3.7

$w=\frac{\sum_{i=1}^{m}y_i(x_i-\bar{x})}{\sum_{i=1}^{m}x_i^2-\frac{1}{m}(\sum_{i=1}^{m}x_i)^2}$

[推导]：令式（3.5）等于0： $0=w\sum_{i=1}^{m}x_i^2-\sum_{i=1}^{m}(y_i-b)x_i$ $w\sum_{i=1}^{m}x_i^2=\sum_{i=1}^{m}y_ix_i-\sum_{i=1}^{m}bx_i$ 由于令式（3.6）等于0可得$b=\frac{1}{m}\sum_{i=1}^{m}(y_i-wx_i)$，又$\frac{1}{m}\sum_{i=1}^{m}y_i=\bar{y}$，$\frac{1}{m}\sum_{i=1}^{m}x_i=\bar{x}$，则$b=\bar{y}-w\bar{x}$，代入上式可得：

$$
\begin{aligned}
w\sum_{i=1}^{m}x_i^2&=\sum_{i=1}^{m}y_ix_i-\sum_{i=1}^{m}(\bar{y}-w\bar{x})x_i\\
w\sum_{i=1}^{m}x_i^2&=\sum_{i=1}^{m}y_ix_i-\bar{y}\sum_{i=1}^{m}x_i+w\bar{x}\sum_{i=1}^{m}x_i\\
w(\sum_{i=1}^{m}x_i^2-\bar{x}\sum_{i=1}^{m}x_i)&=\sum_{i=1}^{m}y_ix_i-\bar{y}\sum_{i=1}^{m}x_i\\
w&=\frac{\sum_{i=1}^{m}y_ix_i-\bar{y}\sum_{i=1}^{m}x_i}{\sum_{i=1}^{m}x_i^2-\bar{x}\sum_{i=1}^{m}x_i}
\end{aligned}
$$

又$\bar{y}\sum_{i=1}^{m}x_i=\frac{1}{m}\sum_{i=1}^{m}y_i\sum_{i=1}^{m}x_i=\bar{x}\sum_{i=1}^{m}y_i$，$\bar{x}\sum_{i=1}^{m}x_i=\frac{1}{m}\sum_{i=1}^{m}x_i\sum_{i=1}^{m}x_i=\frac{1}{m}(\sum_{i=1}^{m}x_i)^2$，代入上式即可得式（3.7）： $w=\frac{\sum_{i=1}^{m}y_i(x_i-\bar{x})}{\sum_{i=1}^{m}x_i^2-\frac{1}{m}(\sum_{i=1}^{m}x_i)^2}$

【注】：式（3.7）还可以进一步化简为能用向量表达的形式，将$\frac{1}{m}(\sum_{i=1}^{m}x_i)^2=\bar{x}\sum_{i=1}^{m}x_i$代入分母可得：

$$
\begin{aligned}
w&=\frac{\sum_{i=1}^{m}y_i(x_i-\bar{x})}{\sum_{i=1}^{m}x_i^2-\bar{x}\sum_{i=1}^{m}x_i}\\
&=\frac{\sum_{i=1}^{m}(y_ix_i-y_i\bar{x})}{\sum_{i=1}^{m}(x_i^2-x_i\bar{x})}
\end{aligned}
$$

又因为$\bar{y}\sum_{i=1}^{m}x_i=\bar{x}\sum_{i=1}^{m}y_i=\sum_{i=1}^{m}\bar{y}x_i=\sum_{i=1}^{m}\bar{x}y_i=m\bar{x}\bar{y}=\sum_{i=1}^{m}\bar{x}\bar{y}$，$\sum_{i=1}^{m}x_i\bar{x}=\bar{x}\sum_{i=1}^{m}x_i=\bar{x}\cdot m\cdot\frac{1}{m}\cdot\sum_{i=1}^{m}x_i=m\bar{x}^2=\sum_{i=1}^{m}\bar{x}^2$，则上式可化为：

$$
\begin{aligned}
w&=\frac{\sum_{i=1}^{m}(y_ix_i-y_i\bar{x}-x_i\bar{y}+\bar{x}\bar{y})}{\sum_{i=1}^{m}(x_i^2-x_i\bar{x}-x_i\bar{x}+\bar{x}^2)}\\
&=\frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{m}(x_i-\bar{x})^2}
\end{aligned}
$$

