Date.....

Gradient Descent in Ridge Regression.....

Loss Function $\Rightarrow \boxed{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \Rightarrow (XW-Y)^T\cdot(XW-Y) + \lambda W^T W$

$[(a-b)^T = a^T - b^T]$

$\Rightarrow (X^T W^T - Y^T)(XW-Y) + \lambda W^T W \Rightarrow$ Multiplying with $\frac{1}{2}$ whole equation

$\Rightarrow \frac{1}{2}\left[X^T W^T \cdot XW - X^T W^T Y - Y^T XW + Y^T Y\right] + \frac{1}{2}\lambda W^T W$ (same)

$\Rightarrow \frac{dL}{dW} \Rightarrow \frac{1}{2}\left[X^T W^T \cdot XW - 2X^T W^T \cdot Y + Y^T Y\right] + \frac{1}{2}\lambda W^T W = 0$

$\Rightarrow$ Derivating the complete equation in terms of "W".

$\frac{dL}{dW} \Rightarrow \frac{1}{2}\left[2X^T XW - 2X^T Y + 0\right] + \frac{1}{2}\left[2\lambda W\right] = 0$

$\frac{dL}{dW} \Rightarrow \boxed{X^T XW - X^T Y + \lambda W = 0}$

$\frac{\Delta L}{\Delta W} \Rightarrow \boxed{X^T XW - X^T Y + \lambda W}$

$\boxed{W_{new} = W_{old} - \eta \frac{\Delta L}{\Delta W}}$ "Updating the value of "W"."

Points to be Remember:-

(i) As the Value of 'λ' ↑ Increases, the Coefficients Values starts shrinking.

(ii) Note:- The coef. values can never be equal to Zero '0'.

(iii) Higher Value of coef. starts higher shrinking in their value as 'λ' value increases. i.e; Higher Values of coef. have higher impact...

(iv) Both "bias" and "Variance" have impact on "λ".