Elastic Net

→ combination of L1 and L2 Regularization.

$$\text{cost } f^n = \frac{1}{m} \sum_{i=1}^{m} \{h_\theta(x)^{(i)} - y^{(i)}\}^2 + \lambda \sum_{i=1}^{m} (\text{Slope})^2 + \lambda \sum_{i=1}^{m} |\text{Slope}|$$

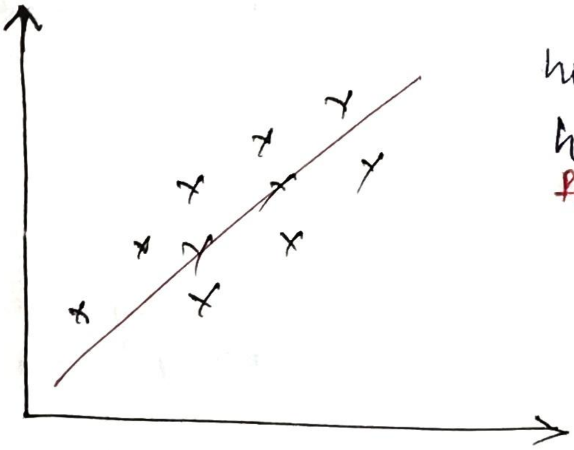
can be changed to MAE, RMSE, MSE

Ridge

Lasso

10/15/2022 ML-Day-3

Linear Regression



$h_\theta(x) = \theta_0 + \theta_1 x$ (Simple Linear Regression)

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

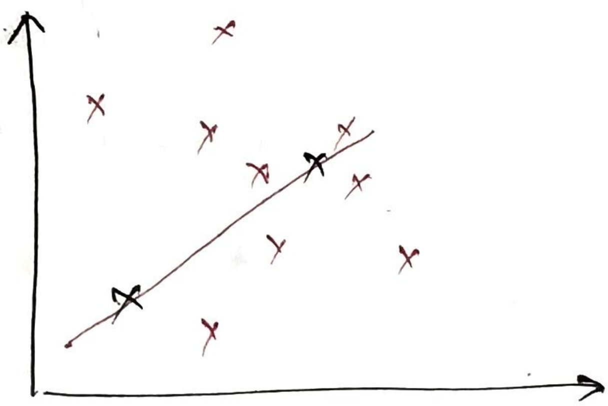
~~Polynomial~~ (Multiple Linear Regression)

$$\boxed{\text{cost function} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2}$$

Gradient Descent

Ridge Regression (L2 Regularization / L2 Norm)

- Used to reduce overfitting.



$\boxed{\text{cost function} = 0}$

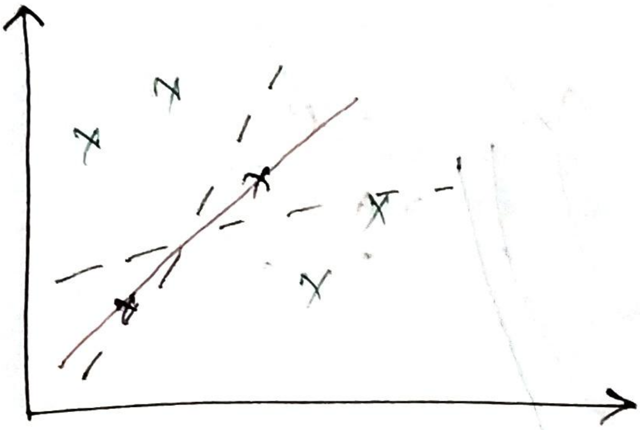
Training Data → Low Bias

Test Data → Low/High Variance

→ If the test data is near to best fit line then performance will be good. [Low Variance]

→ If the test data is far to best fit line then performance will be bad. [High Variance]

AIM: To reduce overfitting.



→ We create multiple lines to ~~reduce~~ ones improve performance of test data.

cost function

$$\boxed{\text{cost function} = \frac{1}{m}\sum_{i=1}^{m}\{h_\theta(x)^{(i)} - y^{(i)}\}^2 + \lambda(\text{slope})^2}$$

λ = hyperparameter

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

slope = θ_1

If multiple features present, then

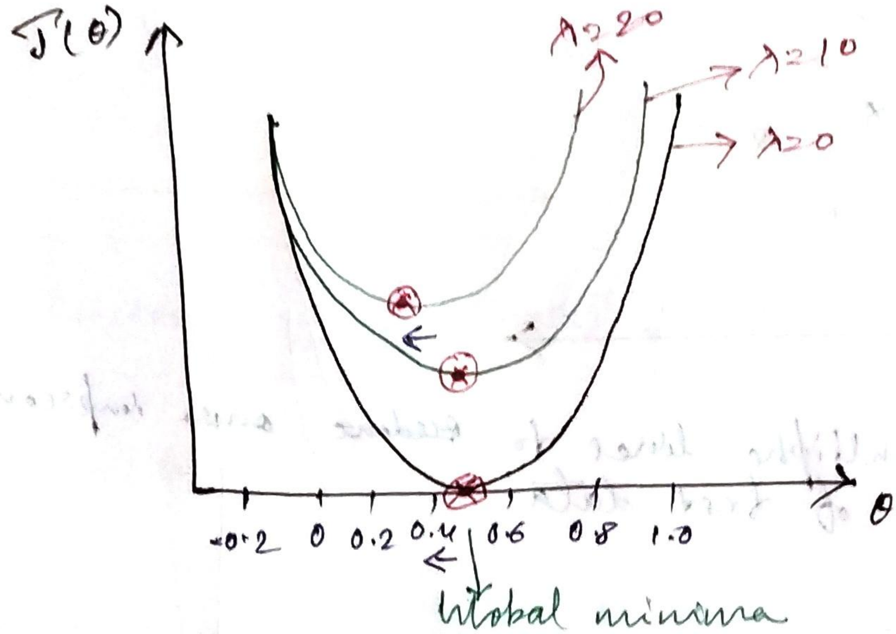
$$\boxed{(\text{slope})^2 = \sum_{i=1}^{m}(\text{slope})^2}$$

slope = slope of different lines.

if $\lambda = 0$

→ cost function is same as linear regression's cost function.

