多特征情况：X为特征矩阵，$\vec{\theta}$为参数向量，$\vec{y}$为真实值向量

偏置已加入

$$h_{\vec{\theta}}(X) = X\vec{\theta}$$

$$L(\vec{\theta}) = \frac{1}{2m}\| h_{\vec{\theta}}(X) - \vec{y} \|^2$$

$$= \frac{1}{2m}(h_{\vec{\theta}}(X) - \vec{y})^T (h_{\vec{\theta}}(X) - \vec{y})$$

$$= \frac{1}{2m}(X^T\vec{\theta}^T X\vec{\theta} - \vec{\theta}^T X^T \vec{y} - \vec{y}^T X\vec{\theta} + \vec{y}^T\vec{y})$$

$$\frac{\partial L(\vec{\theta})}{\partial \vec{\theta}} = \frac{\partial (X^T\vec{\theta}^T X\vec{\theta} - \vec{\theta}^T X^T \vec{y} - \vec{y}^T X\vec{\theta} + \vec{y}^T\vec{y})}{\partial \vec{\theta}}$$

$$= \frac{\partial X^T\vec{\theta}^T X\vec{\theta}}{\partial \vec{\theta}} - \frac{\partial \vec{\theta}^T X^T \vec{y}}{\partial \vec{\theta}} - \frac{\vec{y}^T X\vec{\theta}^T}{\partial \vec{\theta}} + \frac{\vec{y}^T\vec{y}}{\partial \vec{\theta}} = 0$$

$$= 2\vec{\theta}^T X^T X - \vec{y}^T X - \vec{y}^T X + \vec{0}^T = 0$$

$$2\vec{\theta}^T X^T X - 2\vec{y}^T X = 0$$

$$\vec{\theta} = (X^T X)^{-1} X^T \vec{y}$$

$\vec{\theta}$为最优解