Regularized Regression

Saturday, July 8, 2023 10:44 AM

Ordinary Least Squares :- (OLS)

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

$$\text{Error } f^n \quad Z = \sum \left[y_i - (b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k) \right]^2$$

$$\underbrace{\frac{\partial Z}{\partial b_0} = 0, \ \frac{\partial Z}{\partial b_1} = 0, \ \cdots \frac{\partial Z}{\partial b_k} = 0}_{k+1 \text{ equations}}$$

Ridge Regression (l_2)

The coeff of less influential variables, reduce to a lower value $\neq 0$

$$\text{Error } f^n \quad Z = \sum \left[y_i - (b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k) \right]^2 + \alpha \sum_{j=1}^{k} b_j^2$$

$$\underbrace{\frac{\partial Z}{\partial b_0} = 0, \ \frac{\partial Z}{\partial b_1} = 0, \ \cdots \frac{\partial Z}{\partial b_k} = 0}_{k+1 \text{ equations}}$$

α : a non-negative number : Regularization Parameter

$$\sum_{j=1}^{k} b_j^2 = b_1^2 + b_2^2 + \cdots + b_k^2 : l_2 \text{ norm}$$

Lasso Regression :- (l_1)

The coeff of less influential variables, reduce to zero

$$\text{Error } f^n \quad Z = \frac{1}{2n} \sum \left[y_i - (b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k) \right]^2 + \alpha \sum_{j=1}^{k} |b_j|$$

$$\underbrace{\frac{\partial Z}{\partial b_0} = 0, \ \frac{\partial Z}{\partial b_1} = 0, \ \cdots \frac{\partial Z}{\partial b_k} = 0}_{k+1 \text{ equations}}$$

α : a non-negative number : Regularization Parameter

$$\sum_{j=1}^{k} |b_j| = |b_1| + |b_2| + \cdots + |b_k| : l_1 \text{ norm}$$

Elastic net : (l_1 l_2)

n = training obs

$$\text{Error } f^n \quad Z = \frac{1}{2n} \sum \left[y_i - (b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k) \right]^2 + \alpha * l_1\_ratio * \sum_{j=1}^{k} |b_j|$$

$$\frac{1}{2n} L(\quad) + \alpha * l1\_ratio * \sum_{j=1}^{k} |b_j| + 0.5 * \alpha * (1 - l1\_ratio) * \sum_{j=1}^{k} b_j^2$$

$0 \leq l1\_ratio \leq 1$

$b_0, b_1, \ldots b_K$: Parameters

Model	Hyper-Parameters
Ridge	Alpha
Lasso	Alpha
Elastic Net	Alpha, l1_ratio

Step-wise Regression:- $X_1\ X_2 \cdots X_K\ \ y$

a) Forward Selection :

(X_1, y) , (X_1, X_2, y) ---- $(X_1, X_2 \ldots X_K)\ y$

err_1 err_2 err_K

b) Backward Elimination:-

$(X_1, X_2 \ldots X_K, y)$ $(X_1, X_2 \ldots X_{K-1})\ y$ ---- $(X_2\ X_1)\ y$ $X_1\ y$

err_1 err_2 ---- err_{K-1} err_K

c) Step-wise :-

(X_1, y) $(X_1\ X_2\ y)$ $(X_1\ X_2\ X_3\ y)$ -----

err_1 err_2 err_3

$err_2 > err_1$

then drop X_2