Module : Regularization- Ridge & Lasso

Instructor: Dr. Darshan Ingle

HI Y'ALL !!!

STARTING SOON

upGrad

Regularization Overview!

$$y = c + m_1 x_1 + m_2 x_2 + m_3 x_3$$

$$y = 15 + 1.2 x_1 + 20 x_2 + 39 x_3$$

Note! $x_3 \rightarrow$ v.imp to determine y.

$x_3 \rightarrow$ Highest Co-effi., $x_1 \rightarrow$ least imp.

$\because$ Val. of x_3 is v. High

$\therefore$ it becomes a

Computationally expensive.

$\therefore$ Model $\rightarrow$ Complex.

$2 \times 3 = 6$.

$6 \times 9 = 54$

$23 \times 47 = ?$

$992 \times 496 = ?$

ML

$\downarrow$

v. small.

$\rightarrow 0.0012 \times 0.356 = ?$

To reduce the complexity, ∴ we go for Regularization.

<u>OLD</u>: $y = 15 + 1.2x_1 + 20x_2 + 39x_3$

$\Downarrow$ Regularization

$y = 0.9 + 0.7x_1 + 2x_2 + 5x_3$

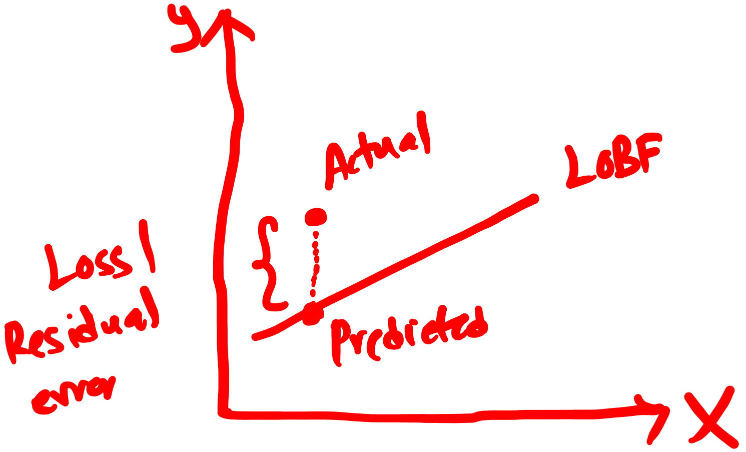
∴ this beams a computationally easy task.

Reg. Types :

- Ridge
- Lasso
- ElasticNet

① **Ridge**: (L2 Regularization Technique)

$$\text{Ridge Regression} = \underline{\underline{\text{Loss}}} + \underbrace{\alpha \cdot [w]^2}_{\text{Penalty}}$$



∴ Greater the ~~##~~ difference, greater the loss.

$$\alpha \cdot (W)^2$$

Penalty: is imposed to reduce the loss.
i.e. we try to compensate for the loss.

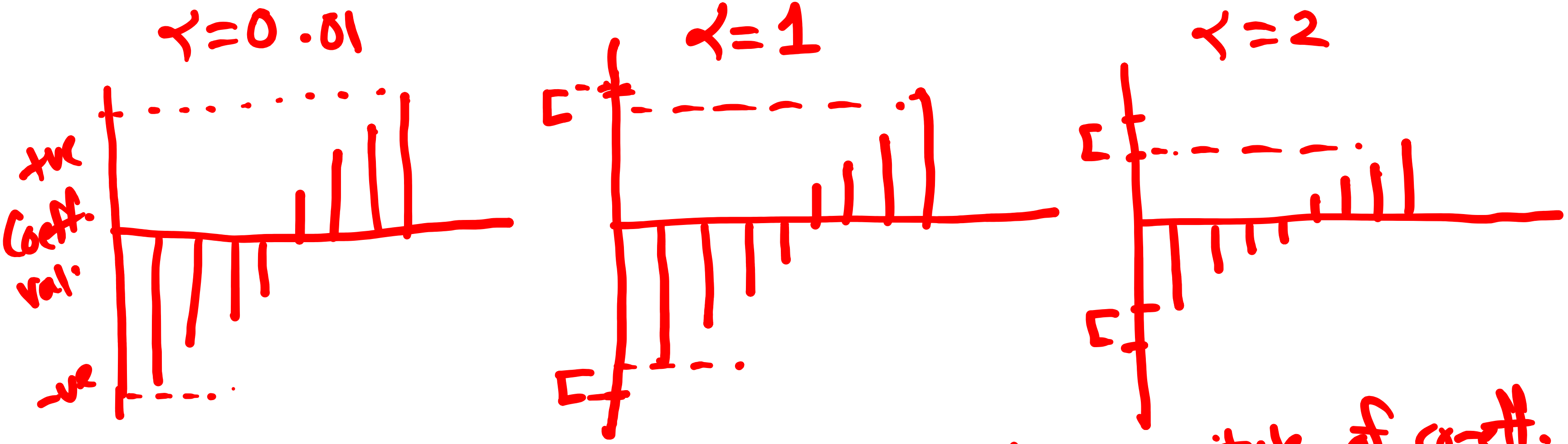
Adv: It helps us to scale down the magnitude of coefficients