Relationship b/w slope and λ



→ Global minima gets shifted towards left with increase in λ.

$\lambda = +ve$, $\theta^2 = +ve$

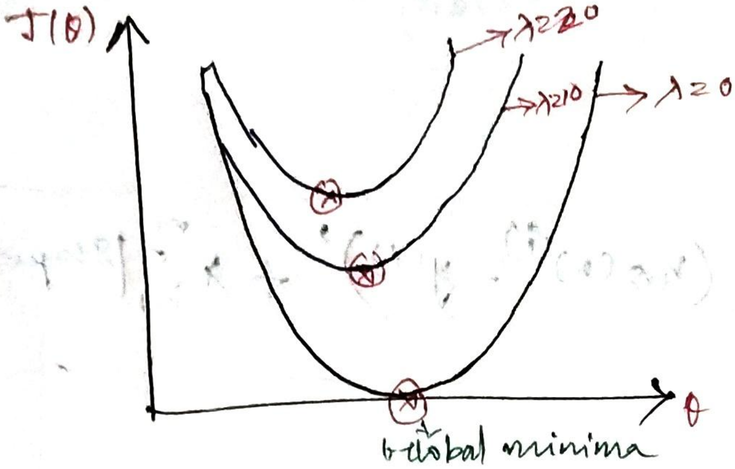
cost fn = 0 + (+ve) = 0 + ve ↓↓↓

→ change θ value to create another best fit line.

$$\boxed{\lambda \propto \frac{1}{\text{slope}}}$$

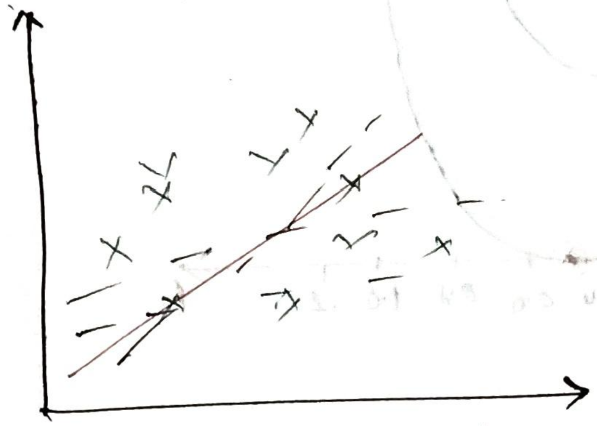
Inversely Proportional

λ = It is used to make sure that our line doesn't overfit.



$\lambda \uparrow \rightarrow \theta \downarrow$

$\lambda \uparrow \rightarrow$ Global minima shifted towards left



Keep changing line until overfitting gets reduced.

θ value never becomes zero

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

$$= \theta_0 + 0.95 x_1 + 0.82 x_2 + 1.5 x_3$$

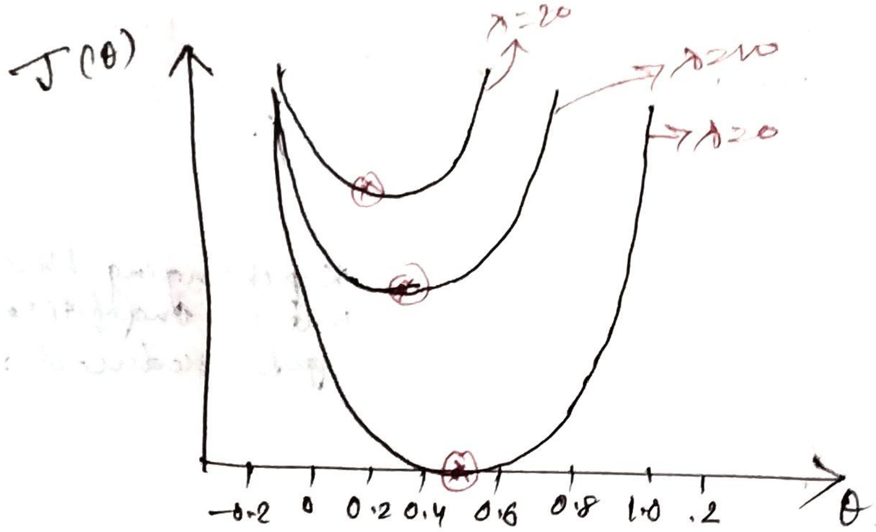
If it is zero then x_3 will get deleted.

Lasso Regression (L1 Regularization / L1 Norm)

→ It is used to reduce the features. It helps in feature selection.

Cost function

$$\boxed{\text{cost } f^n = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda \sum_{i=1}^{m} |\text{slope}|}$$



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

$= \boxed{\theta_0 + 0.54x_1 + 0.23x_2} + \boxed{0.10x_3}$

Imp features — least correlated

[If the dataset has outliers → Use Ridge Regression]