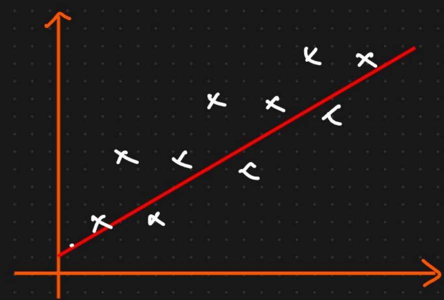
Regression

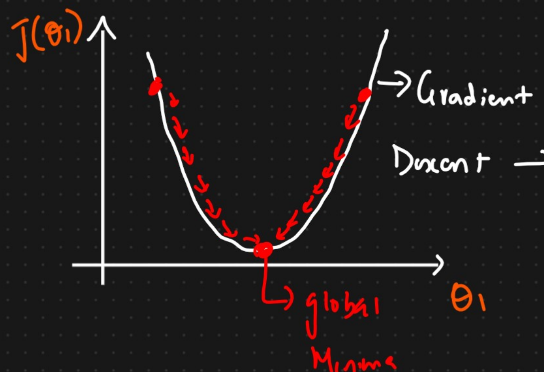
1. Mean Square Error (MSE), MAE, RMSE ✓
2. Ridge Regression ✓
3. Lasso Regression ✓
4. ElasticNet ✓
5. Practicals. [Simple Implementation]



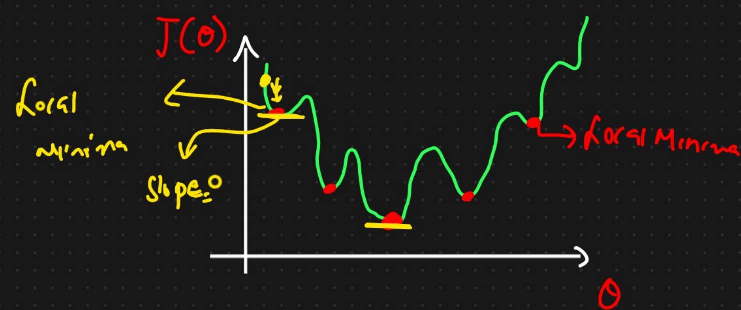
1. Mean Squared Error (MSE)
2. MAE
3. RMSE

$\uparrow\uparrow$ MSE

$$J(\theta_0,\theta_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)\right)^2$$

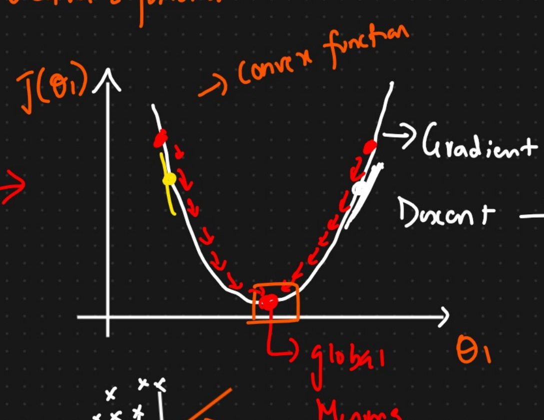


Descent → optimize



1. Mean Square Error [Cost function].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \Rightarrow \text{Quadratic function.}$$



$$(a+b)^2 = a^2 + 2ab + b^2$$

$$\boxed{ax + by + c = 0}$$

$\Downarrow$

$$\boxed{y = mx + c}$$

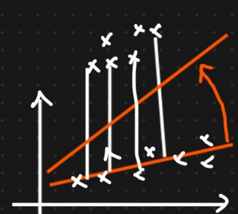
$$\boxed{(y_i - \hat{y}_i)^2}$$

INR $(INR)^2$

US $(US)^2$

ERROR $\uparrow\uparrow$

MSE



$ax + by + c = 0$

$by = -ax - c$

$$y = \boxed{\frac{-a}{b}}x + \boxed{\frac{-c}{b}}$$

$\Downarrow$ mx + c

MSE $(y_i - \hat{y}_i)^2 \uparrow\uparrow\uparrow$

Advantages

① It is differentiable.

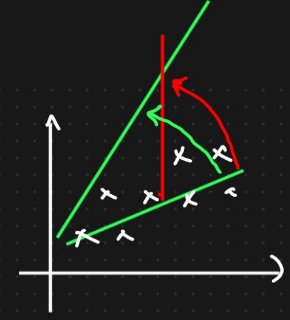
② It has one local and one global minima.

Disadvantages

① Not Robust to outliers

{Affected by outliers}

② MSE changes the unit



Price	Predicted price (↓Model)	$(y-\hat{y}_i)^2$ $(INR)^2$
100	120	400

② Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}| \Rightarrow$$

Advantage

ERROR ↑

But not as much as MSE

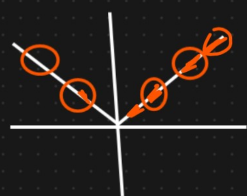
hence

① Robust to outliers

② It will be in same unit

Disadvantage

Time Complexity is more for optimizer.



→ Sub gradient → optimizer

↑ convert

Not getting Gradient Descent

Split gradient by gradient and find then converge

③ Root Mean Square Error

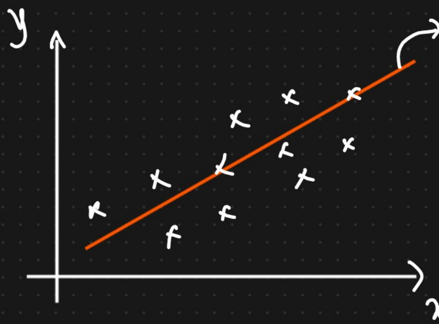
$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \Rightarrow$$ Advantages And Disadvantages.