α: It is H.P. It can be any constant value [0 to any +ve value]

$$(W)^2 = m_1^2 + m_2^2 + m_3^2 + \ldots\ldots + m_n^2$$

W - It is a vector of coeff. $(m_1, m_2, \ldots m_n \rightarrow$ Slope$)$.

$$y = 15 + 1.2x_1 + 20x_2 + 39x_3 \xrightarrow{R.R} y = 0.9 + 0.7x_1 + 2x_2 + 5x_3$$



α value modulates the magnitude of co-eff.

$\boxed{\text{Greater } \alpha \longleftrightarrow \text{Lesser val. of coeff.}}$

Ridge

Note: ① More computationally efficient ② Penalizes large nos. more.
③ More popular ④ Good when encounter colinearity

② Lasso : (L1 Reg. Tech).

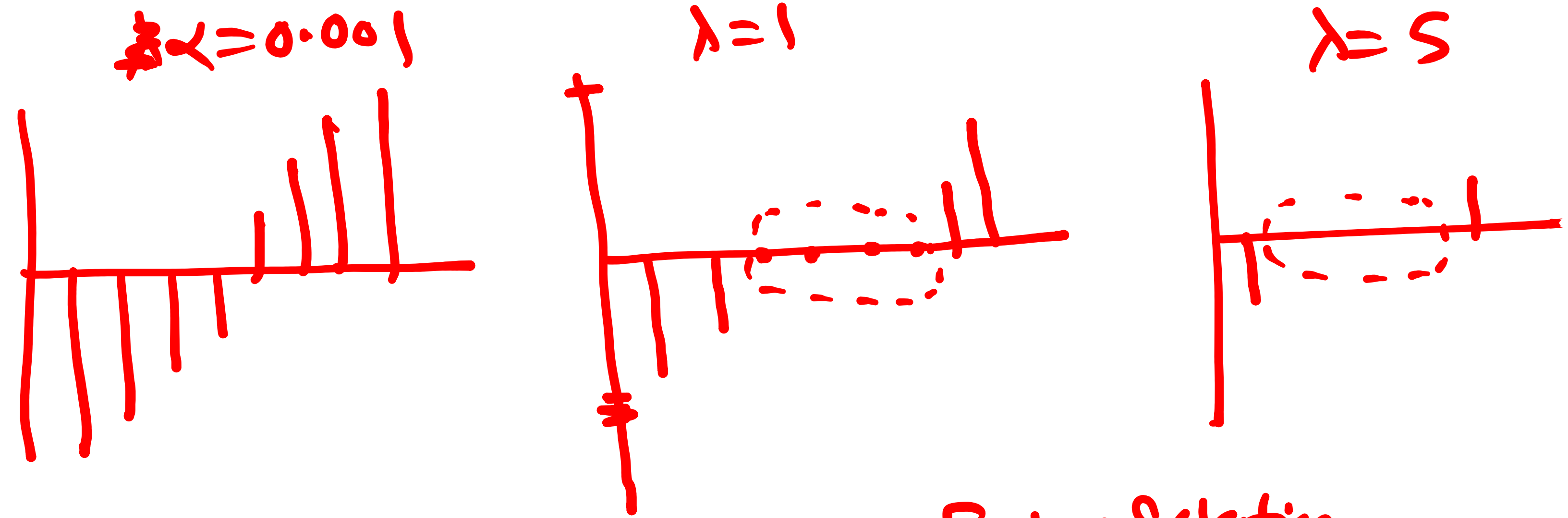
$$\text{Lasso Reg} = \text{Loss} + \underbrace{\alpha|W|}_{\text{Penalty}}$$

