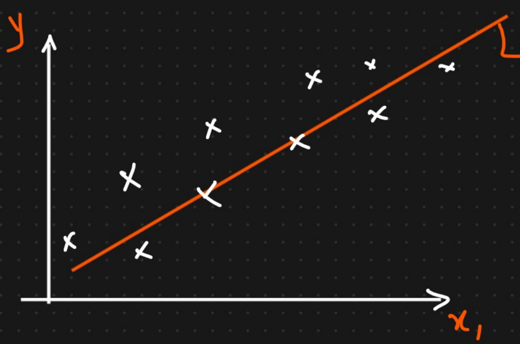
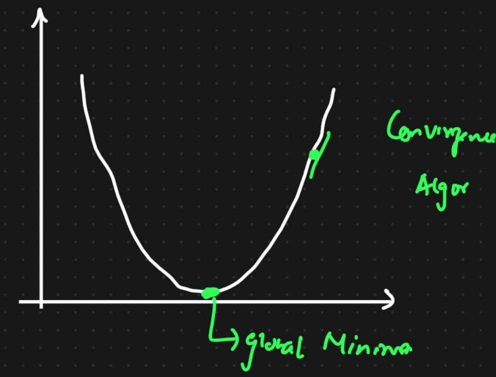
Ridge, Lasso And ElasticNet Regression



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

↓

Independent feature



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

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

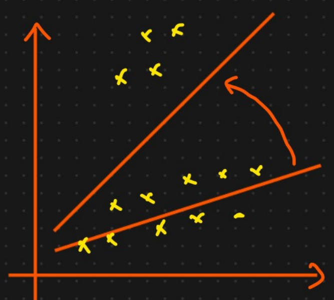
$(a+b)^2 = a^2 + 2ab + b^2$

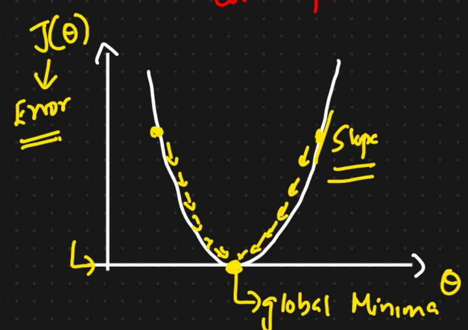
① Mean Squared Error (MSE)

$$\boxed{MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$ ⟶ Quadratic Equation

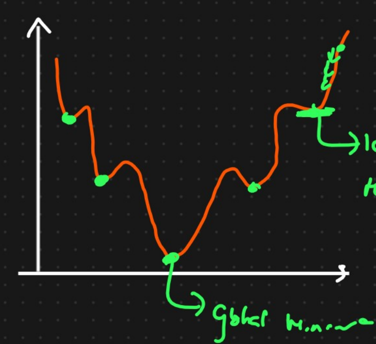
⇓

Convex function





Non Convex function



Advantage

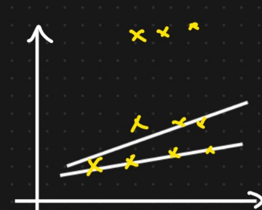
① Equation is differentiable

② It has only one local or global Minima

Disadvantage

① Not Robust to Outliers

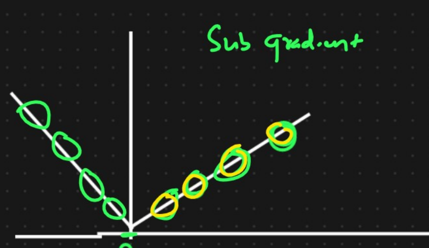
② Error is not in the same unit



② Mean Absolute Error (MAE)

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

Error is in same unit.



