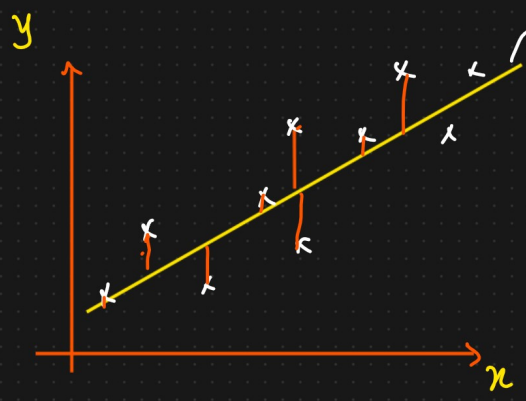
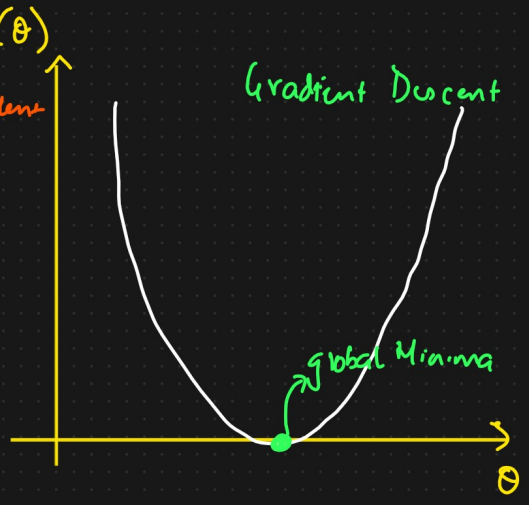
Ridge Regression, Lasso Regression, Elasticnet Regression

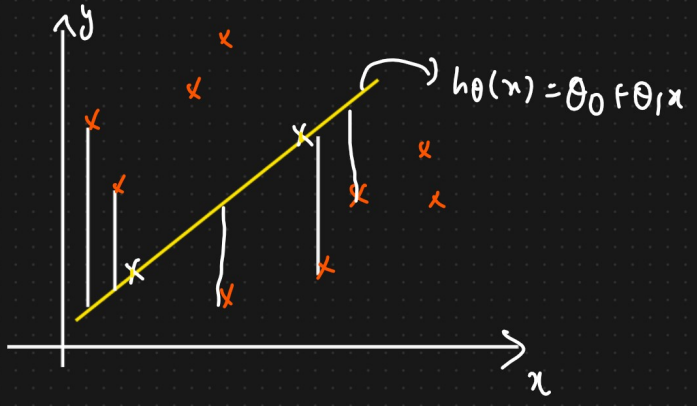
Linear Regression





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

① Ridge Regression (L2 Regularization) → Reducing Overfitting.



DATASET
- TRAIN → TRAIN OUR MODEL
- TEST → TEST OUR MODEL

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

$\downarrow$

0.

