# Ridge Regression part-4

## 5 key understanding under Ridge.

as we know that, in Ridge

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \boxed{\lambda\|w\|^2}$$

→ regularization term or Shrinkage Co-eff

$= \lambda(w_1^2+w_2^2 - - - + w_n^2)$

→ important features & strong shrinkage co-eff.

### Q1. How the Co-eff gets affected?

as $\lambda\uparrow$ ⟶ co-eff ↓ towards zero but never become zero.

### Q2. Higher values are impacted more?

Let's say for example

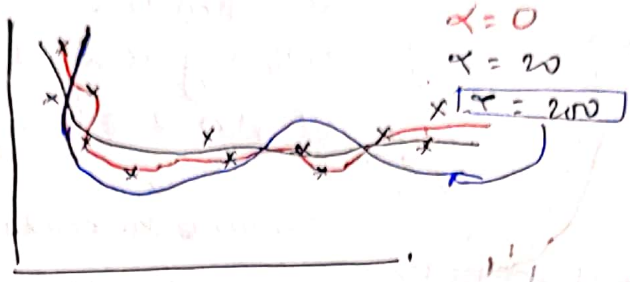
| $X_1$ | $X_2$ | $X_3$ | $Y$ |
|---|---|---|---|
| ↓ | ↓ | ↓ | |
| $w_0 = 1000$ | $w_2 = 10$ | $w_3 = 1$ | |

→ these are co-eff of features we got after applying regg

as we take $\lambda = 1 \to \infty$ increasing its value 1 to $\infty$, How, $w_0, w_2, w_3$ will get affected?

On increasing $\lambda$-value, the co-eff which is as bigger will decrease (or penalize) as fast comparison to smaller co-eff.

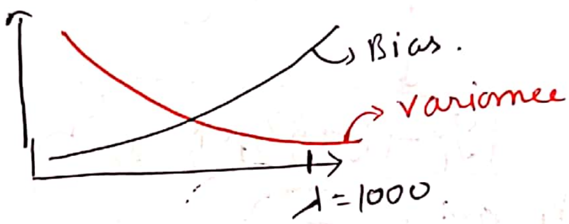
### Q3. Bias Variance Tradeoff




for $\alpha = 0$ → regg. line tries to touch each data point → ~~high~~ Low bias

$\alpha = 20$ → regg. line becoming smooth → ~~low~~ High bias → ~~High~~ Low variance.

$$\boxed{\alpha\uparrow \Rightarrow \text{bias}\uparrow\ \text{variance}\downarrow}$$



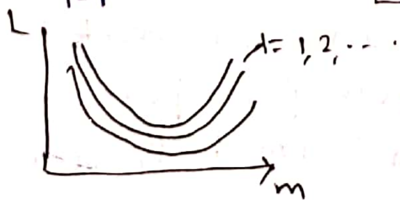

### Q4 Effect of regularization on loss function.

as we know that

$$L = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \boxed{\lambda\|w\|^2}$$

Let's assume b is constant only m is varying due to regularization then eqn of Loss would be

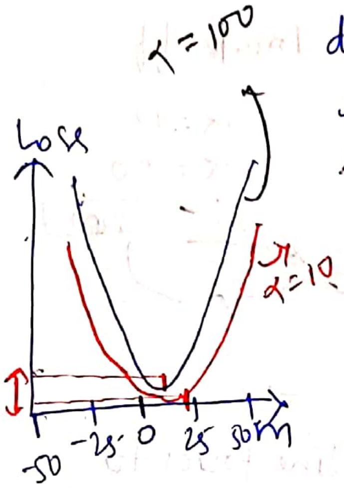
$$L = \sum_{i=1}^{n}(y_i - mx_i)^2 + \boxed{\lambda m^2}$$



$\lambda \uparrow \longrightarrow$ Co-eff $\downarrow \rightarrow$ tends 0

$\Downarrow$

due to which our loss function shifting towards origin 0 of (m, Loss)




$\Downarrow$

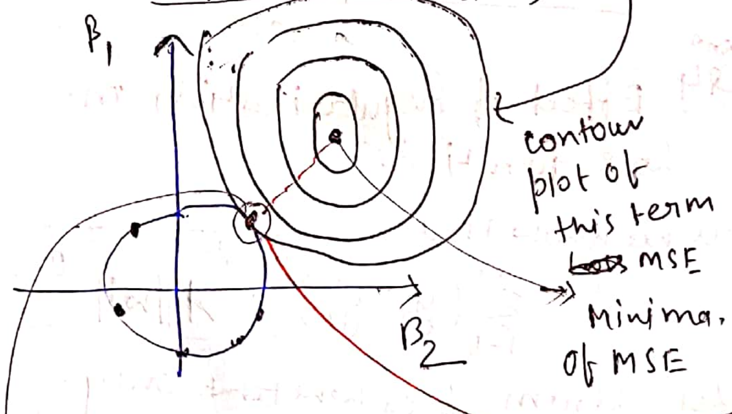
So, on increasing $\lambda$ loss function is ~~reading~~ increasing but m is decreasing.

* why Ridge is called Ridge.

as we know that in Ridge

$$Loss = MSE + \lambda ||w||^2$$

$$(y_i - (\beta_0 + \beta x_{i1} + \beta_2 x_{i2}))^2$$




$\lambda(\beta_1^2 + \beta^2)$

this is point from where MSE minima is nearest to point found on peripheri of Extra term $\lambda(\beta_1^2 + \beta_2^2)$

So, always ~~com~~ common point $(\beta_1, \beta_2)$ lie on peripheri (Ridge)

↳ ∴ its called as

**** Ridge estimation.

that's why ~~this is~~ called as ridge estimation