Problem Set 1

林小斌

2019年3月28日

1 Conditions for Normal Equation

If the matrix A^TA is invertible，matrix $(A^TA)^{-1}$ satisfy the equation

$$A^TA \cdot (A^TA)^{-1} = (A^TA)^{-1} \cdot A^TA = I$$

where A is $n \times m$ matirx， A^TA is $m \times m$ matrix.，I is unit matrix. According to the characteristics of the invertible matrix，The equation $A^TAx = 0$ has only one normal solution.

$$A^TAx = 0 \Longrightarrow x^TA^TAx = 0 \Longrightarrow (Ax)^TAx = 0 \Longrightarrow \|Ax\|^2 = 0$$

Then $Ax = 0$，where $x = 0$. So the columns of A are linearly independent.

if the columns of A are linearly independent，then $Ax = 0$ has only one normal solution $x = 0$. So the null space of $A^TAx = \{0\}$. Since A^TA is a square matrix, this means A^TA is invertible.

2 Newton's Method for Computing Least Squares

(a)

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

thus

$$H = \sum_{t=1}^{m} \begin{pmatrix} x_1^{(t)}x_1^{(t)} & x_1^{(t)}x_2^{(t)} & \cdots & x_1^{(t)}x_{n+1}^{(t)} \\ x_2^{(t)}x_1^{(t)} & x_2^{(t)}x_2^{(t)} & \cdots & x_2^{(t)}x_{n+1}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ x_m^{(t)}x_1^{(t)} & x_m^{(t)}x_2^{(t)} & \cdots & x_m^{(t)}x_{n+1}^{(t)} \end{pmatrix} = \sum_{t=1}^{m} x^{(i)}(x^{(i)})^T \tag{2.1}$$

where $x^{(i)} \in R^{n+1}$.

(b) According to Newton's method, we have the following update

$$\theta^{t+1} = \theta^t - H^{-1} \bigtriangledown_\theta J(\theta) \tag{2.2}$$

Before iteration, suppose $\theta^t = 0$, after first iteration, we get

$$\theta = -H^{-1} \bigtriangledown_\theta J(\theta) = -H^{-1} \sum_{t=1}^{m} (\theta^T x^{(t)} - y^{(i)})x_i^{(t)}$$

Let

$$X = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix} \tag{2.3}$$

where $x^{(i)} = [x_i^{(1)}\ x_i^{(2)}\ \dots\ x_i^{(n+1)}]^T$. Caculate X^TX and $X^T\vec{y}$, then we get

$$X^TX = \begin{bmatrix} x^{(1)}\ x^{(2)}\ \dots\ x^{(m)} \end{bmatrix} \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix} = \sum_{t=1}^{m} x^{(i)}(x^{(i)})^T = H \tag{2.4}$$

$$X^T\vec{y} = \begin{bmatrix} x^{(1)}\ x^{(2)}\ \dots\ x^{(m)} \end{bmatrix} \begin{bmatrix} (y^{(1)}) \\ (y^{(2)}) \\ \vdots \\ (y^{(m)}) \end{bmatrix} = \sum_{t=1}^{m} x^{(i)}y^{(i)} \tag{2.5}$$

And

$$\sum_{t=1}^{m}(\theta^T x^{(t)} - y^{(i)})x_i^{(t)} = \sum_{t=1}^{m} -y^{(i)}x_i^{(t)} = -X^T\vec{y} \tag{2.6}$$

Thus

$$\theta^* = \theta = -H^{-1} \bigtriangledown_\theta J(\theta) = (X^TX)^{-1}X^T\vec{y}$$

3 Prediction using Linear Regression

(a) In linear regression, the least squares method is to try to find a straight line that minimizes the sum of Euclidean distances of all samples onto a straight line.

Hypothesis:

$$E_{(a,b)} = \sum_{i=1}^{m}(y_i - ax_i - b)^2 \tag{3.1}$$

Deriving a and b through $E_{(a,b)}$:

$$\frac{\partial E_{(a,b)}}{\partial a} = 2(a\sum_{i=1}^{m} x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i) \tag{3.2}$$

$$\frac{\partial E_{(a,b)}}{\partial b} = 2(mb - \sum_{i=1}^{m}(y_i - ax_i)) \tag{3.3}$$

Let the formulas (3.2) and (3.3) be zero to get the close-form of a and b optimal solutions.

$$a = \frac{\sum_{i=1}^{m} y_i(x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2} \tag{3.4}$$

$$b = \frac{1}{m}\sum_{i=1}^{m}(y_i - ax_i) \tag{3.5}$$

where $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$.

Solution:

$$\bar{x} = \frac{2005 + 2006 + 2007 + 2008 + 2009}{5} = 2007$$

$$a = \frac{12 \times (-2) + 19 \times (-1) + 29 \times 0 + 37 \times 1 + 45 \times 2}{(2005^2 + 2006^2 + 2007^2 + 2008^2 + 2009^2) - \frac{1}{5}(2005 + 2006 + 2007 + 2008 + 2009)^2}$$

$$= 8.4$$

$$b = \frac{1}{5}[(12 - 16842) + (19 - 16850.4) + (29 - 16858.8) + (37 - 16867.2) + (45 - 16875.6)]$$

$$= -16830.4$$

The resulting linear equation is

$$y = 8.4x - 16830.4 \tag{3.6}$$

(b) Substituting $x = 2012$ into equation 3.6,

$$y = 2012 \times 8.4 - 16830.4 = 70.4(\text{million dollars})$$