三，线性回归

$f(w) = w^T x$ （当数据满足正态分布，最小二乘就是最大似然）

噪声

Loss function: $L(w) = \sum_{i=1}^{N} \|w^T x_i - y_i\|^2$

$\frac{dL(w)}{dw} \Rightarrow \hat{w} = (X^T X)^{-1} X^T Y$

$X_{N \times p}$，N个样本，$x_i \in \mathbb{R}^p$ $N \gg p$

过拟合 ⇒ ① 加数据
② 特征选择/特征提取 (PCA)
③ 正则化

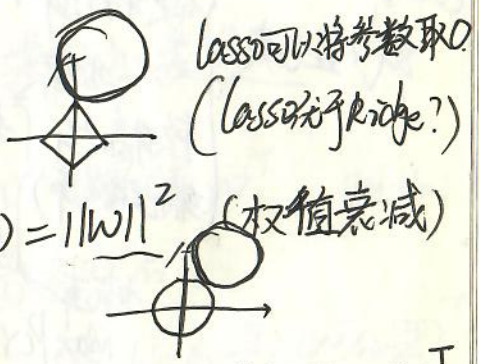
正则化框架

$\text{argmin}[L(w) + \lambda P(w)]$

loss　penalty.

L_1: Lasso, $P(w) = \|w\|$ （lasso可以将参数取0）

L_2: Ridge, 岭回归 $P(w) = \|w\|^2$ （权值衰减）



$X = (x_1, x_2 \cdots x_n)^T$

推导岭回归：

$\text{argmin}_w \sum_{i=1}^{N} \|w^T x_i - y_i\|^2 + \lambda w^T w$

$(w^T x_1 - y_1, w^T x_2 - y_2 \cdots w^T x_n - y_n) \begin{pmatrix} w^T x_1 - y_1 \\ w^T x_2 - y_2 \\ \vdots \\ w^T x_n - y_n \end{pmatrix} = w^T (x_1, x_2 \cdots x_n) - (y_1, y_2 \cdots y_n)$

$= (w^T X^T - Y^T)(Xw - Y) + \lambda w^T w = w^T X^T X w - 2 w^T X^T Y + \lambda w^T w + Y^T Y$