T z_i \geq 1 \ \forall i$。

（2）

$$\begin{aligned}\|\beta_{\text{new}} - \beta_{\text{sep}}\|^2 &= \|\beta_{\text{old}} - \beta_{\text{sep}} + y_i z_i\|^2\\ &= \|\beta_{\text{old}} - \beta_{\text{sep}}\|^2 + \|y_i z_i\|^2 + 2y_i(\beta_{\text{old}} - \beta_{\text{sep}})^T z_i\\ &= \|\beta_{\text{old}} - \beta_{\text{sep}}\|^2 + y_i^2\|z_i\|^2 + 2y_i\beta_{\text{old}}^T z_i - 2y_i\beta_{\text{sep}}^T z_i\\ &\leq \|\beta_{\text{old}} - \beta_{\text{sep}}\|^2 + 1 - 2\\ &= \|\beta_{\text{old}} - \beta_{\text{sep}}\|^2 - 1\end{aligned}$$