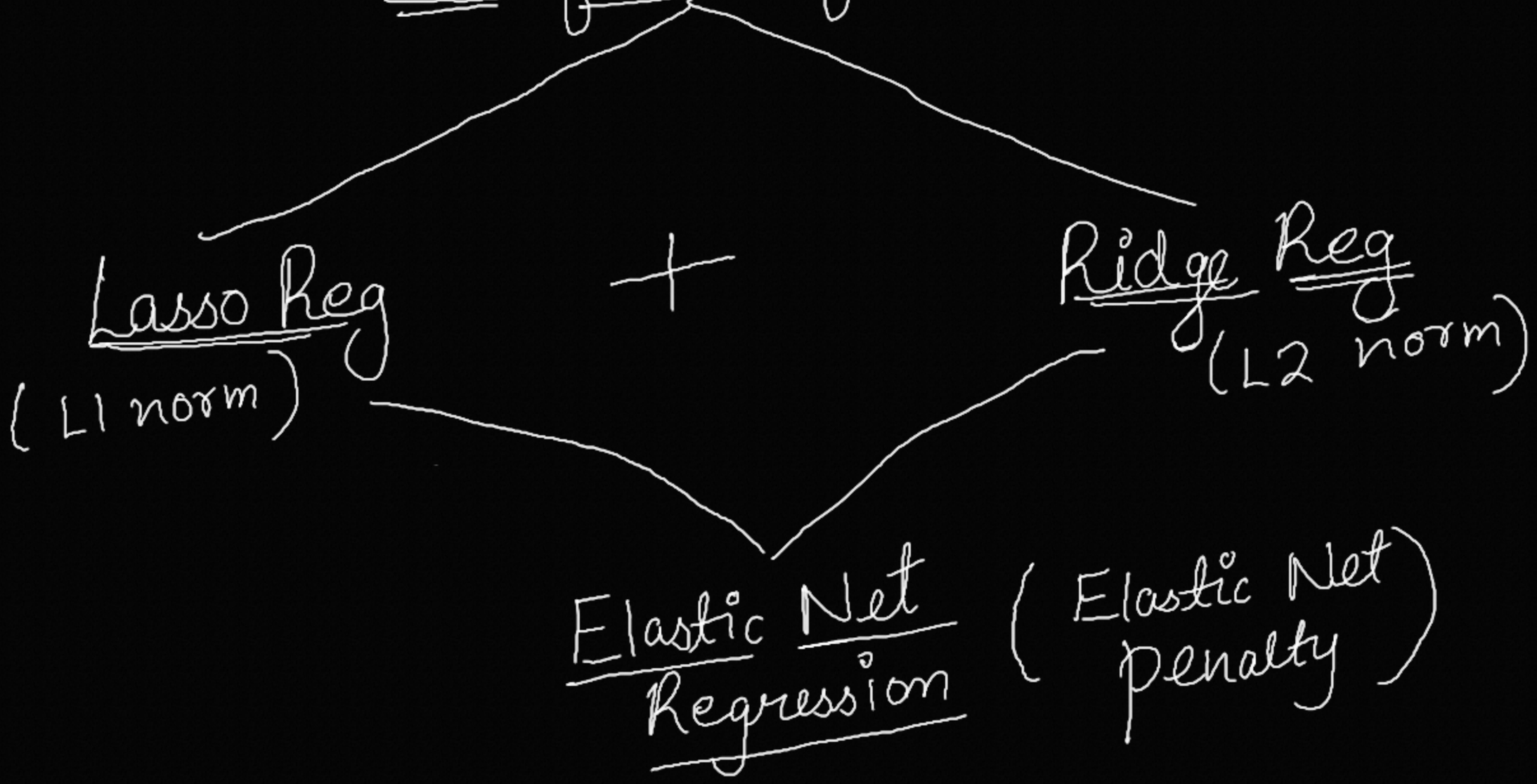
Two forms of Linear Regression



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;

$\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 1$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

