$(Xw-y)^T(Xw-y)$

↓ L2正则

$(Xw-y)^T(Xw-y)+\frac{1}{2}\alpha w^Tw$

↓

$$\begin{cases} \text{无正则}: & w=(X^TX)^{-1}X^Ty \\ \text{有正则} & w=(X^TX+\alpha I)^{-1}X^Ty \end{cases}$$

不同w仅仅是在对角加α，这使矩阵w对角项对应于每个特征输入的协方差，

L2正则化能让学习算法感知到具有较高方差w的特征输入x，因此与特征当目标w十协方差较小w的特征权重将会收缩。