Advantage

1. Robust to outlier
2. It will be in the same unit

Disadvantage

1. Convergence takes more time

③ **Root Mean Squared Error**

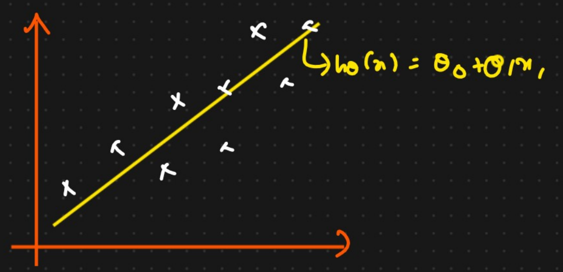
$$RMSE = \sqrt{MSE}$$

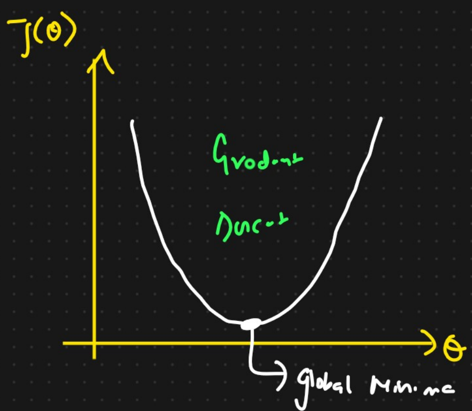
$$\boxed{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}}$$

$\Rightarrow$ Assignment [Advantages And Disadvantages].

Ridge, Lasso And ElasticNet

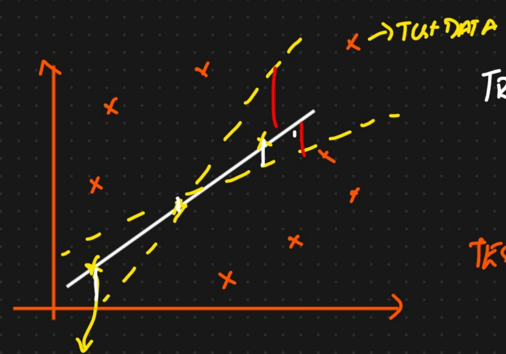
Linear Regression





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

① **Ridge Regression (L2 Regularization)** $\longrightarrow$ Reduce Overfitting



TRAINING Error = 0 Accuracy $R^2 \uparrow\uparrow \Rightarrow 100\%$

TEST Error $\uparrow\uparrow$ Accuracy $\downarrow\downarrow$

$$\boxed{\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \boxed{\lambda\sum_{i=1}^{n}(\text{slope})^2}} \longrightarrow \text{L2 Regularization}$$