$$|W| = |m_1| + |m_2| + |m_3| + \cdots + |m_n|$$

eg: $y = 15 + 1.2x_1 + 20x_2 + 39x_3$

$\Downarrow$ L.R

$$\therefore \underline{\underline{y}} = 0.9 + \underbrace{0x_1}_{\times} + \underbrace{0x_2}_{\times} + 5x_3$$ [Feat. Sel.]



Note: (1) Lasso R. also acts as Feature Selection.

(2) Also, it is more robust to Outliers.

③ Elastic Net Regression: It is taking adv. of both R.R. & L.R.

Ridge $= \text{Loss} + \alpha (w)^2$

Lasso $= \text{Loss} + \alpha |w|$

$$\boxed{\text{Elastic Net Reg.} = \text{Loss} + \alpha (w)^2 + \alpha |w|}$$

When to use what?

1. Ridge R: Used in datasets which hv Multicolinearity
2. Lasso R: Used for feature Selection
3. Elastic Net Reg: → Adv of both. → Hybrid

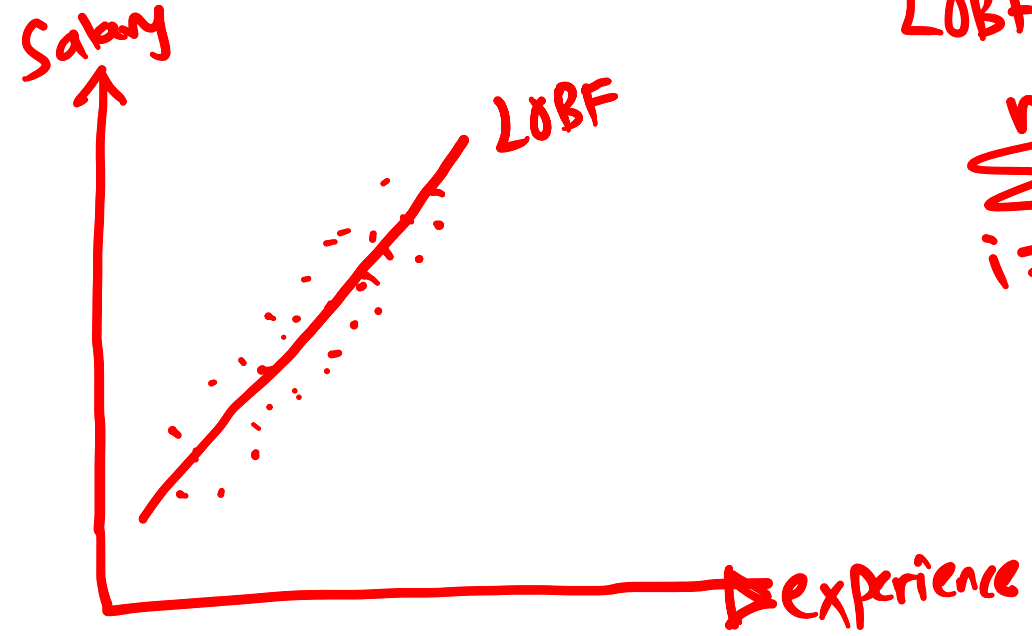
X

Overview Over.

Upcoming

P.T.O. for Math behind.