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

$$\boxed{Y = X_1 + X_2}$$ ⟹ → 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 Set the L1 ratio

L1 ratio ⇒ It is ratio b/w L1 norm and L2 norm

It ranges from 0 to 1

If L1 ratio = 0 ⇒ Model behave as Ridge Regression

L1 ratio = 1 ⇒ Model behave as Lasso Regression

L1 ratio = 0.5 ⇒ Balance b/w Ridge & Lasso.

eg. You set $\alpha = 1$ and L1 ratio = 0.2

So, alpha (Lasso) = 0.2

alpha (Ridge) = 1 − 0.2 = 0.8

(4) Calculate Elastic net penalty (You set L1 ratio = 0.4 and overall $\alpha = 1$)

$$ENP = \alpha_L(|\beta_1| + |\beta_2|) + \alpha_R((\beta_1)^2 + (\beta_2)^2)$$

$$= 0.4(1+1) + 0.6(1+1)$$

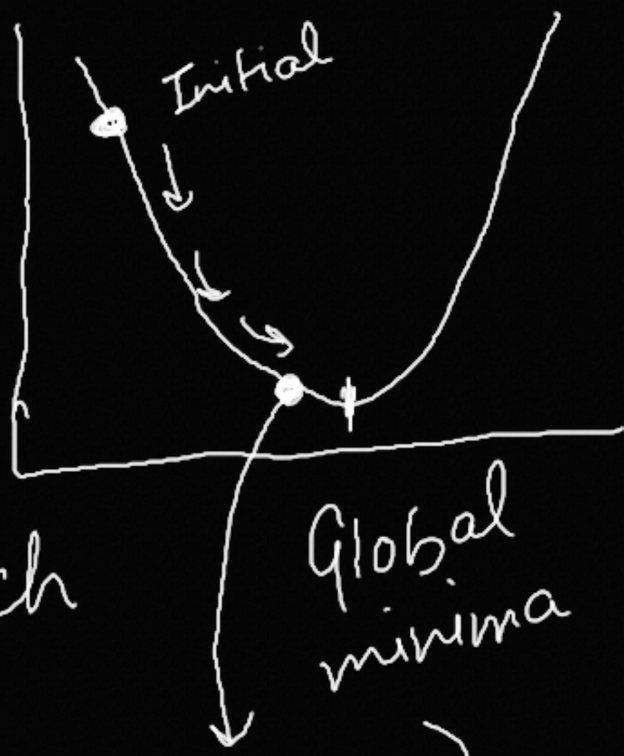
$$= 0.4 \times 2 + 0.6 \times 2$$

$$= 2$$

(5) Calculate Cost fn / objective fn

Cost fn = MSE + ENP

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



Aim: Minimize the cost fn as much 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 l1_ratio ⇒ default = 0.5

(Most imp) You can set it b/w 0 to 1

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.

⑦ selection ⇒ Already discussed in Lasso reg

⑧ precompute ⇒ Already discussed in Lasso reg

Keys	Lasso	Ridge	Elastic Net
Penalty Type	L1 norm (absolute values) $\Rightarrow \alpha(\lvert\beta_1\rvert + \lvert\beta_2\rvert)$	L2 norm (squared values) $\Rightarrow \alpha((\beta_1)^2 + (\beta_2)^2)$	Elastic Net penalty (combination) $\Rightarrow$ L1 norm + L2 norm
Feature Selection (Effect on Coefficients)	Shrinks most of the coefficients to exactly zero Made for agressive feature selection	Shrinks coefficient but not exactly to zero Can't do feature selection	Combines both it may shrink coefficient to exact zero. Moderate feature selection

Handling Multicollinearity	Not good for this task	Most effective in this task	Better than Lasso but not as good as Ridge
Use Case	When you know your data has many irrelevant features. ie. Mainly focused on Feature Selection	When your data has mainly important features and most of features are correlated. ie. Mainly focus on handling Multicollinearity	When you are dealing with highly correlated features as well as you want to do moderate feature selection also.