$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$ 数据矩阵

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$ 标签向量

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$ 回归系数.

$$X^T w = y.$$

已知 X, y

找出使误差最小的 w.

平方误差和： $\sum_{i=1}^{m} (y_i - x_i^T w)^2$

矩阵表示： $(y - Xw)^T (y - Xw)$

$$\frac{\partial (y-Xw)^T(y-Xw)}{\partial w}$$

$(Xw)^T = w^T X^T$

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

分母列向量, 分子行向量, 分母布局 — 分子标量, 转置不变 — 分母布局

$$= 0 - X^T y - \frac{\partial (w^T X^T y)^T}{\partial w} + 2X^T X w$$

$$\frac{\partial y^T X w}{\partial w}$$

$$= 2X^T (Xw - y) = 0.$$

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