Regularization Techniques

→ Regularization Techniques used in linear Regression to address overfitting and improve model generalizability.

DATASET
TRAIN   TEST

MODEL
↓
{ TRAIN → ACC ↑↑ — R^2 ↑
TEST → ACC ↓↓ → R^2 ↓ }
overfitting conditions

① Ridge Regression (L2)

② Lasso Regression (L1)

③ Elastic Regression

Reduce overfitting

Ridge Regression (L2 Reguarlization)

Ridge regression adds a penalty term to the linear regression cost function, where the penalty is the L2 norm of the coefficients multiplied by a regularization parameter, usually denoted as λ (lambda). The goal is to minimize the sum of squared errors between the predicted and actual values of the target variable, while also penalizing large coefficients.

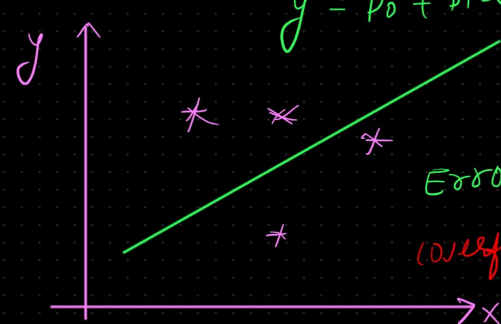
slope

$\hat{y} = \beta_0 + \beta_1 x$

Cost function
↓

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

MSE



Error = 0
(overfitting)

Cost function of Ridge Regression

$$\boxed{J(\theta) = \underset{\downarrow \atop 0}{MSE(\theta)} + \underbrace{\lambda \sum_{i=1}^{n} \theta_i^2}_{\text{penalty term}}}$$

coeff. / slope

$$J(\theta) = 0 + 1[(\theta_1)^2] \Leftarrow \text{penalize the cost func}^n$$

$J(\theta)$ = cost func^n

$MSE(\theta)$ = Mean Square Error

λ = regularization parameter / hyperparameter

θ_i = coefficient / slope of linear Regression model

Goal: Reduces the variance of the model by shrinking the coefficients towards zero, but not necessarily setting them to zero.

② Lasso Regression (L1 Reguralization)

(Feature selection)

Cost func^n

$$\boxed{J(\theta) = MSE(\theta) + \lambda \underbrace{\sum_{i=1}^{n} |\theta_i|}_{L1\ Norm}}$$

Goal: Shrinks coefficients towards zero and can even set some coefficients to exactly zero, performing feature selection

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Lasso regression, similar to Ridge regression, adds a penalty term to the linear regression cost function. However, Lasso uses the L1 norm of the coefficients as the penalty instead of the L2 norm.

③ Elastic Net Regression

Cost func^n

$$\boxed{J(\theta) = MSE(\theta) + \lambda_1 \sum_{i=1}^{n} \theta_i^2 + \lambda_2 \sum_{i=1}^{n} |\theta_i|}$$

$\lambda_1, \lambda_2 \rightarrow$ regularization parameters for Ridge and Lasso

Hyperparameter Tuning

Elastic Net Regression combines both L1 and L2 penalties in the linear regression cost function. This allows it to handle multicollinearity (correlations between predictors) better than either Ridge or Lasso regression alone.

L2 → Overfitting

L1 → Feature selection

L1 and L2 → Elastic → multicollinearity

$$\text{Cost func}^n \; J(\theta) = MSE(\theta) + \underbrace{\lambda_1 \sum_{i=1}^{n} (\theta_i)^2}_{\text{Reduce Overfitting}} + \underbrace{\lambda_2 \sum_{i=1}^{n} |\theta_i|}_{\text{Feature selection}}$$