若令$\boldsymbol{x}=(x_1,x_2,\ldots,x_m)^T$，$\boldsymbol{x}_d$为去均值后的$\boldsymbol{x}$，$\boldsymbol{y}=(y_1,y_2,\ldots,y_m)^T$，$\boldsymbol{y}_d$为去均值后的$\boldsymbol{y}$，其中$\boldsymbol{x}$、$\boldsymbol{x}_d$、$\boldsymbol{y}$、$\boldsymbol{y}_d$均为m行1列的列向量，代入上式可得： $w=\frac{\boldsymbol{x}_d^T\boldsymbol{y}_d}{\boldsymbol{x}_d^T\boldsymbol{x}_d}$

3.10

$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}}=2\mathbf{X}^{T}(\mathbf{X}\hat{\boldsymbol{w}}-\boldsymbol{y})$

[推导]：将$E_{\hat{\boldsymbol{w}}}=(\boldsymbol{y}-\mathbf{X}\hat{\boldsymbol{w}})^{T}(\boldsymbol{y}-\mathbf{X}\hat{\boldsymbol{w}})$展开可得：$E_{\hat{\boldsymbol{w}}}=\boldsymbol{y}^{T}\boldsymbol{y}-\boldsymbol{y}^{T}\mathbf{X}\hat{\boldsymbol{w}}-\hat{\boldsymbol{w}}^{T}\mathbf{X}^{T}\boldsymbol{y}+\hat{\boldsymbol{w}}^{T}\mathbf{X}^{T}\mathbf{X}\hat{\boldsymbol{w}}$ 对$\hat{\boldsymbol{w}}$求导可得：
$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}}=\frac{\partial \boldsymbol{y}^{T}\boldsymbol{y}}{\partial \hat{\boldsymbol{w}}}-\frac{\partial \boldsymbol{y}^{T}\mathbf{X}\hat{\boldsymbol{w}}}{\partial \hat{\boldsymbol{w}}}-\frac{\partial \hat{\boldsymbol{w}}^{T}\mathbf{X}^{T}\boldsymbol{y}}{\partial \hat{\boldsymbol{w}}}+\frac{\partial \hat{\boldsymbol{w}}^{T}\mathbf{X}^{T}\mathbf{X}\hat{\boldsymbol{w}}}{\partial \hat{\boldsymbol{w}}}$ 由向量的求导公式可得：
$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}}=0-\mathbf{X}^{T}\boldsymbol{y}-\mathbf{X}^{T}\boldsymbol{y}+(\mathbf{X}^{T}\mathbf{X}+\mathbf{X}^{T}\mathbf{X})\hat{\boldsymbol{w}}$ $\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}}=2\mathbf{X}^{T}(\mathbf{X}\hat{\boldsymbol{w}}-\boldsymbol{y})$

3.27

$l(\beta)=\sum_{i=1}^{m}(-y_i\beta^{T}\hat{\boldsymbol{x}}_i+\ln(1+e^{\beta^{T}\hat{\boldsymbol{x}}_i}))$

