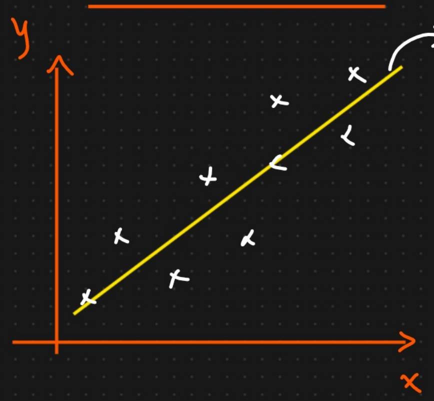
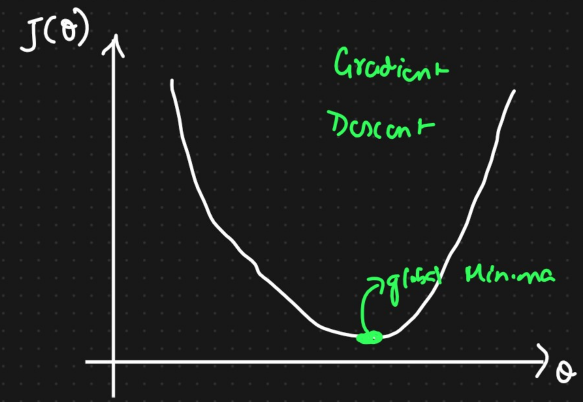
Ridge, Lasso And Elasticnet Regression

Linear Regression



Independent

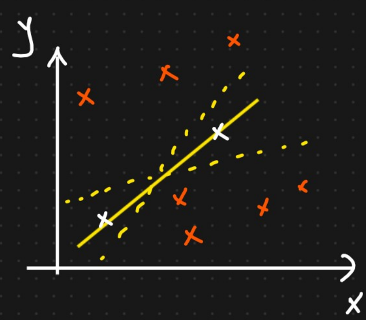
$h_\theta(x) = \theta_0 + \theta_1 x$



$\hat{y}_i$

Cost fn : $\frac{1}{n} \sum_{i=1}^{n} (y_i - h_\theta(x)_i)^2$ [Mean Squared Error]

① Ridge Regression (L2 Regularization) ⟶ Reducing Overfitting



Error = 0

Error ↑↑

Test DATA

DATASET

TRAIN TEST

Ridge regression is a type of linear regression where the coefficients are determined not only by minimizing the sum of squared residuals, like in ordinary least squares (OLS) regression, but also by adding a penalty term that shrinks the coefficients towards zero. This penalty term is a regularization parameter multiplied by the square of the coefficients, which helps to reduce overfitting and improve the model's generalization ability.

Overfitting

TRAIN ⟶ Acc ↑↑ ⟶ R^2

TEST ⟶ Acc ↓↓ ⟶ R^2

$h_\theta(x) = \theta_0 + \theta_1 x_1$

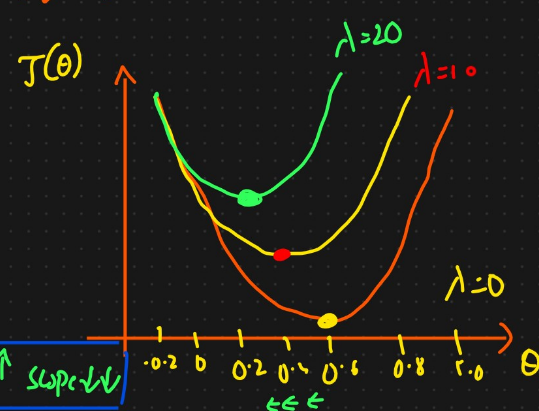
$\lambda = 1$

Cost fn = $\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boxed{\lambda \sum_{i=1}^{n} (\text{slope})^2}$

⇓

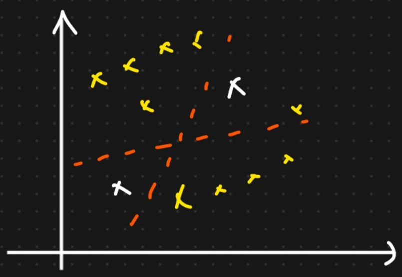
0

Hyperparameter



$\boxed{\lambda \uparrow\uparrow \text{ slope} \downarrow\downarrow}$

$= \quad 0 + 1 [(\theta_1)^2]$ ⇐ penalize the cost function



$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

$$\boxed{\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 + \lambda\left[(\theta_1)^2 + (\theta_2)^2 + (\theta_3)^2\right]}$$

② Lasso Regression (L_1 Regularization) → Feature Selection

$$\underline{\text{Cost fn}} : \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|}$$

In Lasso regression, the penalty term is the absolute sum of the coefficients of the features multiplied by a constant, often denoted as alpha or lambda. This penalty encourages sparsity in the coefficients, meaning it tends to force some coefficients to be exactly zero, effectively selecting a subset of the features and performing feature selection.

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$

feature remove

$= 0.52 + \underbrace{0.65x_1 + 1.5x_2}_{\text{prediction}} + \boxed{\cancel{0.2x_3}}$

feature selection

$x_2 \rightarrow 1$ unit

$y \rightarrow 1.5x_2$

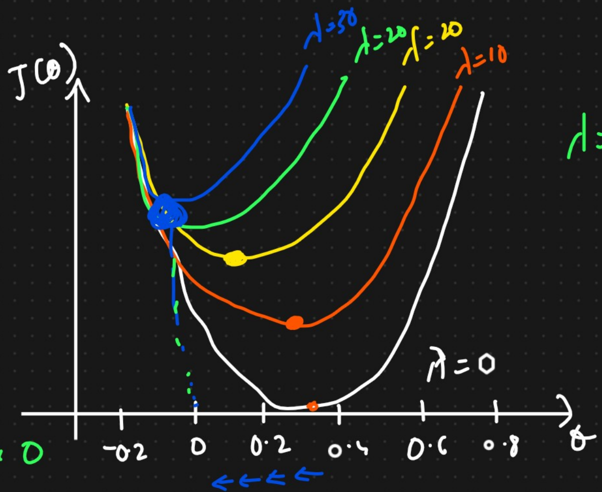
$x_3 \rightarrow 1$ unit

$y \rightarrow 0.2x_3$

$\lambda = 1$

$$\underline{\text{Cost fn}} : \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|}$$

$= \text{Error} + 1\left[|\theta_1| + |\theta_2| + |\theta_3|\right]$



λ = Hyperparameter

θ & λ

$\lambda\uparrow\uparrow \quad \theta\downarrow\downarrow$

↓

It will become 0

③ ElasticNet Regression

- Reducing Overfitting → Ridge
- Feature Selection → Lasso

$$\boxed{\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{i=1}^{n}(\text{slope})^2 + \lambda_2 \sum_{i=1}^{n}|\text{slope}|.}$$

$\Downarrow$ MSE + $\Downarrow$ Reducing overfitting + $\Downarrow$ Feature Selection

λ_1, λ_2 {Hyperparameter Tuning}.

Elastic Net Regression is a type of linear regression that combines penalties from both Lasso (L1 regularization) and Ridge (L2 regularization) methods. It's particularly useful when dealing with high-dimensional data where there may be multicollinearity among the predictors.

The Elastic Net penalty term is a linear combination of the L1 and L2 penalties.

Elastic Net Regression can help with feature selection by pushing the coefficients of irrelevant features towards zero and by encouraging grouping effects when there are correlated predictors.