Problem Set 1

1.先证：当矩阵A的列向量组线性无关，则矩阵A^TA可逆。
设$A^TAX=0$，如果A^TA可逆则方程有唯一解$X=0$，
原命题等价于证明当矩阵A的列向量组线性无关，则$A^TAX=0$有唯一解$X=0$，
有$X^TA^TAX=0$，变换得$(AX)^TAX=0$，$AX=0$，
设$A=[a_1,a_2,...,a_n]$，$X=[x_1,x_2,...,x_n]^T$，
有$a_1x_1+a_2x_2+...+a_nx_n=0$（1），
因为A的列向量组线性无关，所以方程有唯一解$[x_1,x_2,...,x_n]=[0,0,...,0]$，即$X=0$，原命题得证。

若矩阵A的列向量组不满足线性无关的条件，观察式（1），不保证有唯一解$X=0$，因此矩阵A^TA可逆当且仅当矩阵A的列向量组线性无关。

2.(a)根据Hessian矩阵定义：

$$H_{i,j}=\frac{\partial^2 J(\theta)}{\partial\theta_i\partial\theta_j}=\frac{\partial}{\partial\theta_j}(\frac{\partial J(\theta)}{\partial\theta_i})=\frac{\partial}{\partial\theta_j}(\sum_{k=1}^{m}(\theta^Tx^{(k)}-y^{(k)})x_i^{(k)})=\sum_{k=1}^{m}x_i^{(k)}x_j^{(k)}$$

(b)设$X=[x^{(1)}x^{(2)}...x^{(m)}]^T$，$Y=[y^{(1)}y^{(2)}...y^{(m)}]$，
可得$H=X^TX$，
对于$\bigtriangledown_\theta J(\theta)$，根据梯度的含义：
$\bigtriangledown_\theta J(\theta)=[\frac{\partial J(\theta)}{\partial\theta_1}\frac{\partial J(\theta)}{\partial\theta_2}...\frac{\partial J(\theta)}{\partial\theta_n}]^T$
$\bigtriangledown_\theta J(\theta)_i=\sum_{k=1}^{m}(\theta^Tx^{(k)}-y^{(k)})x_i^{(k)}=\sum_{k=1}^{m}\theta^Tx^{(k)}x_i^{(k)}-\sum_{k=1}^{m}y^{(k)}x_i^{(k)}$
$=(X^TX\theta-X^T\overrightarrow{y})_i$
因此$\bigtriangledown_\theta J(\theta)=X^TX\theta-X^T\overrightarrow{y}$，
根据牛顿法迭代公式$\theta:=\theta-H^{-1}\bigtriangledown_\theta J(\theta)$，
第一轮迭代后$\theta^*=\theta-(X^TX)^{-1}(X^TX\theta-X^T\overrightarrow{y})=(X^TX)^{-1}X^T\overrightarrow{y}$.

3.(a)根据定理，最小二乘的唯一解为 $\theta=(X^TX)^{-1}X^TY$，
根据题意设
$$X=\begin{bmatrix}1 & 2005\\1 & 2006\\1 & 2007\\1 & 2008\\1 & 2009\end{bmatrix},\ Y=\begin{bmatrix}12\\19\\29\\37\\45\end{bmatrix},$$
代入计算得 $\theta=[-1.68e+04;8.40]$，

即最小二乘回归直线为$y = 8.40a - 1.68e4$.

(b)预测公司2012年的销售额为70.4百万美元。

4.根据Hessian矩阵定义：

$\frac{\partial J(\theta)}{\partial \theta_i} = -\frac{1}{m}\sum_{k=1}^{m}(1 - g(\theta^T y^{(k)} x^{(k)}))y^{(k)} x_i^{(k)}$

$\frac{\partial J(\theta)}{\partial \theta_i \partial \theta_j} = \frac{1}{m}\sum_{k=1}^{m} g(\theta^T y^{(k)} x^{(k)})(1 - g(\theta^T y^{(k)} x^{(k)}))(y^{(k)})^2 x_i^{(k)} x_j^{(k)} = H_{i,j}$

注意到$y^{(k)} \in \{-1, 1\}$，$(y^{(k)})^2 = 1$，

以及$h_\theta(x) = g(\theta^T x)$，$h_\theta(x) = h_\theta(-x)$，

$H_{i,j} = \frac{1}{m}\sum_{k=1}^{m} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))x_i^{(k)} x_j^{(k)}$，

进一步可得$H = \frac{1}{m}\sum_{k=1}^{m} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))x^{(k)} x^{(k)T}$，

$z^T H z = \frac{1}{m}\sum_{k=1}^{m} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))z^T x^{(k)} x^{(k)T} z$

$= \frac{1}{m}\sum_{k=1}^{m} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))(z^T x^{(k)})^2 \geq 0$，命题得证。