Linear Regression

↓

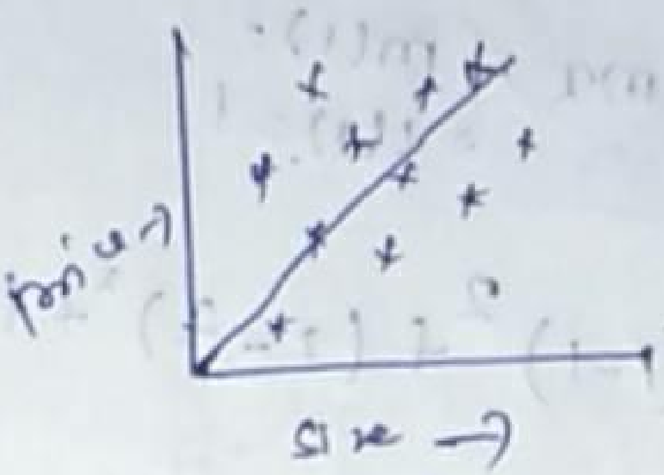
Predict by the use of known

Parameter

$y = mx + c$ → line equation use

this equation to find best bit line

m → slope, c → intercept



$s = 0$

$p = 0$

slope = 0

Because passes value by origin.

Sum of all errors.

Select best bit line

↓ through

minimal error

↓

Cost function

distance b/w real p & predicted

$$y = \frac{1}{2m}(\hat{y} - y_i)^2$$

$\hat{y} = mx + c$

Cost function : $\frac{1}{2m}(\hat{y} - y_i)^2$

$= \hat{y} = mx + c$

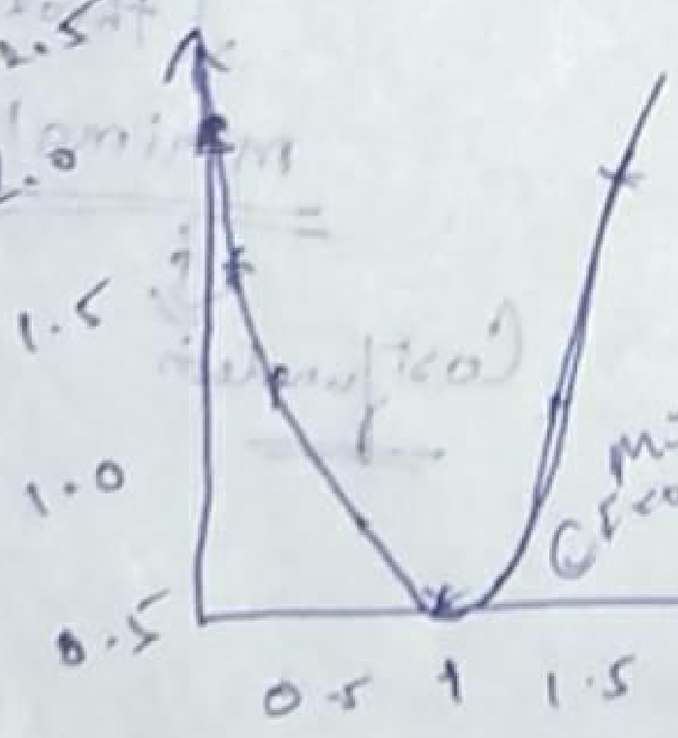
$y = mx = m(1)$

$= 1(1) = 1$

$= \frac{1}{2(3)} = \left((1-1)^2 + (2-2)^2 + (3-3)^2\right)$

$= 1/6\,(0) = 0$

Cost function $= 0$



$m = 0.5$

$\hat{y} = mx$

$= 0.5(1) = 0.5$

$= 0.5(2) = 1.0$

$0.5(3) = 1.5$

$y = \frac{1}{2m}(\hat{y} - y)^2$

$= \frac{1}{2(3)}$

$(0.5-1)^2 +$

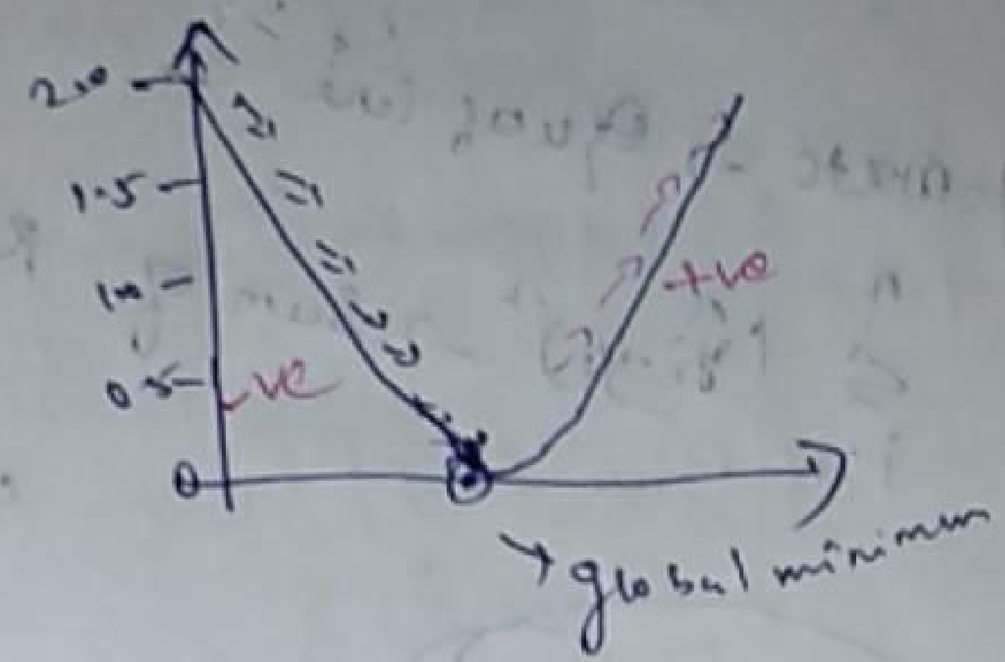
$(1.0-2)^2 +$

$(1.5-3)^2$

$= 0.25 + 1 + 2.25$

$= 0.5$

Gradient descent:



Converging theorem → help to reach global minimum.

$$m := m - \left(\frac{dm}{dm}\right) * \alpha$$

↓ learning rate

Subtract the slope and prod of α.

m = m - (-ve) * smaller

m = m + (+ve) * sR.

Ridge and lasso Regression

$y = mx + c$ → Eqn of line

$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2$ → Sum of Residuals

↓

cost function

Generalized model → high variance – low variance

Ridge:

⇒ Reduce cost function

⇒ Reduce overfit

Ridge add one more parameter in cost function

$$(y - \hat{y})^2 + \lambda * (\text{slope})^2$$

Steep slope → punish increase in x
huge move in y

$\lambda \rightarrow 0$ to any (+ve)

$= 0 + 1 (1.3)^2$

$= 0 + 1.69 =$ (1.69)

Penalize → along with (+ ve)
$\lambda * (\text{slope})^2$

highslope - lowslope.

Regression → selected best fit line value.

Slope line → close to zero
not exactly zero.

Lasso: (magnitude of slope)

↓ help of feature selection

$y = m_1 x_1 + m_2 x_2 + m_3 x_3 +$
$m_4 x_4 + \dots c.$

$\lambda + |m_1 + m_2 + m_3 + m_4|$

Ridge → towards zero not exactly zero

Lasso → towards to zero.

Lasso → helps to feature - selection

↓ how

$m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4$ ...

Slope is slow that feature will be removed.

that removable feature not helpto prediction

when use magnitude moving towards zero or atleast one time reach zero