Ordinary Least Squares

Targets $\underline{y}$ Design matrix X $(n \times p)$

Linear comb of features : $X\underline{w}$

$\hat{\underline{w}}_{OLS}$ minimizes $\|y - X\underline{w}\|^2$.

i^{th} example $y_i = x_{i1} w_1 + x_{i2} w_2 + \cdots + x_{ip} w_p + \epsilon_i$

OLS : minimize $\sum \epsilon_i^2$.

If $\epsilon_i \sim N(0, \sigma^2)$, ϵ_i independent

Ridge Regression

$\underline{y}$ X

(1) $\hat{w}_{Ridge} = \arg\min\limits_{\underline{w}} \|\underline{y} - X\underline{w}\|^2$

subject to $\|\underline{w}\|^2 \leq B$. —— (v1)

Completely equivalent

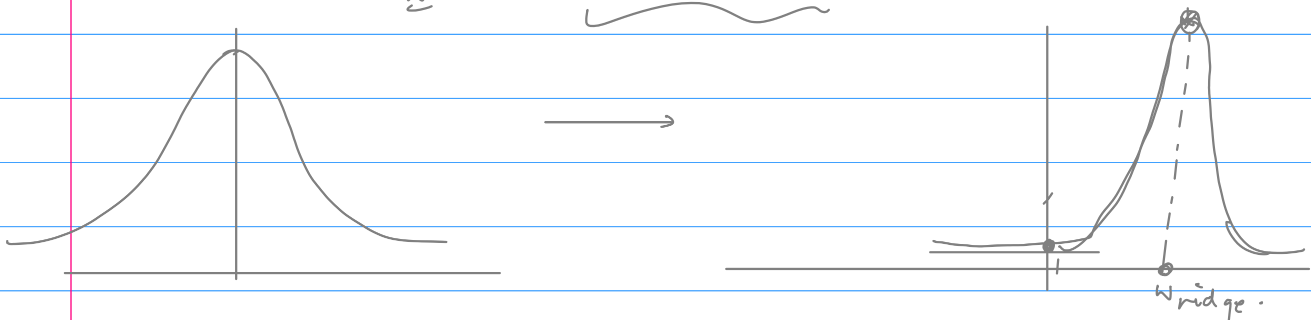
$\arg\min\limits_{\underline{w}} \|\underline{y} - X\underline{w}\|^2 + \lambda \|\underline{w}\|^2$. —— (v2)

(2) $y_i = x_{i1} w_1 + x_{i2} w_2 + \cdots + x_{ip} w_p + \epsilon_i$

$\underline{w} \sim N(\underline{0}, I/\lambda)$.

$\{$ If $\underline{w} \sim N(0, \Sigma)$.

$\arg\min\limits_{\underline{w}} \|y - X\underline{w}\|^2 + \underline{w}^T \Sigma^{-1} \underline{w}$ $\}$



$$\hat{w}_{OLS} = (X^T X)^{-1} X^T \underline{y} = \arg\min_{\underline{w}} \|y - X\underline{w}\|^2.$$

Predict on a new test case $\underline{z}$: $\underline{z}^T \hat{\underline{w}}_{OLS}$

$$\hat{w}_{ridge} = (X^T X + \lambda I)^{-1} X^T \underline{y} = \arg\min_{w} \|y - X\underline{w}\|^2 + \lambda \|\underline{w}\|^2.$$

Predict on a new test case $\underline{z}$: $\underline{z}^T \hat{w}_{ridge}$.

$$\text{But} \quad \underline{z}^T \underline{\hat{w}}_{ridge} = \underline{z}^T (X^T X + \lambda I)^{-1} X^T \underline{y}$$
$$= z^T X^T \underbrace{(X X^T + \lambda I)^{-1}}\, \underline{y}. \quad \Big\} \text{ Matrix inversion lemma.}$$

$$X = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \underline{x}_3^T \\ \underline{x}_4^T \end{bmatrix} \qquad X^T = \begin{bmatrix} | & | & | & | \\ \underline{x}_1 & \underline{x}_2 & \underline{x}_3 & \underline{x}_4 \\ | & | & | & | \end{bmatrix}.$$

$$\underline{z}^T X^T = \left(\underline{z}^T \right) \begin{pmatrix} | & | & | & | \\ \underline{x}_1 & \underline{x}_2 & \underline{x}_3 & \underline{x}_4 \\ | & | & | & | \end{pmatrix} = \left(\underline{z}^T \underline{x}_1 \quad z^T x_2 \quad \cdots \quad z^T x_4 \right).$$

$$X X^T = \begin{bmatrix} \underline{x}_1^T \underline{x}_1 & \underline{x}_1^T x_2 & \underline{x}_1^T x_3 & \underline{x}_1^T x_4 \\ & \vdots & & \\ & \vdots & & \\ & & & \underline{x}_4^T \underline{x}_4 \end{bmatrix} \longrightarrow k(X, X) = \begin{bmatrix} k(\underline{x}_1, \underline{x}_2) & & \\ & \ddots & \\ k(x_3, x_1) & & \\ & & k(x_4, x_4) \end{bmatrix}$$

$$\underline{x}_1 \longrightarrow \phi(\underline{x}_1)$$
$$\underline{x}_2 \longrightarrow \phi(x_2)$$