[推导]：将式（3.26）代入式（3.25）可得：$l(\beta)=\sum_{i=1}^{m}\ln(y_i p_1(\hat{\boldsymbol{x}}_i;\beta)+(1-y_i)p_0(\hat{\boldsymbol{x}}_i;\beta))$ 其中
$p_1(\hat{\boldsymbol{x}}_i;\beta)=\frac{e^{\beta^{T}\hat{\boldsymbol{x}}_i}}{1+e^{\beta^{T}\hat{\boldsymbol{x}}_i}},p_0(\hat{\boldsymbol{x}}_i;\beta)=\frac{1}{1+e^{\beta^{T}\hat{\boldsymbol{x}}_i}}$，代入上式可得：

$$\begin{aligned}
l(\beta)&=\sum_{i=1}^{m}\ln\left(\frac{y_i e^{\beta^{T}\hat{\boldsymbol{x}}_i}+1-y_i}{1+e^{\beta^{T}\hat{\boldsymbol{x}}_i}}\right)\\
&=\sum_{i=1}^{m}\left(\ln(y_i e^{\beta^{T}\hat{\boldsymbol{x}}_i}+1-y_i)-\ln(1+e^{\beta^{T}\hat{\boldsymbol{x}}_i})\right)
\end{aligned}$$

由于y_i=0或1，则：

$l(\beta)=\begin{cases}\sum_{i=1}^{m}(-\ln(1+e^{\beta^{T}\hat{\boldsymbol{x}}_i})), & y_i=0\\ \sum_{i=1}^{m}(\beta^{T}\hat{\boldsymbol{x}}_i-\ln(1+e^{\beta^{T}\hat{\boldsymbol{x}}_i})), & y_i=1\end{cases}$ 两式综合可得：$l(\beta)=\sum_{i=1}^{m}\left(y_i\beta^{T}\hat{\boldsymbol{x}}_i-\ln(1+e^{\beta^{T}\hat{\boldsymbol{x}}_i})\right)$ 由于此式仍为极大似然估计的似然函数，所以最大化似然函数等价于最小化似然函数的相反数，也即在似然函数前添加负号即可得式（3.27）。

【注】：若式（3.26）中的似然项改写方式为$p(y_i|\boldsymbol{x}_i;\boldsymbol{w},b)=[p_1(\hat{\boldsymbol{x}}_i;\beta)]^{y_i}[p_0(\hat{\boldsymbol{x}}_i;\beta)]^{1-y_i}$，再将其代入式（3.25）可得：$l(\beta)=\sum_{i=1}^{m}\left(y_i\ln(p_1(\hat{\boldsymbol{x}}_i;\beta))+(1-y_i)\ln(p_0(\hat{\boldsymbol{x}}_i;\beta))\right)$ 此式显然更易推导出式（3.27）

3.30

$\frac{\partial l(\beta)}{\partial \beta}=-\sum_{i=1}^{m}\hat{\boldsymbol{x}}_i(y_i-p_1(\hat{\boldsymbol{x}}_i;\beta))$

[解析]：此式可以进行向量化，令$p_1(\hat{\boldsymbol{x}}_i;\beta)=\hat{y}_i$，代入上式得：

$$\begin{aligned}
\frac{\partial l(\beta)}{\partial \beta}&=-\sum_{i=1}^{m}\hat{\boldsymbol{x}}_i(y_i-\hat{y}_i)\\
&=\sum_{i=1}^{m}\hat{\boldsymbol{x}}_i(\hat{y}_i-y_i)\\
&=\boldsymbol{X}^{T}(\hat{\boldsymbol{y}}-\boldsymbol{y})\\
&=\boldsymbol{X}^{T}(p_1(\boldsymbol{X};\beta)-\boldsymbol{y})
\end{aligned}$$

