# Problem Set 1

## 1 Conditions for Normal Equation

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

## 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 $\theta^*$

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

(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.)

## 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 $y = ax + b$.

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

## 4 Logistic Regression

Consider the average empirical loss for logistic regression:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \log\left(1 + e^{-y^{(i)}\theta^T x^{(i)}}\right) = -\frac{1}{m}\sum_{i=1}^{m} \log\left(h_\theta(y^{(i)}x^{(i)})\right)$$

where $y^{(i)} \in \{-1, 1\}$ $h_\theta(x) = g(\theta^T x)$ 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

$$z^T H z \geq 0$$

*Hint: You might want to start by showing the fact that* $\sum_i \sum_j z_i x_i x_j z_j = (x^T z)^2 \geq 0$.