Data Science & AI

Machine Learning

Supervised Learning

Lecture No.- 02

By- Krish Naik Sir

Recap of Previous Lecture

Topic Introduction

Topic Regression & Its types

Topic

Topic

Topic

Topics to be Covered

Topic

Topic

Topic

Topic

Topic

Ridge
Lasso
Elastic
Regression
Cost Function

Machine Learning

1. Multiple Linear Regression ✓
2. R Squared, Adjusted R Squared [Metrics] → Correlation ✓.
3. Types of Cost function
   - i) Mean Squared Error (MSE)
   - ii) Mean Absolute Error (MAE)
   - iii) Root Mean Squared Error (RMSE) ✓.
4. Overfitting, Underfitting ✓
5. Ridge, Lasso, Elastinet ML Algorithms ✓.

1. Multiple Linear Regression

Intercept

Slope

Simple Linear Regression → $h_\theta(x) = \beta_0 + \beta_1 x_1$ → I/p feature ⇓ Independent feature

⇓

We have 1 i/p feature

House Price Dataset

I/p feature

↓ O/p

No.of Rooms	Size of Rooms	Price
x_1	x_2	

$h_\theta(x) \to$ 1 unit ↑↑

$0.5x_1 \quad 0.75x_2$

$\theta_1 = 0.5 \qquad \theta_2 = 0.75$

Intercept

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

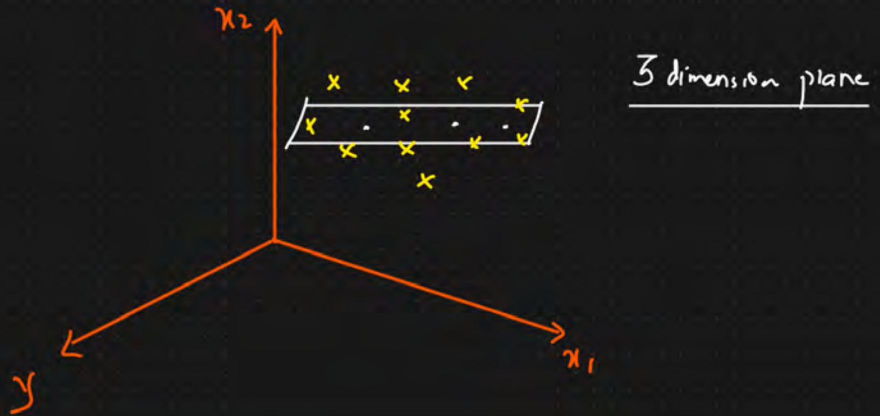
$$h_\theta(x) = \theta_0 + 0.5x_1 + 0.75x_2$$

$\theta_1 \to$ Coefficient of x_1

$\theta_2 \to$ Slope or coefficient of x_2.

Gradient Descent 3d

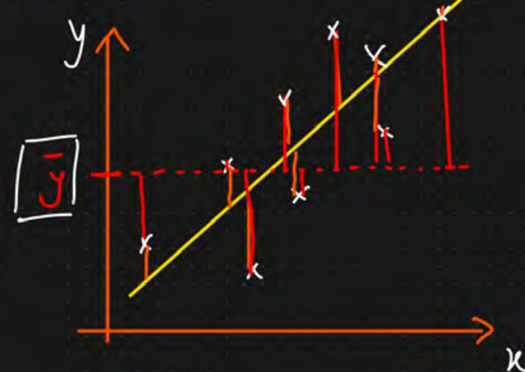
Error → Minimal



$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 \cdots + \theta_n x_n}$$

② Performance Metrics [R Squared, Adjusted R Squared]

① R Squared



$$R\ \text{Squared} = 1 - \frac{SS_{Res}}{SS_{Total}}$$

SS_{Res} = Sum of Squares Residual [Error] = $\sum_{i=1}^{n} (y_i - h_\theta(x)_i)^2$

↳ $\hat{y}_i \Rightarrow$ predicted

SS_{Total} = Sum of Square Total = $\sum_{i=1}^{n} (y_i - \bar{y}_i)^2$

If best fit line $\boxed{SS_{Total} >>> SS_{Res}.}$

$$R\ \text{Squared} = 1 - \frac{SS_{Res}}{SS_{Total}} \{\text{small}\} \approx 1$$

$\boxed{\text{If } SS_{Res} >> SS_{Total} \Rightarrow R\ \text{Squared} \Rightarrow -\text{ve value}}$