3.32

$J=\frac{\boldsymbol{w}^{T}(\mu_0-\mu_1)(\mu_0-\mu_1)^{T}\boldsymbol{w}}{\boldsymbol{w}^{T}(\Sigma_0+\Sigma_1)\boldsymbol{w}}$

$$
\begin{aligned}
J &= \frac{\left|\left|\boldsymbol{w}^T\mu_0 - \boldsymbol{w}^T\mu_1\right|\right|_2^2}{\boldsymbol{w}^T(\Sigma_0+\Sigma_1)\boldsymbol{w}} \\
&= \frac{\left|\left|(\boldsymbol{w}^T\mu_0 - \boldsymbol{w}^T\mu_1)^T\right|\right|_2^2}{\boldsymbol{w}^T(\Sigma_0+\Sigma_1)\boldsymbol{w}} \\
&= \frac{\left|\left|(\mu_0 - \mu_1)^T\boldsymbol{w}\right|\right|_2^2}{\boldsymbol{w}^T(\Sigma_0+\Sigma_1)\boldsymbol{w}} \\
&= \frac{[(\mu_0 - \mu_1)^T\boldsymbol{w}]^T(\mu_0 - \mu_1)^T\boldsymbol{w}}{\boldsymbol{w}^T(\Sigma_0+\Sigma_1)\boldsymbol{w}} \\
&= \frac{\boldsymbol{w}^T(\mu_0 - \mu_1)(\mu_0 - \mu_1)^T\boldsymbol{w}}{\boldsymbol{w}^T(\Sigma_0+\Sigma_1)\boldsymbol{w}}
\end{aligned}
$$

[推导]：

3.37

$\boldsymbol{S}_b\boldsymbol{w} = \lambda\boldsymbol{S}_w\boldsymbol{w}$

[推导]：由3.36可列拉格朗日函数： $l(\boldsymbol{w}) = -\boldsymbol{w}^T\boldsymbol{S}_b\boldsymbol{w} + \lambda(\boldsymbol{w}^T\boldsymbol{S}_w\boldsymbol{w} - 1)$ 对$\boldsymbol{w}$求偏导可得：

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{w})}{\partial \boldsymbol{w}} &= -\frac{\partial(\boldsymbol{w}^T\boldsymbol{S}_b\boldsymbol{w})}{\partial \boldsymbol{w}} + \lambda\frac{(\boldsymbol{w}^T\boldsymbol{S}_w\boldsymbol{w} - 1)}{\partial \boldsymbol{w}} \\
&= -(\boldsymbol{S}_b + \boldsymbol{S}_b^T)\boldsymbol{w} + \lambda(\boldsymbol{S}_w + \boldsymbol{S}_w^T)\boldsymbol{w}
\end{aligned}
$$

又$\boldsymbol{S}_b = \boldsymbol{S}_b^T, \boldsymbol{S}_w = \boldsymbol{S}_w^T$，则： $\frac{\partial l(\boldsymbol{w})}{\partial \boldsymbol{w}} = -2\boldsymbol{S}_b\boldsymbol{w} + 2\lambda\boldsymbol{S}_w\boldsymbol{w}$ 令导函数等于0即可得式3.37。

3.43

$$
\begin{aligned}
\boldsymbol{S}_b &= \boldsymbol{S}_t - \boldsymbol{S}_w \\
&= \sum_{i=1}^{N} m_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T
\end{aligned}
$$

