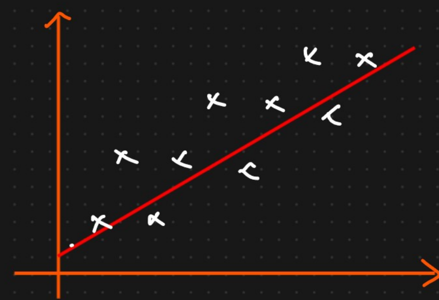
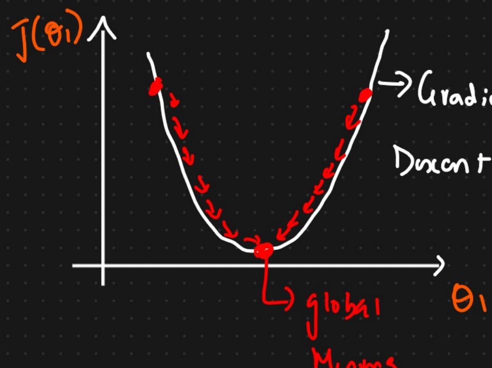
# Regression

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



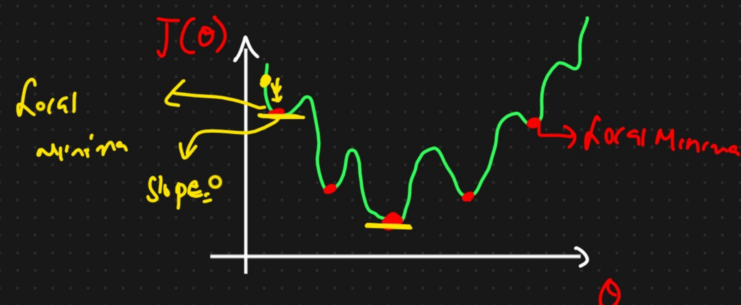
1. Mean Squared Error (MSE)
2. MAE
3. RMSE




Descent → optimize

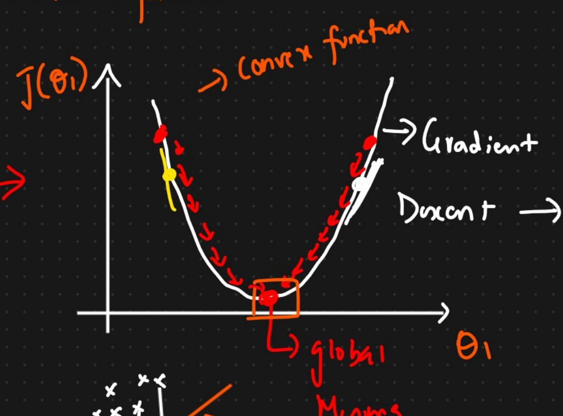
↑↑ MSE

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




## 1. Mean Square Error [Cost function].

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




Descent →

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

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

⇓

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

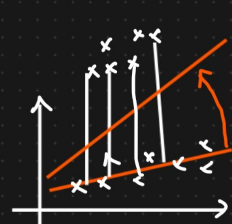
$$\boxed{(y_i - \hat{y}_i)^2}$$

INR $(INR)^2$

$\boxed{US}$ $(US)^2$

ERROR ↑↑

MSE



$ax + by + c = 0$

$by = -ax - c$

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

⇓ ⇓

$mx + c$

MSE $(y_i - \hat{y}_i)^2$ ↑↑↑

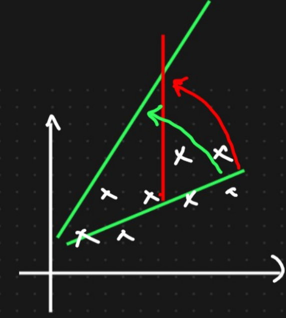
Advantages

1. It is differentiable.
2. It has one local and one global minima.

Disadvantages

1. Not Robust to outliers

{Affected by outliers}



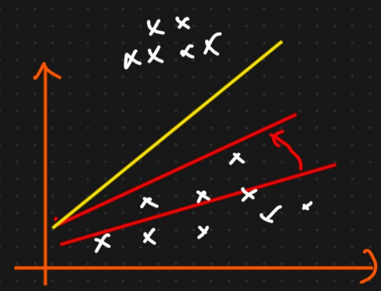
2. MSE changes the unit

| Price | ↓Model Predicted price | $(INR)^2$ $(y-\hat{y}_1)^2$ |
|---|---|---|
| 100 | 120 | 400 |

## ② Mean Absolute Error (MAE)

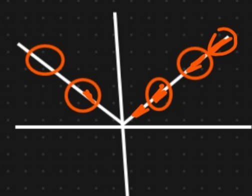
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y-\hat{y}| \Rightarrow$$

ERROR ↑



Advantage

1. Robust to outliers
2. It will be in same unit



→ Sub gradient

Disadvantage

Time Complexity is more for optimizer.