$\Downarrow$ MSE

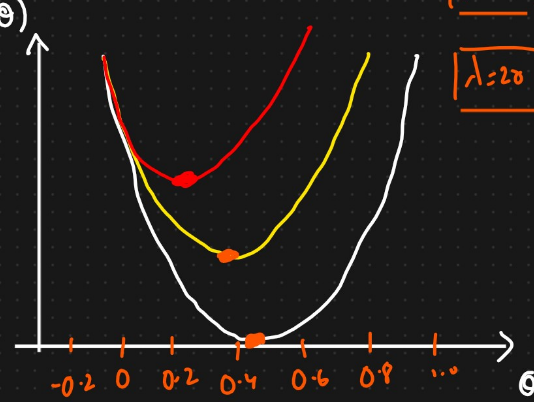
Hyperparameter

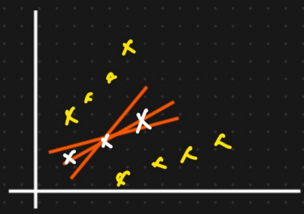
Error is getting adding

$\lambda=0$

$\lambda=10$

$\lambda=20$





$\lambda = 1$

$\text{Cost fn} = 0 + 1\left[(\theta_1)^2\right]$ ⇐ penalizing the cost fn

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

$[x_1, x_2, x_3]$

$[\theta_1, \theta_2, \theta_3]$

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

② Lasso Regression (L1 Regularization) $\longrightarrow$ Feature Selection

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

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

feature selection

$x_1 \rightarrow 0.65$

$$= 0.52 + 0.65x_1 + 1.5x_2 + \boxed{0.2}x_3$$

$\Downarrow$

$$0.52 + 0.54x_1 + 1.2x_2 + \boxed{0 \times x_3}$$

$\Downarrow$ 0

x_3 $\Downarrow$ removed $\Downarrow$ Correlation

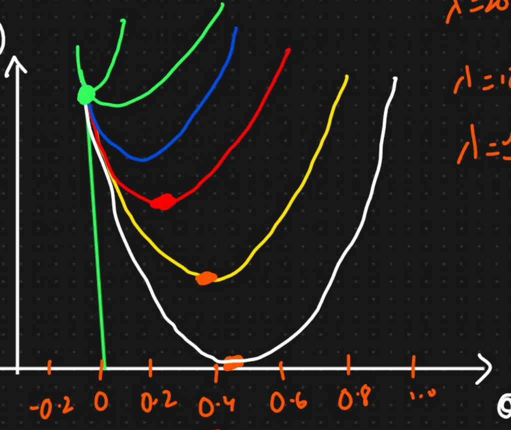
$$h_\theta(x) = \boxed{0.52 + 0.54x_1 + 1.2x_2}$$

$\lambda=30$

$\lambda=20$

$\lambda=10$

$\lambda=1$



③ ElasticNet Regression

- → Reducing overfitting → Ridge
- → Feature Selection → Lasso

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

⇓ MSE + ⇓ Reduce overfitting + ⇓ Feature Selection

λ_1, λ_2 {Hyperparameter}.

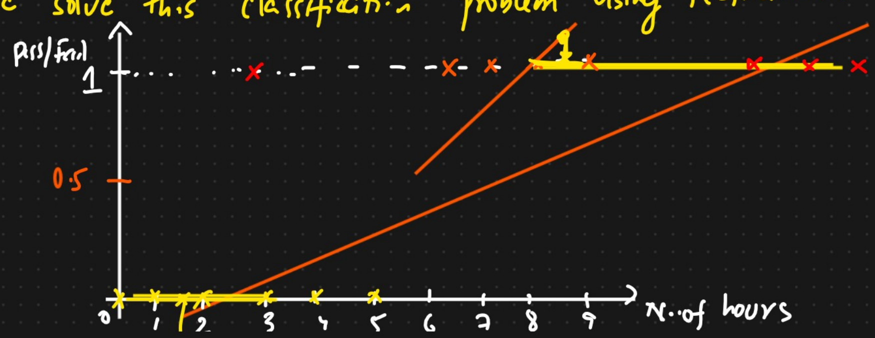
④ Logistic Regression [Classification → Binary classification]

[UPSC].

DATASET.

I/p feature: No. of study hours	Dependent feature: Pass/Fail
1	0
2	0
3	0
4	0
5	0
6	1
7	1
8	1
9	1

Can we solve this classification problem using Regression?



$$\boxed{\frac{1}{1+e^{-z}}}$$

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

⇓

Sigmoid function.

① Best fit line changes because of outliers ⟶ prediction goes wrong.

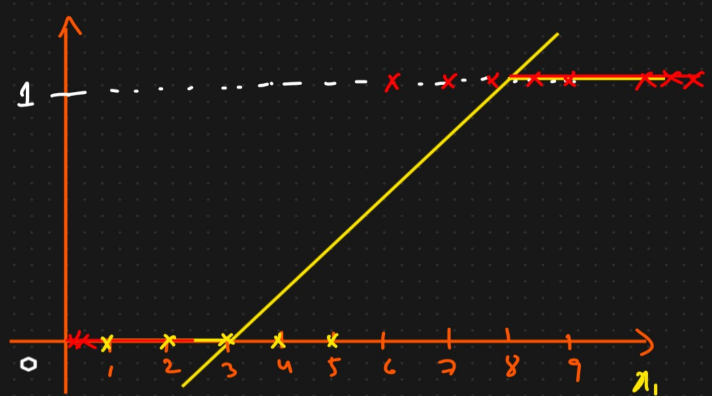
② Outcome comes >1 and <0 [Real output should be between 0 and 1]

To solve this problem, we use Logistic Regression

⇓

$\boxed{0 \text{ to } 1}$ ⇒ Squashing Technique

How Logistic Regression Solve Classification Problem



$z = \boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$ ⟶ Best fit line

⇓

[Sigmoid Activation function].

