Date.................

Ridge Regression ["n" dimensions]

$$\boxed{L = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \implies$$ For 'n' dimensions

$$L = (XW_i - Y_i)^T (XW_i - Y_i)$$

$$(\hat{Y} - Y) * (\hat{Y} - Y)$$

$L = (XW - Y)^T \cdot (XW - Y)$ {In Matrix Form}

Simplifying the Loss Function..

As per the Formula:-
$(a-b)^T = a^T - b^T$

$$\Rightarrow [XW^T - Y^T] * (XW - Y) + \underbrace{\lambda \|W\|^2}$$ → Ridge Regularization Addition

$$\Rightarrow [XW^T - Y^T] * (XW - Y) + \lambda WW^T$$

⊙ This $\lambda \|W\|^2$ is written as:-

$$\Rightarrow \{\lambda W_0^2 + \lambda W_1^2 + \lambda W_2^2 + \lambda W_n^2\}$$

$$\Rightarrow \lambda [W_0^2 + W_1^2 + W_2^2 + \dots W_n^2]$$

$$\Rightarrow \boxed{\lambda [W \cdot W^T]}$$

$$\Rightarrow [XW^T - Y^T] * (XW - Y) + \lambda W^T W$$

$$\Rightarrow [X^T W^T - Y^T] * (XW - Y) + \lambda W^T W$$

$$\Rightarrow X^T W^T \cdot XW - X^T W^T \cdot Y - Y^T \cdot XW + Y^T \cdot Y + \lambda W^T W$$

Same $Y^T Y$

$$\Rightarrow \boxed{X^T W^T \cdot XW - 2 X^T W^T \cdot Y + \lambda W^T W}$$ Loss Function....

Some Common derivations.

Scaler derivate	Vector derivative
$f(x) \Rightarrow \frac{df}{dx}$	$f(x) \Rightarrow df/dx$
$bx \longrightarrow b$	$x^T B \longrightarrow B$
$bx \longrightarrow b$	$x^T b \longrightarrow b$
$x^2 \longrightarrow 2x$	$x^T x \longrightarrow 2x$
$bx^2 \longrightarrow 2bx$	$x^T Bx \longrightarrow 2Bx$

~~$\frac{\partial L}{\partial W}$ = ...~~

$$\frac{\partial L}{\partial W} \Rightarrow \cancel{2} X^T XW - \cancel{2} X^T Y + 0 + \cancel{2} \lambda W = 0$$

$$\Rightarrow X^T XW + \lambda W = X^T Y$$

$$W(X^T X + \lambda I) = X^T Y$$

~~$W = X^T X$ ...~~

$$\boxed{W = (X^T X + \lambda I)^{-1} \cdot X^T Y}$$ → Identity Matrix....

Spiral

Date..................

So, finally Formula for Ridge Regression ['n' dimensions]

$$\Rightarrow \boxed{W = \left(X^T \cdot X + \lambda \cdot I\right)^{-1} \cdot \left(X^T \cdot Y\right)}$$