$\boxed{-\text{infinity} \longleftrightarrow 1}$ $\boxed{0 \text{ to } 1}$

House Size ↑ No of Bedrooms ↑ Location ↑ No of people → Price ↑↑

R^2 Squared = 0.76 ⇒ 76%

R^2 Squared = 0.80 ↑↑

R^2 Squared = 0.83 ↑↑

R^2 Squared = 0.84 ↑↑

X	y
2	3
3	4
5	6
7	8

X↑ y↑
X↓ y↓ ⇒

⇒ Mathematically

Adjusted R Squared = 0.82 ↓↓

X	y
1	5
2	4
3	3
4	2
5	1

X↑ y↓
X↓ y↑ ⇒

① Covariance → It quantifies the relationship between X and Y

$$\text{Variance}(X) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\text{Covariance}(X,Y) = Cov_{(X,Y)} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}_1)$$

⇓

$\boxed{-\text{inf to } +\text{inf}}$

$$\boxed{Cov(X,X) = Var(X)}$$

X	y
2	3
3	4
5	6
7	8

$\boxed{\begin{matrix} X\uparrow y\uparrow \\ X\downarrow y\downarrow \end{matrix}}$

+ve Covariance

X	y
1	5
2	4
3	3
4	2
5	1

⟹ X↑ y↓
X↓ y↑

⇒ -ve Covariance

⇒ 0 Covariance

Covariance ÷ −infinity ⟷ + infinity.

0.92 +300

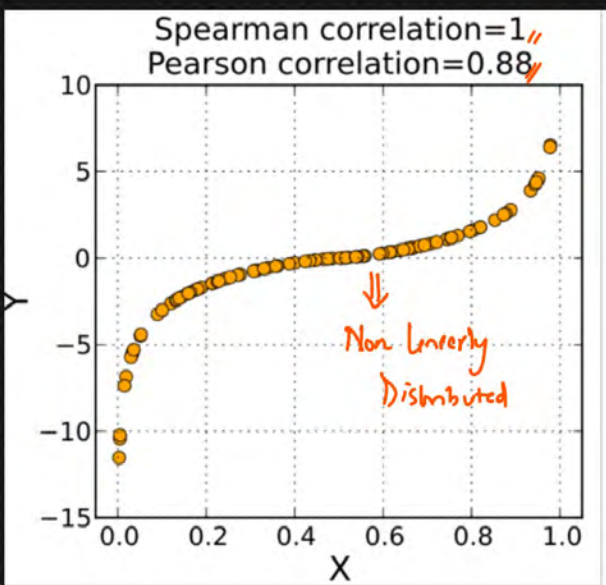
X Y Z ⇒ Pearson Correlation Coefficient

+450 0.95

② Pearson Correlation Coefficient ÷ [−1 to 1]

$$\rho_{x,y} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \Rightarrow -1 \text{ to } 1.$$

③ Spearman Rank Correlation [−1 to +1].



⇓.

X↑ Y↑
X↓ Y↓. ⇒ 1

$$r_s = \rho_{R(X),R(Y)} = \frac{cov(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}},$$

where

X	Y	R(X)	R(Y)
1	2	1	
5	3	2	
7	8	3	
9	1	4	

② Adjusted R Squared

$$\text{Adjusted R squared} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

N → No. of data points.

R^2 → R squared

P → No. of Independent feature

④ Types of cost function

① Mean Squared Error (MSE) ✓

② Mean Absolute Error (MAE) ✓

③ Root Mean Squared Error (RMSE) → Assignment

① Mean Squared Error (MSE)

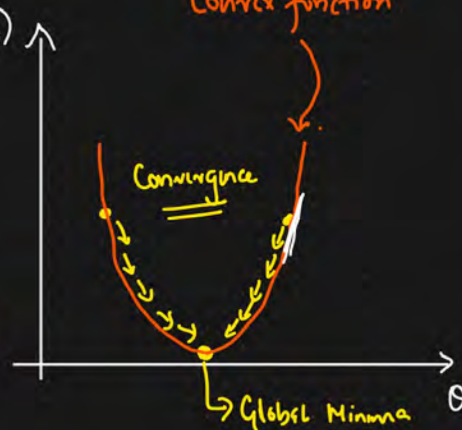
$$\boxed{MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)_i\right)^2}$$ → Quadratic Equation

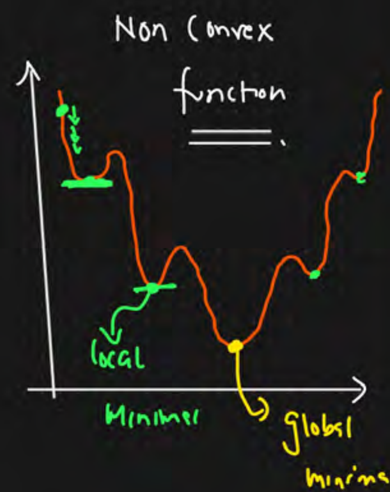
⇓

Convex function

MSE↑

Penalize the cost fn





Advantage

① Equation is differentiable

② It has only one local or global minima.