[推导]：由式3.40、3.41、3.42可得：

$$
\begin{aligned}
\boldsymbol{S}_b &= \boldsymbol{S}_t - \boldsymbol{S}_w \\
&= \sum_{i=1}^{m}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T - \sum_{i=1}^{N}\sum_{\boldsymbol{x}\in X_i}(\boldsymbol{x} - \boldsymbol{\mu}_i)(\boldsymbol{x} - \boldsymbol{\mu}_i)^T \\
&= \sum_{i=1}^{N}\left(\sum_{\boldsymbol{x}\in X_i}\left((\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T - (\boldsymbol{x} - \boldsymbol{\mu}_i)(\boldsymbol{x} - \boldsymbol{\mu}_i)^T\right)\right) \\
&= \sum_{i=1}^{N}\left(\sum_{\boldsymbol{x}\in X_i}\left((\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x}^T - \boldsymbol{\mu}^T) - (\boldsymbol{x} - \boldsymbol{\mu}_i)(\boldsymbol{x}^T - \boldsymbol{\mu}_i^T)\right)\right) \\
&= \sum_{i=1}^{N}\left(\sum_{\boldsymbol{x}\in X_i}\left(\boldsymbol{x}\boldsymbol{x}^T - \boldsymbol{x}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{x}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{x}\boldsymbol{x}^T + \boldsymbol{x}\boldsymbol{\mu}_i^T + \boldsymbol{\mu}_i\boldsymbol{x}^T - \boldsymbol{\mu}_i\boldsymbol{\mu}_i^T\right)\right) \\
&= \sum_{i=1}^{N}\left(\sum_{\boldsymbol{x}\in X_i}\left(-\boldsymbol{x}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{x}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{x}\boldsymbol{\mu}_i^T + \boldsymbol{\mu}_i\boldsymbol{x}^T - \boldsymbol{\mu}_i\boldsymbol{\mu}_i^T\right)\right) \\
&= \sum_{i=1}^{N}\left(-\sum_{\boldsymbol{x}\in X_i}\boldsymbol{x}\boldsymbol{\mu}^T - \sum_{\boldsymbol{x}\in X_i}\boldsymbol{\mu}\boldsymbol{x}^T + \sum_{\boldsymbol{x}\in X_i}\boldsymbol{\mu}\boldsymbol{\mu}^T + \sum_{\boldsymbol{x}\in X_i}\boldsymbol{x}\boldsymbol{\mu}_i^T + \sum_{\boldsymbol{x}\in X_i}\boldsymbol{\mu}_i\boldsymbol{x}^T - \sum_{\boldsymbol{x}\in X_i}\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T\right) \\
&= \sum_{i=1}^{N}\left(-m_i\boldsymbol{\mu}_i\boldsymbol{\mu}^T - m_i\boldsymbol{\mu}\boldsymbol{\mu}_i^T + m_i\boldsymbol{\mu}\boldsymbol{\mu}^T + m_i\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T + m_i\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T - m_i\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T\right) \\
&= \sum_{i=1}^{N}\left(-m_i\boldsymbol{\mu}_i\boldsymbol{\mu}^T - m_i\boldsymbol{\mu}\boldsymbol{\mu}_i^T + m_i\boldsymbol{\mu}\boldsymbol{\mu}^T + m_i\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T\right) \\
&= \sum_{i=1}^{N} m_i\left(-\boldsymbol{\mu}_i\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}_i^T + \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}_i\boldsymbol{\mu}_i^T\right) \\
&= \sum_{i=1}^{N} m_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T
\end{aligned}
$$

3.44

$$\max_{\mathbf{W}}\frac{tr(\mathbf{W}^T\boldsymbol{S}_b\mathbf{W})}{tr(\mathbf{W}^T\boldsymbol{S}_w\mathbf{W})}$$

[解析]：此式是式3.35的推广形式，证明如下： 设$\mathbf{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_i,\ldots,\boldsymbol{w}_{N-1}]$，其中$\boldsymbol{w}_i$为$d$行1列的列向量，则：

$$
\begin{cases}
tr(\mathbf{W}^T\boldsymbol{S}_b\mathbf{W}) = \sum\limits_{i=1}^{N-1}\boldsymbol{w}_i^T\boldsymbol{S}_b\boldsymbol{w}_i \\
tr(\mathbf{W}^T\boldsymbol{S}_w\mathbf{W}) = \sum\limits_{i=1}^{N-1}\boldsymbol{w}_i^T\boldsymbol{S}_w\boldsymbol{w}_i
\end{cases}
$$

所以式3.44可变形为： $\max\limits_{\mathbf{W}}\frac{\sum_{i=1}^{N-1}\boldsymbol{w}_i^T\boldsymbol{S}_b\boldsymbol{w}_i}{\sum_{i=1}^{N-1}\boldsymbol{w}_i^T\boldsymbol{S}_w\boldsymbol{w}_i}$ 对比式3.35易知上式即为式3.35的推广形式。