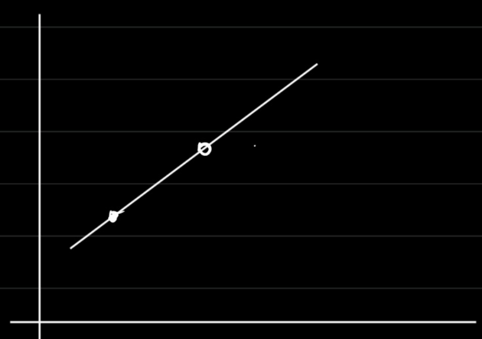
* Ridge Regression (L2 Penalty, L2 norm)

$$SLR \Rightarrow CF = \frac{1}{n}\sum_{i=1}^{n}(y_{act} - y_{pred})^2$$

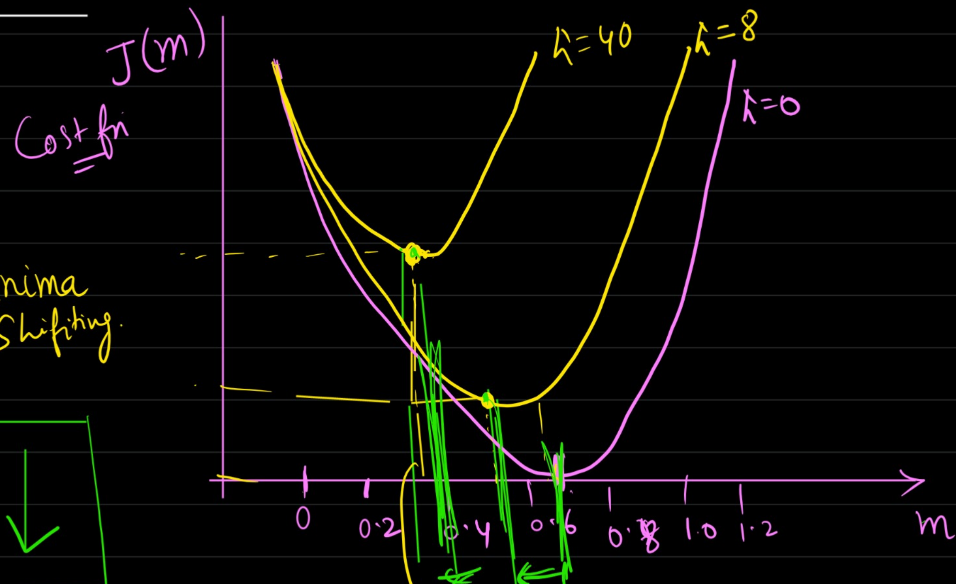
Hyperparameter

$$Ridge \Rightarrow CF = \frac{1}{n}\sum_{i=1}^{n}(y_{act} - y_{pred})^2 + \lambda\,(slope^2)$$

$y_{pred} = mx + c$ $\lambda = 1$



$$0 \quad + \quad 1 \times m^2 \rightarrow \text{Penalizing the Cost Funct}$$



λ ↑ Cost fun (J(m)) ↑ minima Shifting.

$\boxed{\lambda \uparrow \text{ slope } \downarrow}$

→ optimal m.

→ Slope (m) will never become 0 in Ridge.

Simple linear regression

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$Cf = \frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x)_i)^2 + \underbrace{\lambda((\theta_i)^2)}_{\text{L2 norm}}$$

* Multiple linear Regression

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$\text{Cost fn} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x_i)\right)^2 + \lambda\left((\theta_1^2)+(\theta_2^2)+(\theta_3^2)\right)$$

* Ridge advantage

→ Reduces Overfitting → by reducing the coeff (coeff will be close to 0)

→ effective in handling multi collinearity.

$$y = 2.3x_1 + 4.2x_2 + 0.01x_3 + 2$$

Disadvantage

→ doesn't make least important feature 0 → feature selection.

Unit change in x_1

2.3 unit change in y

$0 \times x_3$

* Lasso Regression (L_1 Regularization, L_1 norm) → Feature Selection.

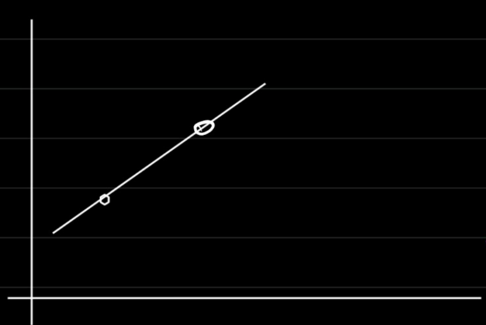
(Least Absolute Shrinkage and Selection operator)

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)_i\right)^2 + \lambda\sum_{i=1}^{n}|\text{Slope}|$$

↳ L_1 regulari

$$\Downarrow \qquad\qquad \Downarrow$$

$$\text{Error} + \lambda\left(|\theta_1| + |\theta_2| + |\theta_3|\right)$$



$\lambda \uparrow J(\theta) \uparrow$ Global minima will shift

$\lambda \uparrow \theta \downarrow$

$\theta \leftarrow \theta_{old}$ $\theta \leftarrow \theta$

$$h_\theta(x) = 2.3x_1 + 4.2x_2 + \cancel{0.01x_3} + 2$$

$\to 0$

At one point $\theta \approx 0$

$$h_\theta(x) = 2.3x_1 + 4.2x_2 + 2$$

* Advantage

→ It helps us in removing insignificant feature (Sparsity) ← Automatic feature selection

$\theta = 0$

→ multicollinear

→ overfitting. (To some extent)

③ Elastic Net. → Reduce Overfitting (Ridge)
→ Feature Selection (Lasso)

$$\text{Cost Fn} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i - h_\theta(x_i))^2}_{\text{MSE}} + \underbrace{\lambda_1 \sum_{i=1}^{n}(\text{slope}^2)}_{L2} + \underbrace{\lambda_2 \sum_{i=1}^{n}|\text{slope}|}_{L1}$$

L2: Reduce (Overfitting)

L1: (Feature Selection)

λ_1, λ_2 → Hyperparameter tuning.

* Occam's Razor Principle → The simpler models are the best models as compared to complex model.