## ③ Root Mean Square Error

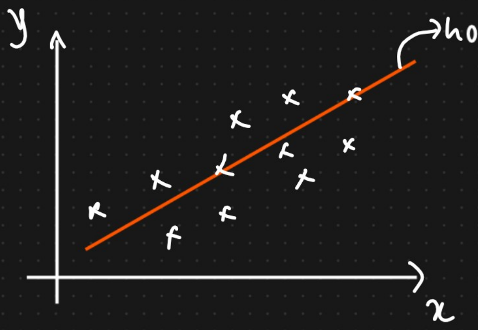
$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \Rightarrow$$ Advantages And Disadvantages.

Performance Metric = $R^2$ and Adjusted $R^2$

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

# Ridge, Lasso and ElasticNet

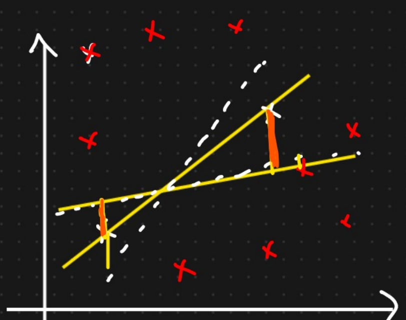
## Linear Regression




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

## ① Ridge Regression

(L2 Regularization) ⟶ Reducing Overfitting }



TRAINING

ACC ⇒ 100%

⇓

Overfitting

TRAININ DATA 100 ↑↑

Test Data ACC ↓↓.

DATASET

⇓ DATA Leakage

Book ⇓ ⇒ TRAINING — TRAIN Model

⇓ Exam — TEST — Test the Accuracy or performance of the model

⇓. ⇗ Ridge Regression

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

⇓ 0

$\lambda$ → Hyperparameter

slope → Coefficients

$\boxed{\lambda \geq 0}$

$\lambda = 1$

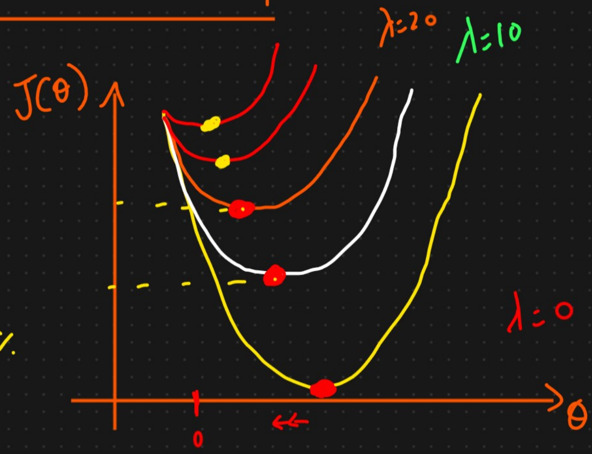
$+ 1 * [(\theta_1)^2 + (\theta_2)^2 + (\theta_3)^2]$ ⇒ +ve Value

Cost fn = +ve value ↓↓↓

Coefficients ↓↓

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

## Relationship between $\lambda$ and slope




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

$\uparrow 0$

$\lambda \uparrow\uparrow$ slope ↓↓

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

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

COST ↑↑↑

TRAINING

Accuracy = 100%

TRAINING ERROR

## ② Lasso Regression ($L_1$ Regularization) ⟶ Feature Selection

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

unit change in $x_1$

0.65 change in output

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

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

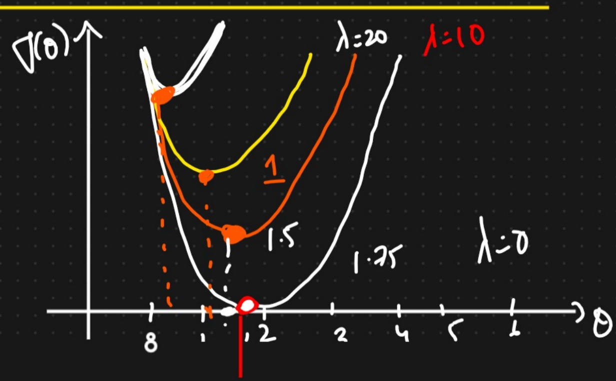
⇓

Lasso Regression

unit change in $x_3$

0.12 change in output

## Relationship between $\lambda$ and |slope|




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

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

⇓

0

$\lambda$ - . . .

$\boxed{0.12, 0.20, 0.30}$


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

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

③ ElasticNet

- 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.

$\lambda_1, \lambda_2$ [Hyperparameters].