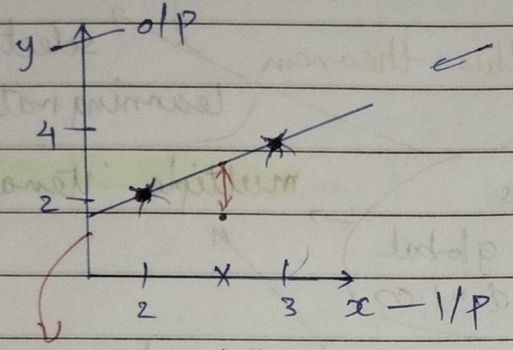
Ridge and Lasso Regression

→ Ridge regression (also called Tikhonov regularization) is a regularized version of linear regression



← This is my Training set

x	y
1	2
3	3

⇒ linear regression

$x \rightarrow \square \rightarrow y$

• ⇒ predicted value

x ⇒ actual value

error = (f − f̂)

error = 0 − 0

error = 0

1. Overfitting

Train Accuracy = 90%

Test Accuracy = 70% ↓

→ Low Bias & High Variance

2. Underfitting

Train Accuracy = 60%

Test Accuracy = 62% } ↓

High Bias & High Variance

3. Generalized model

low bias — Train Accuracy = 90%

low variance — Test Accuracy = 89%

★ L1 Regularization — Lasso Regression (Cost function)

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

→ Overfitting, feature selection

★ L2 Regularization — Ridge Regression (Cost function)

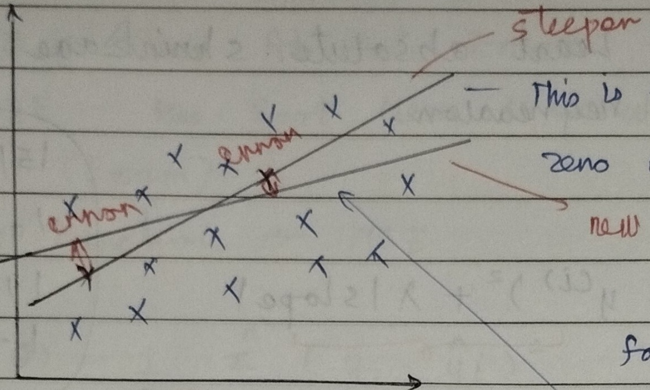
$$-\frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda\,(slope)^2$$

→ Overfitting

$\frac{1}{2m} \sum_{i=1}^{m} (y - \hat{y})^2 + \lambda (slope)^2$

Ridge Regression - L2 Regularization

$x \rightarrow$ new data points



— This is my current best fit line with zero error

new best line

for eg if our $(h_\theta(x) - y)^2 = 0$

and $\lambda = 1$ (for eg)

slope = 2

Cost function or Residual Error

$= \cancel{\frac{1}{2m} \sum_{i=1}^{m} (h_{\theta}(x) - (y)}$

$\rightarrow \lambda \sum \theta_j^2$

$= \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^{(i)} - y^{(i)})^2 + \lambda \underbrace{(slope)^2}_{\text{This term is called penalizing parameter}}$ — m

↓

formula

$= 0 + 1(2)^2$

$= 0 + 4$

$= 4$:- we need to reduce this thing

$m = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{(X_i - \bar{X})^2}$

↓

for simple linear regression

this will be just a small value

↓

{small value} + $1(1.3)^2$

$\Rightarrow 1.69 \Rightarrow$ here we have reduced the value

↓

and this will not overfit the training dataset and works the same for new or unseen data

in the first the slope was more steeper now will create another slope(m) which will reduce the value of current slope based on how flatten the line is for eg now value is 1.3

↳ this will become our new best fit line which covers most of the data points

→ θ_1 (or m_1) = coefficient or weight or parameter

→ $\theta_1 x_1$ or $m_1 x_1$ = feature term

Lasso Regression - L1 Regularization

~~Lasso is also called~~

Lasso full form :- Least absolute shrinkage and selection operator regression

$|5| = 5$
$|-5| = 5$
$|10| = 10$
$|-10| = 10$

Cost function

$$\frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x)^{(i)} - y^{(i)})^2 + \underbrace{\lambda |slope|}$$

$$\frac{1}{2m} \sum_{i=1}^{m} (y - \hat{y})^2 + \lambda |slope|$$

This term is called Penalizing parameter

→ Prevents Overfitting & help's in feature selection by neglecting the value of ~~theta~~ θ

→ Feature selection :- If the Feature term has small values those values will be neglected

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

has ⇒ small values small values

this will be neglected or removed

→ for not participating to help to create best fit line because of having small values → they neglects themselves using this equation or penalizing parameter $\lambda |slope|$