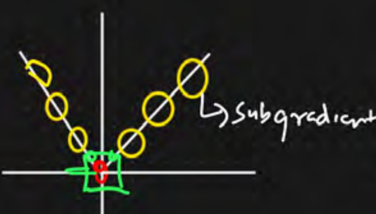
Disadvantage

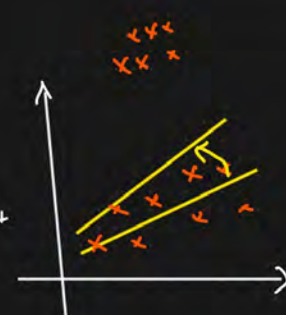
① Not Robust to outliers

② Units gets changed

② Mean Absolute Error

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





Advantage

① Robust to outliers

Disadvantage

① Convergence it will take time

(4) Unit are same

(5) Root Mean Squared Error [Assignment].

$$RMSE = \sqrt{MSE}$$

(x) Ridge, Lasso And Elastic Regression

$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x)_i)^2 \quad [MSE]$$

(1) Ridge Regression [L2 Regularization] → Reduce Overfitting ←

DATASET

TRAIN → Model TRAIN

TEST → Model Validate or Test

Overfitting = TRAIN → Acc ↑↑ ⇒ 93%

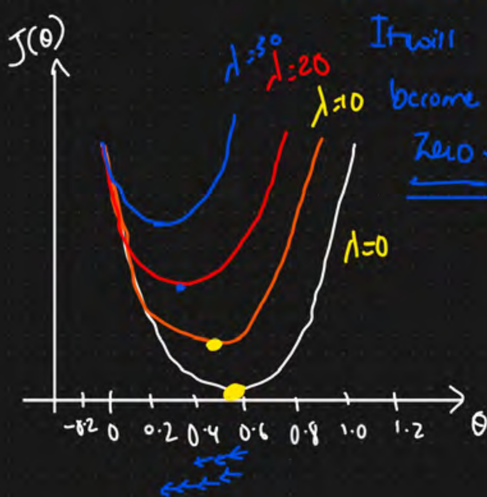
TEST → Acc ↓↓ ⇒ 50%

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

⇓ 0

Hyperparameter

$\boxed{\lambda = 1, 2, 3, 4, 5, 6, 7}$



λ ↑↑ θ ↓↓ It will become Zero.

② Lasso Regression (L_1 Regularization) → Feature Selection

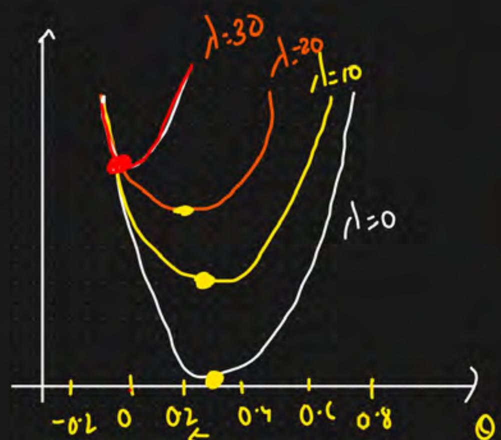
$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \boxed{\lambda \sum_{i=1}^{n}|\text{slope}|}$$

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

$= 0.52 + 0.65x_1 + 1.5x_2 + 0.2x_3.$

The least important feature is x_3 because the coefficient is less.

$\lambda \Rightarrow$ {Experimentation}.



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

$= 0.52 + 0.65x_1 + 1.5x_2 + 0.2x_3.$

$0.52 + 0.50x_1 + 1.3x_2 + 0.1x_3$

$\boxed{h_\theta(x) = 0.52 + 0.40x_1 + 1.1x_2} + \boxed{0 * \cancel{x_3}}$

Coefficient 0

③ Elastic Net

- → Reduce Overfitting → Ridge
- → Feature Selection → Lasso

$$\text{Cost fn} : \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{\text{MSE}} + \underbrace{\lambda_1 \sum_{i=1}^{n}(\text{slope})^2}_{\text{Reduce Overfitting}} + \underbrace{\lambda_2 \sum_{i=1}^{n}|\text{slope}|}_{\text{Feature Selection}}$$

MSE + Reduce Overfitting + Feature Selection.

THANK - YOU