Overfitting

Low Bias

$\downarrow$

TRAIN DATASET → ACC ↑↑

TEST DATASET → ACC ↓↓

$\downarrow$

High Variance

Ridge Regression

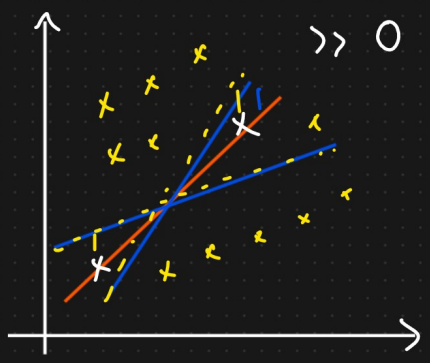
hyperparameter

Lets $\lambda = 1$

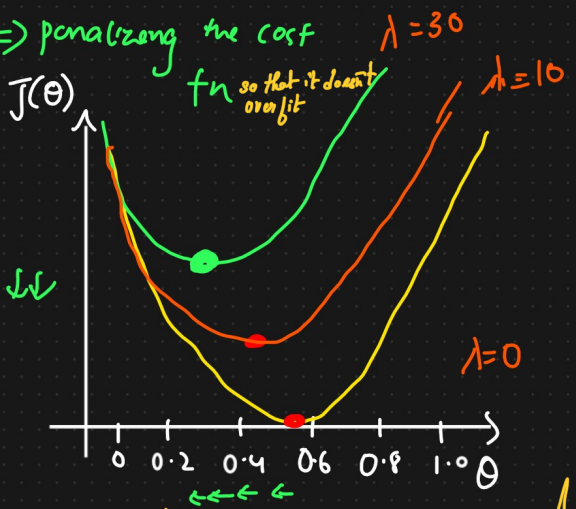
$$J(\theta)\ \text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)_i\right)^2 + \lambda \sum_{j=1}^{n}(\text{slope})^2$$

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

$$= 0 + (1)\left[(\theta_1)^2\right] \Rightarrow \text{penalizing the cost fn so that it doesn't overfit}$$



$\lambda \uparrow$ Slope $\downarrow$



with inc of λ, slope is decreasing

if is hyper parameter and we have to find the right value of this

L2 norm.

Simple Linear Regression

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

$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)_i\right)^2 + \lambda\left[(\theta_1)^2\right]$$

Multiple Linear Regression

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

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

② Lasso Regression (L_1 Regularization) → Feature Selection

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

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

unit change in x_1 → 0.65 change in y

$$= 0.52 + 0.65 x_1 + 0.72 x_2 + 0.12 x_3$$

x_3: very small value so x_3 has near 0 impact on y ("Corelation") we can remove this

⇓ Lasso Regression

$\lambda = 1$

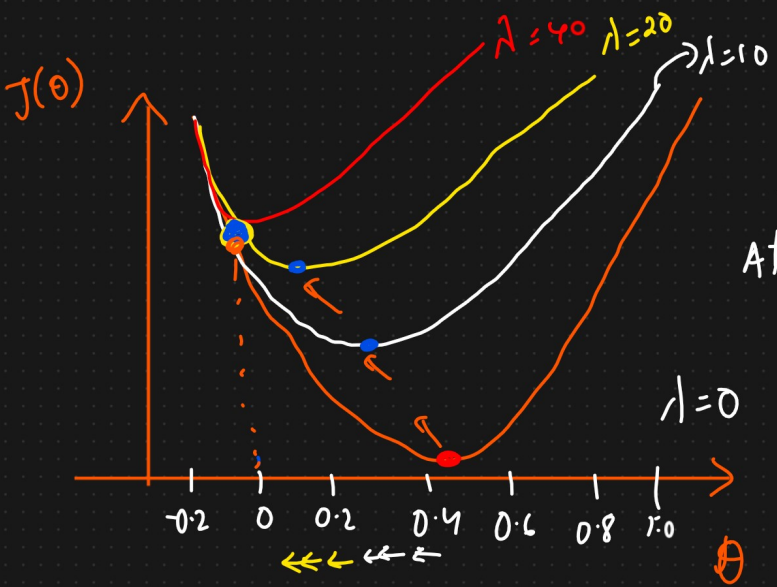
$$h_\theta(x) = 0.52 + 0.65 x_1 + 0.72 x_2 + 0.12 x_3$$

Lasso Regression.

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

⇓

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



we are changing the λ value

λ↑↑ slope↓↓

At one point $\theta \approx 0$

At one point it will be zero

The feature which is not much corelated will be zero

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

$$h_\theta(x) = 0.52 + 0.65x_1 + 0.72x_3$$

③ Elastic Net Regression

→ ① Reduce Overfitting } Ridge
→ ② Feature Selection } Lasso

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

MSE + Reduce Overfitting + Feature Selection

λ_1, λ_2 {Hyperparameter Tuning}