1、

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad \text{特征1 特征2 特征n}$$

$$\sim \hat{y} = X \cdot w$$

$$loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - X \cdot w)^2$$

目标：$\min(loss)$

2、求解（只有w）

$$\frac{\partial loss}{\partial w} = \frac{\partial \|y - X \cdot w\|_2^2}{\partial w}$$

$$= \frac{\partial (y - X \cdot w)^T (y - X \cdot w)}{\partial w}$$

$$= \frac{\partial (y^T - w^T X^T)(y - X \cdot w)}{\partial w}$$

$$= \frac{\partial y^T y - \partial y^T X w - \partial w^T X^T y + \partial w^T X^T X w}{\partial w}$$

$$= 0 - X^T y - X^T y + \left[X^T X + (X^T X)^T \right] w$$

$$= -2X^T y + 2X^T X w = 0$$

$$X^T X w = X^T y$$

$$(X^T X)^{-1} (X^T X) w = (X^T X)^{-1} X^T y$$

$$w = (X^T X)^{-1} X^T y$$

a为常数

$$\frac{\partial a}{\partial A} = 0 \quad \frac{\partial A^T B^T C}{\partial A} = B^T C \quad \frac{\partial C^T B A}{\partial A} = B^T C$$

$$\frac{\partial A^T B A}{\partial A} = (B + B^T) A$$