Two forms of Linear Regression

Lasso Reg

(L1 regularization)

↓

L1 norm

Ridge Reg

(L2 regularization)

(L2 norm)

Regularization Techniques

↓

(Prevent overfitting) → by adding penalty term

Ridge Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - - - -$$

Coefficient ⇒ relationship b/w input & o/p feature

Liner reg ⇒ $y = 0 + 5x_1 + 22x_2$

Ridge ⇒ $y = 0 + 0.2x_1 + 1.2x_2$

Reg

Main use

Shrinkage of coefficients

Ridge → Also called Tikhonov regularization reg

Also known as L2 regularization (addition of penalty term)

⇒ Penalty is called L2 norm

- Difference b/w lasso and Ridge regression

Lasso

① Automatic feature selection

② Penalty is L1 norm

ie

$$\underset{\text{norm}}{L1} = \alpha(|B_1| + |B_2|)$$

(absolute values)

Ridge

No automatic feature selection

Penalty is L2 norm

ie

$$\underset{\text{norm}}{L2} = \alpha\left((B_1)^2 + (B_2)^2\right)$$

(squared values)

③ Some coefficients may get shrunk to exactly zero.	No coefficient will shrink to exact zero
Use * When your data have irrelevant or unimportant features	* When your data have mostly important features in it

X_1	X_2	Y_{actual}	$Y_{predicted}$
1	2	3	3
2	1	4	3
3	3	7	6

Mathematical

① Perform linear regression & calculate coefficients.

Assume;

$B_0 = 0$, $B_1 = 1$, $B_2 = 1$

$$Y = B_0 + B_1 X_1 + B_2 X_2$$

$$Y = 0 + 1 \cdot X_1 + 1 \cdot X_2$$

$$\boxed{Y = X_1 + X_2} \longrightarrow$$ Linear reg completed.

② Calculate MSE

$$MSE = \frac{1}{n} \sum (Y_{actual} - Y_{predicted})^2$$

$$= \frac{1}{3} \left[(3-3)^2 + (4-3)^2 + (7-6)^2 \right]$$

$$= \frac{1}{3} \times 2$$

$$MSE = \frac{2}{3}$$

③ Calculate L2 norm (penalty)

$$\text{L2 norm} = \alpha\left((B_1)^2 + (B_2)^2\right) = 1\left(1^2 + 1^2\right)$$

$$= 2$$

$\alpha \Rightarrow$ hyperparameter (default = 1).

α should not be too big or not too small

Too Big ⇒ Increase penalty
It can lead imp features coeffcient to zero

Too small ⇒ Irrelevant feature will remain

④ Calculate Cost fn / objective fn

Cost fn = MSE + L1 norm

$= \frac{2}{3} + 2 = \frac{8}{3}$

Aim: Minimize the cost fn as much as possible.

→ Optimization Techniques :- Gradient Descent

Hyperparameters

① alpha (Most imp)
default : 1
Higher value of alpha ⇒ more coefficient will shrink

② fit_intercept ⇒ Already discussed
"

③ copy_X ⇒

④ solver ⇒ by default = "auto"
(which algorithm to use for computation)

⑤ positive ⇒ default : False
Set to True ⇒ Avoid negative coefficients.

⑥ max_iter ⇒ default : 1000
Higher the max_iter ⇒ More accurate soln but increase time for computation.