Performance Metric = R^2 and Adjusted R^2

Cost function = MSE, MAE, RMSE, $\boxed{\text{Huber loss}}$ ← DL.

Ridge, Lasso and ElasticNet

Linear Regression



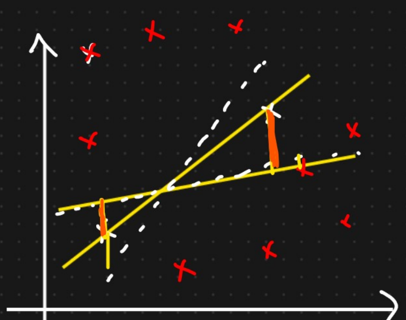
$h_\theta(x) = \theta_0 + \theta$

$$\underline{\text{Cost function}} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - h_\theta(x)\right)^2$$

① Ridge Regression

L2

(L2 Regularization) ⟶ Reducing Overfitting }



TRAINING

Acc ⇒ 100%

⇓

Overfitting

TRAININ

DATA 100 ↑↑

Test Data

Acc ↓↓.

DATASET

Book ⇓ ⇒ TRAINING ⇑ TRAIN Model

TEST ⇑ Test the Accuracy or performance of the model

⇓ DATA Leakage

⇓ Exam

⇓. ⇗ Ridge Regression

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

$\frac{1}{n}$ ⇓ 0

λ → Hyperparameter

slope → Coefficients

$\boxed{\lambda \geq 0}$

$\lambda = 1$

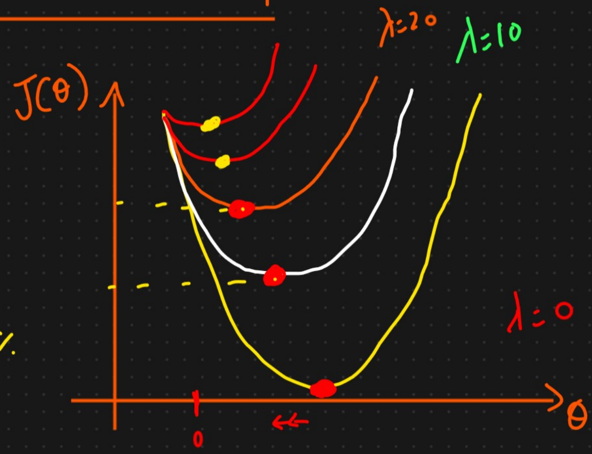
$+ 1 * \left[(\theta_1)^2 + (\theta_2)^2 + (\theta_3)^2\right]$ ⇒ +ve value

Cost fn = +ve value ↓↓↓

Coefficients ↓↓

$\lambda = 1, 2, 3, 4, 10, 20, 30, 40, 50$

Relationship between λ and slope



TRAINING

Accuracy = 100%

TRAINING ERROR

$$\text{Cost} = \boxed{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} + \boxed{\lambda \cdot \sum_{i=1}^{n}(\text{slope})^2} \uparrow 0$$

$\lambda \uparrow\uparrow$ slope ↓↓

$\lambda \uparrow\uparrow\uparrow$ slope ↓↓↓↓

$\lambda \uparrow\uparrow\uparrow$ ERROR ↑↑↑

COST ↑↑↑

② Lasso Regression (L1 Regularization) ⟶ Feature Selection

$$\text{Cost fn}: \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x)_i)^2 + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|} \Leftarrow L_1 \text{ Regularization}$$

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

$h_\theta(x) = 0.52 + \boxed{0.65x_1} + 0.72x_2 + \boxed{0.12x_3}$

⇓

Lasso Regression

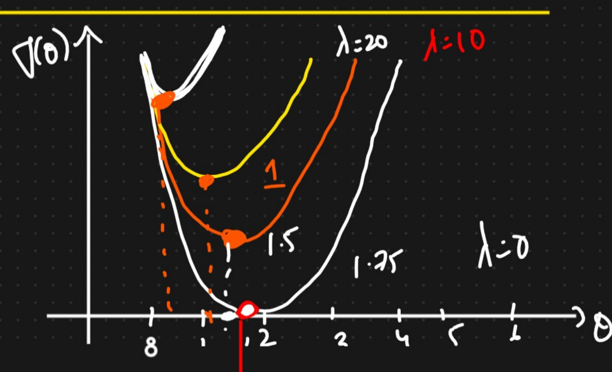
unit change in x_1

0.65 change in output

unit change in x_3

0.12 change in output

Relationship between λ and |slope|



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

$h_\theta(x) = 0.52 + \underbrace{\boxed{0.65x_1} + \boxed{0.72x_2}} + \boxed{\cancel{0.12x_3}}$

0

λ - . . .

$\boxed{0.12, 0.20, 0.30}$

Global Minima.

$\boxed{\theta} \Rightarrow$ Not Important.

$$h_\theta(x) = 0.52 + \boxed{0.65}\,x_1 + \boxed{0.72}\,x_2$$

$$= 0.52 + 0.45x_1 + 0.32x_2 + \boxed{}$$

③ <u>ElasticNet</u>

→ ① Reducing overfitting
→ ② Feature Selection

} Ridge + Lasso

$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x_i))^2 + \boxed{\lambda_1 \sum_{i=1}^{n}(\text{slope})^2} + \boxed{\lambda_2 \sum_{i=1}^{n}|\text{slope}|}$$

$\Downarrow$ Reducing overfitting

$\Downarrow$ Feature selection.

λ_1, λ_2 [Hyperparameters].