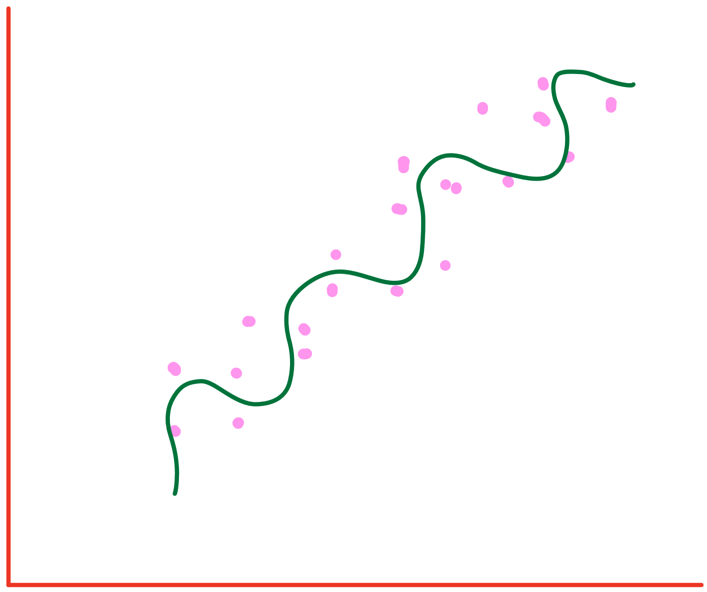
Regularization



$$\hat{y} = w_0 + w_1 f_1 + w_2 f_2 + w_3 f_3 \cdots$$

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} \Rightarrow \text{reduce Weights}$$

$w_2 = 0$ and $w_3 = 0$

⇒ New Loss Function

Loss = MSE + Regularization

$$\sum_{i=1}^{n} \frac{1}{n}(y_i - \hat{y}_i)^2 + \sum_{j=1}^{d} (\omega_j)^2$$

(Sum of w's squared)

(Ridge Regression)

$$\frac{\partial (\omega_1^2 + \omega_2^2 + \omega_3^2 \; --- \; \omega_d^2)}{\partial \omega_{j=1}} \Rightarrow 2\omega_j$$

$|\omega_j| \Rightarrow 5.0$

$|\omega_j| \Rightarrow 0.09$

$j=1 \Rightarrow 2\omega_1$

100 → 80

⇒ Loss = MSE + Regularization

⇒ reducing Loss function

→ 0

$\omega_1, \omega_2, \omega_0 \Rightarrow 0$

$[w] \Rightarrow 0$

MSE $\uparrow$

$\downarrow$ Regularization

Regularization Parameter

(Bias) $\uparrow$ (variance)

$$\text{Loss} = \text{MSE} + \lambda \ \text{Regularization}$$

$\downarrow$

≥ 0

0.01

0.02

0.03

$\downarrow$

Hyperparameter tuning $\Rightarrow \lambda$

* Reg Term $\Rightarrow 0 \Rightarrow$

$\downarrow$

overfit or Underfit ?

Lasso Regression

Loss = MSE + L1 Regularization

$$\sum_{i=1}^{n} \frac{1}{n}(y_i - \hat{y}_i)^2 + \sum_{j=1}^{d} |(w_j)|$$

$$\frac{\partial (\text{L1 Regularization})}{w_j} \Rightarrow \begin{cases} 1 \\ \text{Undefined} \\ -1 \end{cases} \rightarrow 0$$

at $w_j = 0$

$\partial w_j \in (-\infty, \infty) \Rightarrow \partial w_j$ (more stable)

$\partial w_j \ni (-1, 0, 1) \Rightarrow$ Better chance of Becoming 0

Elastic Net Regularization

$$\text{Loss} \Rightarrow MSE + \lambda_1 |w| + \lambda_2 w_j^2$$

Topic Remaining

1. Hyperparameter Tuning

2. Cross Validation

3. K-fold Cross Validation