第三章（上）

3.1

A，错误。方差一致就行。
B，正确。
C，错误。线性回归常用来预测连续的独立变量。
D，错误。违背基本假设的计量经济学模型还是可以估计的，只是不能使用普通最小二乘法进行估计。

3.2

C
最小二乘法的解析解为：

$$W=(X^TX)^{-1}X^TY$$
$$\hat{Y}=XW$$

根据伪逆可知：

$$X^{\dagger}=(X^TX)^{-1}X^T$$

从而：

$$\hat{Y}=XW=XX^{\dagger}Y$$

故选C。

3.3

A、D不相关，C负相关，故选B。

3.4

（a）由于$x(t)$被$w(t)$误分类，因此

$y(t)\neq sign(w^T(t)x(t))$
当$w^T(t)x(t)>0$，即$sign(w^T(t)x(t))=1$时，$y(t)=-1$
当$w^T(t)x(t)<0$，即$sign(w^T(t)x(t))=-1$时，$y(t)=1$
因此$y(t)w^T(t)x(t)<0$
（b）$y(t)w^T(t+1)x(t)=y(t)(w^T(t)+y(t)x^T(t))x(t)=y(t)w^T(t)x(t)+y^2(t)x^T(t)x(t)$
由于$x(t)$的第一个分量为1，因此$x^T(t)x(t)\geqslant 1$，并且$y^2(t)=1$，所以$y^2(t)x^T(t)x(t)\geqslant 1$
故$y(t)w^T(t+1)x(t)>y(t)w^T(t)x(t)$
（c）由（a）可知，当分类错误时，$y(t)w^T(t)x(t)<0$，而由（b）可知每次更新后
$y(t)w^T(t+1)x(t)>y(t)w^T(t)x(t)$，
yw^Tx朝着正方向前进，因此若数据集是线性可分的，必然经过有限次更新后，使得$yw^Tx>0$，因此前进的方向是正确的。

3.5

当使用梯度下降法更新权重：

$\frac{\partial \frac{1}{N}\sum(w_1x_1+w_2x_2+b-y_i)^2}{\partial w_1}=\frac{1}{N}x_1*\sum 2(f(x_i)-y_i)$

$\frac{\partial \frac{1}{N}\sum(w_1x_1+w_2x_2+b-y_i)^2}{\partial b}=\frac{1}{N}*\sum 2(f(x_i)-y_i)$

$w_1=w_1-\frac{1}{N}x_1*\sum 2(f(x_i)-y_i)$

$b = b - \frac{1}{N} * \sum 2(f(x_i) - y_i)$

如果是最小二乘法更新:

分别对w和b求偏导数，使偏导数为0，即有:

$w_1 = \frac{\sum y_i x_1 - (x_1 x_2 w_2 + b x_1)}{\sum x_1^2}$

$b = \frac{1}{N} \sum y_i - w_1 x_1 - w_2 x_2$

3.6

初始化$w_0 = \binom{1}{1}, b_0 = 1$
$\hat{y_1} = sign(7) = 1 = y_1$
$\hat{y_2} = sign(7) = 1 = y_2$
$\hat{y_3} = sign(2) = 1 \neq y_3$
更新: $w_1 = \binom{1}{1} - \binom{0}{0.5} = \binom{1}{0.5}$
$b_1 = 1 - 0.5 = 0.5$
$\hat{y_3} = sign(1) = 1 \neq y_3$
更新: $w_2 = \binom{1}{0.5} - \binom{0}{0.5} = \binom{1}{0}$
$b_2 = 0.5 - 0.5 = 0$
$\hat{y_3} = sign(0) = 0 \neq y_3$
更新: $w_3 = \binom{1}{0} - \binom{0}{0.5} = \binom{1}{-0.5}$
$b_3 = 0 - 0.5 = -0.5$
$\hat{y_1} = sign(-0.5) = -1 \neq y_1$
更新: $w_4 = \binom{1}{-0.5} + \binom{1}{2} = \binom{2}{1.5}$
$b_4 = -0.5 + 0.5 = 0$
$\hat{y_3} = sign(1.5) = 1 \neq y_3$
更新: $w_5 = \binom{2}{1.5} - \binom{0}{0.5} = \binom{2}{1}$
$b_5 = 0 - 0.5 = -0.5$
$\hat{y_3} = sign(0.5) = 1 \neq y_3$
更新: $w_6 = \binom{2}{1} - \binom{0}{0.5} = \binom{2}{0.5}$
$b_6 = -0.5 - 0.5 = -1$
对于所有点成立，因此$w = \binom{2}{0.5}, b = -1$

第三章（下）

3.7

由题意可知: $p(y=1|x,w,b) = \frac{1}{1+exp(-(w^T x+b))}$，$p(y=-1|x,w) = \frac{1}{1+exp(w^T x+b)}$
则易得: $p(y|x,w,b) = \frac{1}{1+exp(-y(w^T x+b))}$
则对数似然函数
$L(w,b) = log(\prod_{i=1}^N \frac{1}{1+exp(-y_i(w^T x_i+b))}) = \sum_{i=1}^N log(\frac{1}{1+exp(-y_i(w^T x_i+b))}) = -\sum_{i=1}^N log(1+exp(-y_i(w^T x_i+b)))$

3.8

由题意得：$\theta(x) = \frac{1}{1+e^x}$
则当分类正确时，$y(w^Tx) > 0$；反之，$y(w^Tx) < 0$
故对于$x > 0$，有$\theta(x) < \theta(-x)$
因此，当分类正确时$|-y_nx_n\theta(y_nw^Tx_n)|$比分类错误时更小，得证。

3.9

假设最终$\hat{w} = (1, 1, -1)$
由题意可知，$R = max\ ||\hat{x}_i|| = \sqrt{(-3)^2 + (-9)^2 + 1} = \sqrt{91}$
$\gamma = \frac{1}{\sqrt{3}}(1+3-1) = \sqrt{3}$
因此，$k \leq (\frac{R}{\gamma})^2 = \frac{91}{3} = 31$