⇓

0 to 1

$$\boxed{\sigma = \frac{1}{1+e^{-z}}} \Rightarrow 0 \text{ to } 1.$$

z

⇑

$$\boxed{h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1)}$$

$$\boxed{h_\theta(x) = \frac{1}{1+e^{-z}}} \Rightarrow z = \theta_0 + \theta_1 x_1 \Rightarrow h_\theta(x) = \boxed{\frac{1}{1+e^{-(\theta_0+\theta_1 x_1)}}}$$

Linear Regression Cost fn

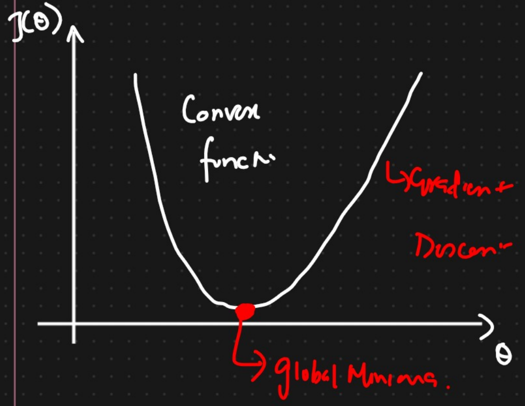
$$J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - h_\theta(x)_i)^2$$

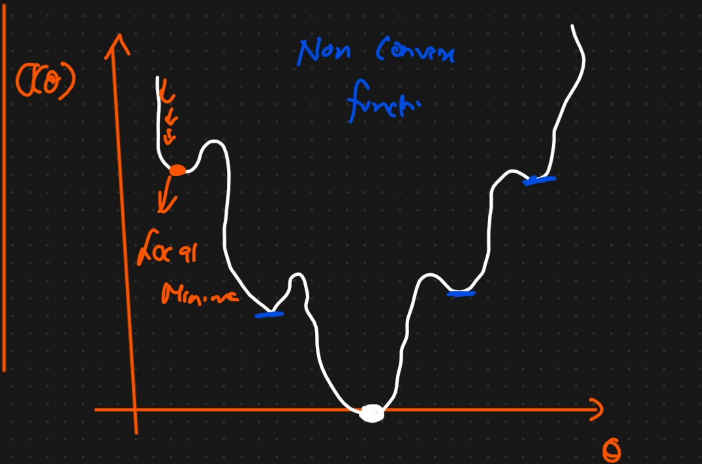
$$h_\theta(x) = \theta_0 + \theta_1 x_1$$

Logistic Regression

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

$$h_\theta(x) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\theta_0+\theta_1 x_1)}}$$





Log Loss Cost fn

$$J(\theta_0,\theta_1) = \begin{cases} -\log h_\theta(x) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0. \end{cases}$$

$$h_\theta(x) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\theta_0+\theta_1 x_1)}}$$

$$\boxed{J(\theta_0,\theta_1) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))} \Rightarrow \text{Convex function}$$

↳ Logistic Regression Cost fn.

Final Aim

Minimize Cost fn $J(\theta_0,\theta_1)$ by changing θ_0, & θ_1

Convergence Algorithm

Repeat

{

$$\theta_j := \theta_j - \alpha \frac{\partial(J(\theta_0,\theta_1))}{\partial\theta}$$

}

$\theta_1\ \theta_2\ \theta_3\ \theta_j$

Logistic Regression With Regularization Parameters

$$\boxed{\text{Cost fn} = -y \log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))}$$

$$h_\theta(x) = \frac{1}{1+e^{-(\theta_0+\theta x_1)}}$$

$$J(\theta_0,\theta_1) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$

L2 Regularization ⇒ Reduce Overfitting

⇓

$$J(\theta_0,\theta_1) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x)) + \boxed{\lambda \sum_{i=1}^{n} (\text{slope})^2}$$

$$J(\theta_0,\theta_1) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x)) + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|}$$

→ L1 Regularization

⇓

feature selection.

$$J(\theta_0,\theta_1) = -y\log(h_\theta(x)) - (1-y)\log(1-h_\theta(x)) + \boxed{\lambda_1 \sum_{i=1}^{n} (\text{slope})^2 + \lambda_2 \sum_{i=1}^{n} |\text{slope}|}$$

⇓

ElasticNet

C and λ

$$\boxed{\lambda = \frac{1}{C}} \qquad \boxed{C = \frac{1}{\lambda}}$$