Problem Set 1

1 Conditions for Normal Equation

> Prove the following theorem: The matrix A^TA is invertible if and only if the columns of A are linearly independent.

该题即证明：A^TA可逆 $\Leftrightarrow$ A线性无关。

充分性：

已知A^TA可逆，若A线性无关，则有：$AX=0$仅有唯一解$X=0$。

所以要证明充分性，即证明在A^TA可逆的条件下，$AX=0$有唯一解$X=0$即可。

若$AX=0$，则有：$A^TAX=0$，由于A^TA可逆，故该等式有唯一解$X=0$。

即：$AX=0$有唯一解$X=0$。

故充分性得证。

必要性：

已知A线性无关，若A^TA可逆，则有：$A^TAX=0$有唯一解$X=0$。

所以要证明必要性，即证明在A线性无关的条件下，$A^TAX=0$有唯一解$X=0$即可。

若$A^TAX=0$，则有：$X^TA^TAX=0$，即：$(AX)^TAX=0$，故$AX=0$。

由于A线性无关，故$X=0$。

即$A^TAX=0$有唯一解$X=0$。

故必要性得证。

或者也可由可逆矩阵的相关性质简单证明：

若A^TA可逆，则A^T、A均可逆，故A线性无关。

若A线性无关，则A、A^T均可逆，故A^TA可逆。

2 Newton's Method for Computing Least Squares

> In this problem, we will prove that if we use Newton's method solve the least squares optimization problem, then we only need one iteration to converge to the optimal parameter θ^*

> (a) Find the Hessian of the cost function $J(\theta)=\frac{1}{2}\sum_{i=1}^m(\theta^Tx^{(i)}-y^{(i)})^2$

首先对$J(\theta)$求θ_j的偏导：

$$\frac{\partial J(\theta)}{\partial \theta_j}=\sum_{i=1}^m(\theta^Tx^{(i)}-y^{(i)})x_j^{(i)}$$

再求得θ_k的偏导：

$$\frac{\partial^2 J(\theta)}{\partial\theta_j\partial\theta_k} = \sum_{i=1}^{m}\frac{\partial}{\partial\theta_k}(\theta^T x^{(i)} - y^{(i)})x_j^{(i)} = \sum_{i=1}^{m} x_j^{(i)} x_k^{(i)} = (X^T X)_{jk}$$

因此，$J(\theta)$的Hessian即为$H = X^T X$。

> (b) Show that the first iteration of Newton's method gives us $\theta^* = (X^T X)^{-1} X^T \vec{y}$, the solution to our least square problem. ($\vec{y}$ denotes the vector of the features.)

对于给定的$\theta^{(0)}$，由牛顿法：

$$\begin{aligned}\theta^{(1)} &= \theta^{(0)} - H^{-1}\nabla_\theta J(\theta^{(0)}) \\ &= \theta^{(0)} - (X^T X)^{-1}(X^T X\theta^{(0)} - X^T\vec{y}) \\ &= \theta^{(0)} - \theta^{(0)} + (X^T X)^{-1} X^T \vec{y} \\ &= (X^T X)^{-1} X^T \vec{y}\end{aligned}$$

3 Prediction using Linear Regression

> The sales of a company (in million dollars) for each year are shown in the table below.
>
> | x (year) | 2005 | 2006 | 2007 | 2008 | 2009 |
> |---|---|---|---|---|---|
> | y (sales) | 12 | 19 | 29 | 37 | 45 |

> (a) Find the least square regression line $y = ax + b$

对题中数据进行处理，得到：

x (year)	0	1	2	3	4
y (sales)	12	19	29	37	45

X、Y：

$$X = \begin{bmatrix}1 & 0\\1 & 1\\1 & 2\\1 & 3\\1 & 4\end{bmatrix} Y = \begin{bmatrix}12\\19\\29\\37\\45\end{bmatrix}$$

由least square公式$\theta = (X^T X)^{-1} X^T Y$计算$\theta$:

$$(X^T X)^{-1} = (\begin{bmatrix}1 & 1 & 1 & 1 & 1\\0 & 1 & 2 & 3 & 4\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\\1 & 2\\1 & 3\\1 & 4\end{bmatrix})^{-1} = (\begin{bmatrix}5 & 10\\10 & 30\end{bmatrix})^{-1} = \begin{bmatrix}\frac{3}{5} & -\frac{1}{5}\\-\frac{1}{5} & \frac{1}{10}\end{bmatrix}$$

$$X^TY=\begin{bmatrix}1&1&1&1&1\\0&1&2&3&4\end{bmatrix}\begin{bmatrix}12\\19\\29\\37\\45\end{bmatrix}=\begin{bmatrix}142\\368\end{bmatrix}$$

$$\theta=(X^TX)^{-1}X^TY=\begin{bmatrix}\frac{3}{5}&-\frac{1}{5}\\-\frac{1}{5}&\frac{1}{10}\end{bmatrix}\begin{bmatrix}142\\368\end{bmatrix}=\begin{bmatrix}\frac{58}{5}\\\frac{42}{5}\end{bmatrix}$$

即：least square regression line 为：$y=\frac{42}{5}x+\frac{58}{5}$。

> (b) Use the least squares regression line as a model to estimate the sales of the company in 2012.

