Solve Optimal Para of Least Square using Matrix Calculus

Shuai Li

December 12, 2013

Definition Suppose we have a $m \times n$ matrix X, a $n \times 1$ vector θ and a $m \times 1$ vector y. We want to make $X\theta$ be closest to y.

If we use Euclidean distance as distance metric, then the goal is to minimize $\frac{1}{2}(X\theta - y)(X\theta - y)^T$.

Since it is convex, the minimum is where the gradient is zero. The problem is to solve:

$$J(\theta) = \nabla_\theta \frac{1}{2}(X\theta - y)(X\theta - y)^T = 0 \tag{1}$$

Let's simplify it.

$$\begin{aligned}\nabla_\theta \frac{1}{2}(X\theta - y)(X\theta - y)^T &= \nabla_\theta \frac{1}{2}(X\theta - y)(\theta^T X^T - y^T) \\ &= \nabla_\theta \frac{1}{2}(X\theta\theta^T X^T - y\theta^T X^T - X\theta y^T + yy^T)\end{aligned}$$

Note the following fact:

- yy^T does not contain θ, $\nabla_\theta yy^T$ is zero.
- $J(\theta)$ is a real number. The trace of $J(\theta)$ is itself.

Then left side of eq. (1) becomes:

$$\begin{aligned}\nabla_\theta \frac{1}{2}(X\theta\theta^T X^T - y\theta^T X^T - X\theta y^T + yy^T) &= \frac{1}{2}\nabla_\theta tr(X\theta\theta^T X^T - y\theta^T X^T - X\theta y^T) \\ &= \frac{1}{2}(\nabla_\theta trX\theta\theta^T X^T - \nabla_\theta try\theta^T X^T - \nabla_\theta trX\theta y^T)\end{aligned} \tag{2}$$

For each term:

$$\begin{aligned}\nabla_\theta trX\theta\theta^T X^T &= \nabla_\theta tr\theta\theta^T X^T X \\ &= \nabla_\theta tr\theta I\theta^T X^T X \\ &= X^T X\theta + X^T X\theta \\ &= 2X^T X\theta\end{aligned}$$

$$\begin{aligned}\nabla_\theta try\theta^T X^T &= \nabla_\theta trX\theta y^T \\ &= \nabla_\theta tr\theta y^T X \\ &= X^T y\end{aligned}$$

$$
\begin{aligned}
\nabla_\theta tr X\theta y^T &= \nabla_\theta tr \theta y^T X \\
&= X^T y
\end{aligned}
$$

Combine them, eq. (1) becomes:

$$
\begin{aligned}
\frac{1}{2}(2X^T X\theta - 2X^T y)) &= 0 \\
X^T X\theta &= 2X^T y \\
\theta &= (X^T X)^{-1} X^T y
\end{aligned}
$$

This is the close form solution of least square method. ∎