1. Ridge Reg:

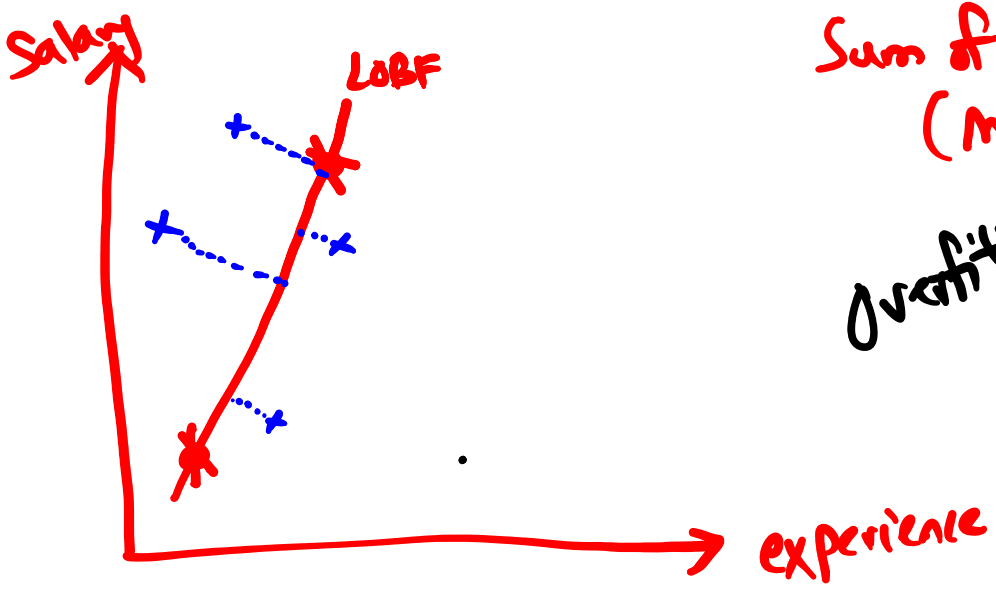


LOBF (least residual error)

$$\sum_{i=1}^{n} \left(y_{act} - y_{pred} \right)^2$$

$$\Downarrow$$

minimum



Sum of Residuals (minimize) $= \sum_{i=1}^{n} (y_{act} - y_{pred})^2$

Overfitting

- Training Data: Error = 0 (No Bias)
- Testing Data: Significant amt of error (High Variance)

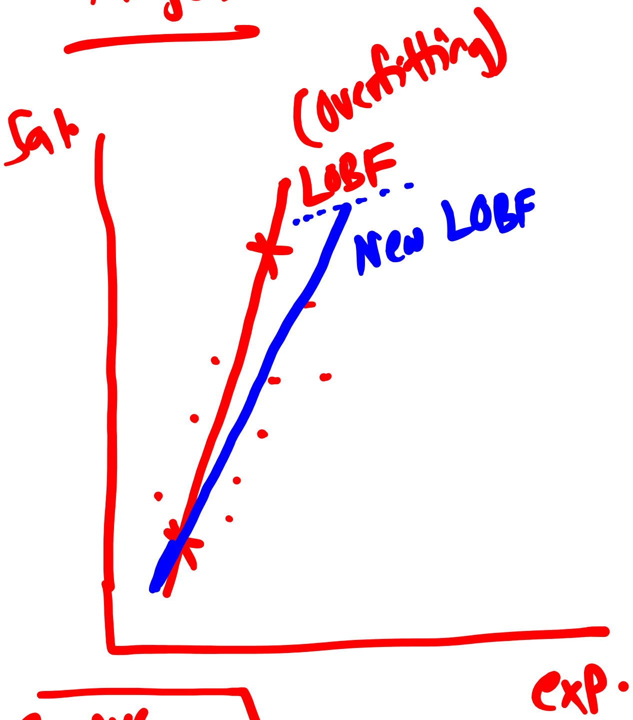
We expect: Model → (Low Bias & Low Variance)

Ridge:

$$= \text{Loss} + \alpha (W)^2$$

$$= \text{Loss} + \alpha (\text{Slope})^2$$

$$= \sum_{i=1}^{n} (y-\hat{y})^2 + \alpha \cdot (\text{Slope})^2$$



$\therefore$ $\boxed{\text{assume } \alpha = 1}$

Slope = 1.2	Slope = 1
$= 0 + 1(1.2)^2$	$= 0 + 1(1)$
$= 0 + 1.44$	$= 0 + 1$
$\boxed{\text{error} = 1.44}$	$\boxed{\text{Error} = 1}$

Tr. Er. ↑

Test. Er ↓

(overfit solved)

upGrad
#LifeKoKaroLift

Thank You!