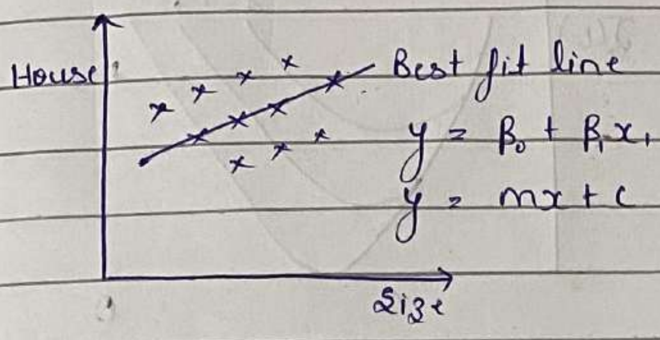
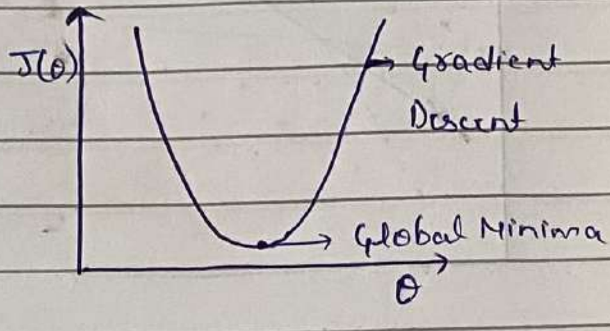
Elastic Net Regression

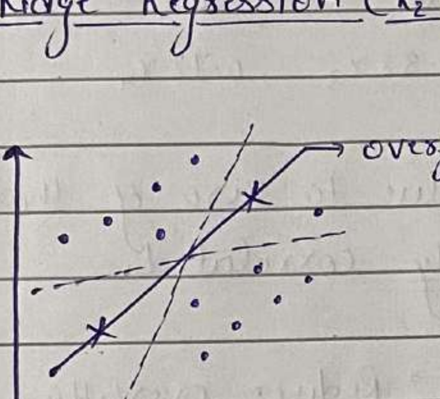
→ **Linear Regression :**

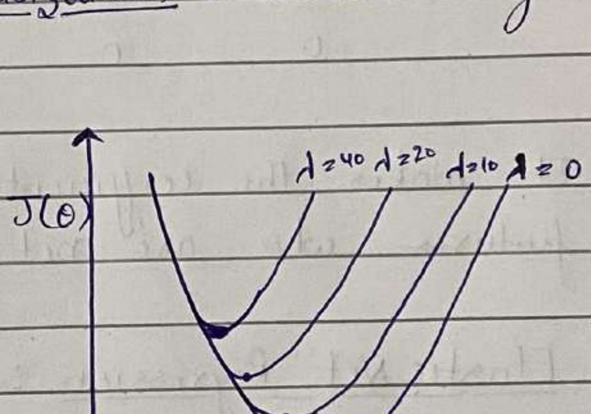




Cost Function : $\frac{1}{2m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$ [Mean Square Error]

→ **Ridge Regression (L_2 Regularization) :** Reduce overfitting





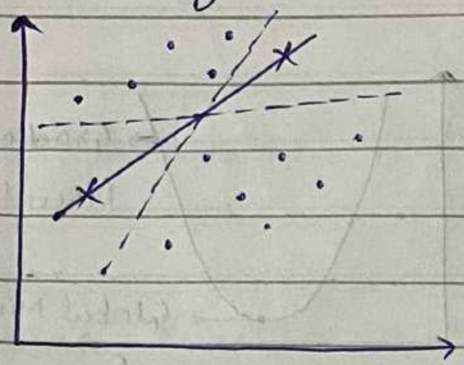
Cost Function : $\frac{1}{2m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{m} (\theta_j)^2$

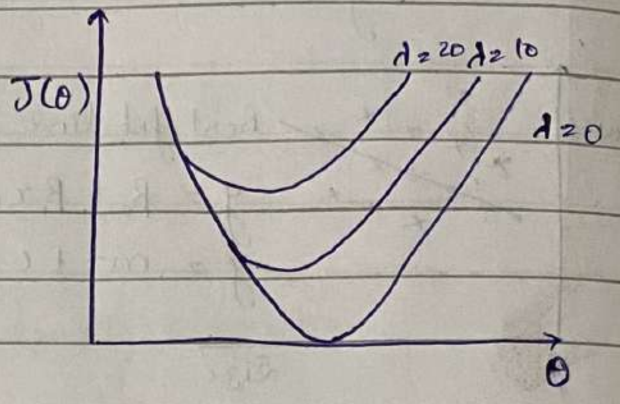
$$\downarrow$$

$0 + \Sigma(\theta_j)^2 \rightarrow$ +ve value

By multiplying with lambda (λ), the coefficient's value (weightage) will reduce but never become zero.

→ **Lasso Regression (L_1 Regularization) : Feature Selection**





Cost Function : $\frac{1}{2m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2+\lambda\sum_{j=1}^{m}|\theta_j|$

$h_\theta(x) = 0.01x_1 + 0.02x_2 + 0.57x_3 + 0.96x_4 + 0.45$

$\downarrow$: $0.01x_1 \to 0$, $0.02x_2 \to 0$, $0.57x_3 \to 0.38x_3$, $0.96x_4 \to 0.72x_4$

It makes the coefficient's value to zero of those features who are not highly correlated.

→ **Elastic Net Regression** : → Reduce overfitting
→ Feature Selection

Cost Function : $\frac{1}{2m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2+\lambda_1\sum_{j=1}^{m}(\theta_j)^2+\lambda_2\sum_{k=1}^{m}|\theta_k|$