当year为2012时，x应为：$2012-2005=7$：

将$x=7$带入least square regression line：

$$y=\frac{42}{5}\times 7+\frac{58}{5}=\frac{352}{5}\approx 70.4$$

4 Logistic Regression

> Consider the average empirical loss for logistic regression:
>
> $$J(\theta)=\frac{1}{m}\sum_{i=1}^{m}log(1+e^{-y^{(i)}\theta^Tx^{(i)}})=-\frac{1}{m}\sum_{i=1}^{m}log(h_\theta(y^{(i)}x^{(i)}))$$
>
> where $y^{(i)}\in\{-1,1\}$ $h_\theta(x)=g(\theta^Tx)$ and $g(z)=1/(1+e^{-z})$. Find the Hessian H of this function, and show that for any vector z, it holds true that
>
> $$z^THz\geq 0$$
>
> Hint: You might want to start by showing the fact that $\sum_i\sum_j z_ix_ix_jz_j=\ (x^Tz)^2\geq 0$.

对$J(\theta)$求θ_j的偏导：

$$\begin{aligned}\frac{\partial J(\theta)}{\partial\theta_j}&=-\frac{1}{m}\sum_{i=1}^{m}\frac{1}{h_\theta(y^{(i)}x^{(i)})}\cdot\frac{\partial h_\theta(y^{(i)}x^{(i)})}{\partial\theta_j}\\&=-\frac{1}{m}\sum_{i=1}^{m}\frac{1}{h_\theta(y^{(i)}x^{(i)})}\cdot\frac{exp(-y^{(i)}\theta^Tx^{(i)})y^{(i)}x_j^{(i)}}{(1+exp(-y^{(i)}\theta^Tx^{(i)}))^2}\\&=-\frac{1}{m}\sum_{i=1}^{m}\frac{1}{h_\theta(y^{(i)}x^{(i)})}\cdot h_\theta^2(y^{(i)}x^{(i)})\cdot\frac{1-h_\theta(y^{(i)}x^{(i)})}{h_\theta(y^{(i)}x^{(i)})}\cdot y^{(i)}x_j^{(i)}\\&=-\frac{1}{m}\sum_{i=1}^{m}(1-h_\theta(y^{(i)}x_j^{(i)}))y^{(i)}x_j^{(i)}\end{aligned}$$

再求对θ_k的偏导：

$$
\begin{aligned}
\frac{\partial^2 J(\theta)}{\partial\theta_j \partial\theta_k} &= -\frac{1}{m}\sum_{i=1}^{m}(-\frac{exp(-y^{(i)}\theta^T x^{(i)})y^{(i)}x_k^{(i)}}{(1+exp(-y^{(i)}\theta^T x^{(i)}))^2})y^{(i)}x_j^{(i)} \\
&= \frac{1}{m}\sum_{i=1}^{m} h_\theta^2(y^{(i)}x^{(i)}) \cdot \frac{1-h(y^{(i)}x^{(i)})}{h_\theta(y^{(i)}x^{(i)})}(y^{(i)})^2 x_j^{(i)} x_k^{(i)} \\
&= \frac{1}{m}\sum_{i=1}^{m} h_\theta(y^{(i)}x^{(i)})(1-h_\theta(y^{(i)}x^{(i)}))(y^{(i)})^2 x_j^{(i)} x_k^{(i)}
\end{aligned}
$$

得到：$H = (\frac{1}{m}\sum_{i=1}^{m} h_\theta(y^{(i)}x^{(i)})(1-h_\theta(y^{(i)}x^{(i)})))(y^{(i)})^2 X^T X$

由于$h(x) \in (0,1)$，有：$h_\theta(y^{(i)}x^{(i)}) > 0$、$(1-h_\theta(y^{(i)}x^{(i)})) > 0$，且$(y^{(i)})^2 > 0$，因此H的系数均为正数，Hessian矩阵中各项的正负性取决于X。

对于任意向量z，有：

$$
z^T H z = (\frac{1}{m}\sum_{i=1}^{m} h_\theta(y^{(i)}x^{(i)})(1-h_\theta(y^{(i)}x^{(i)})))(y^{(i)})^2 z^T X^T X z
$$

要判断上式的正负性，即判断$z^T X^T X z$的正负性，而对于该矩阵中的每一项，有：

$$
\sum_i \sum_j z_i x_i x_j z_j = (x^T z)^2 \geq 0
$$

故$z^T X^T X z \geq 0$，原题得证。