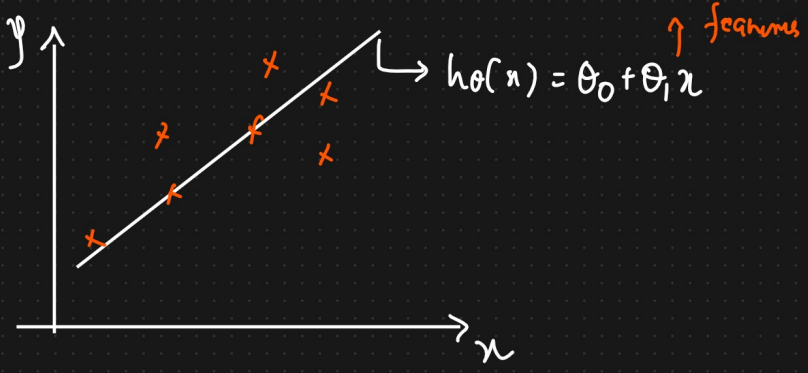
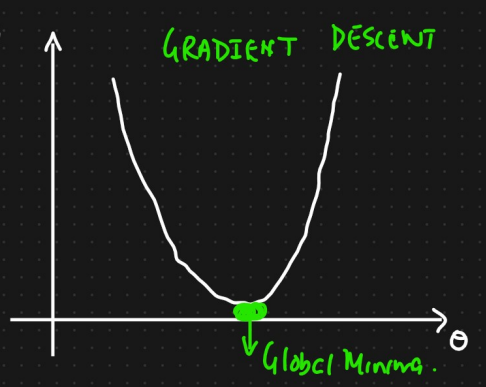
Ridge Regression, Lasso Regression, Elasticnet Regression

Linear Regression





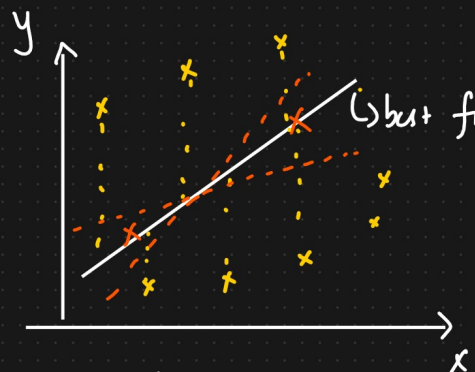
$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

⇓

Mean Squared Error

① Ridge Regression (L2 Regularization) → Reduce Overfitting

Overfitting



- Traindata → Acc ↑ → low Bias
- Test data → Acc ↓ → High Variance

$h_\theta(x) = \theta_0 + \theta_1 x_1$

$\lambda = 0$

$\lambda = 1$

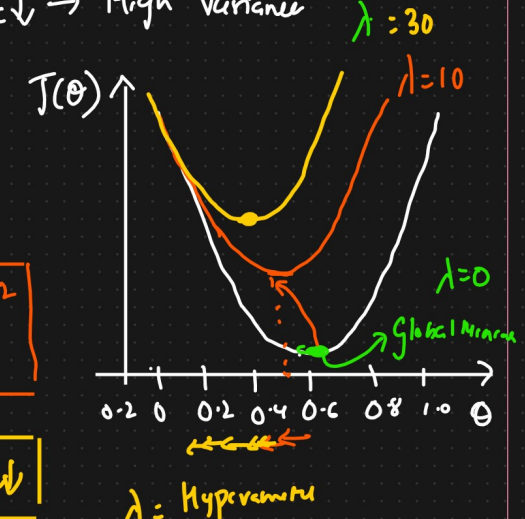
$\lambda = 10$

Hyperparameter

$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \lambda\sum_{i=1}^{n}(\text{slope})^2$$

$$= 0 + (1)\left[(\theta_1)^2\right]$$

$\lambda \uparrow$ Slope ↓



$\lambda = 1$

> 0

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

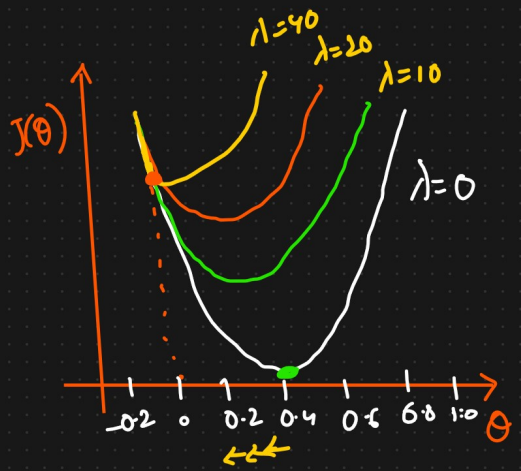
0.24

$$= 0.34 + 0.52x_1 + 0.48x_2 + 0.24x_3$$

$$= 0.34 + 0.40x_1 + 0.38x_2 + \boxed{0.14x_3}$$

② Lasso Regression (L_1 Regularization) → Feature Selection

$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right) + \boxed{\lambda \sum_{i=1}^{n} |\text{slope}|}$$



$$h_\theta(x) = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_3 + \theta_4x_4$$

$$h_\theta(x) = 0.52 + 0.65x_1 + 0.72x_2 + 0.34x_3 + \boxed{0.12x_4}$$

⇓ Lasso Regression

⇓ → 0

$$= 0.52 + 0.51x_1 + 0.60x_2 + 0.14x_3 + \boxed{0x_4}$$

③ ElasticNet Regression

→ ① Reduce Overfitting

→ ② Feature Selection

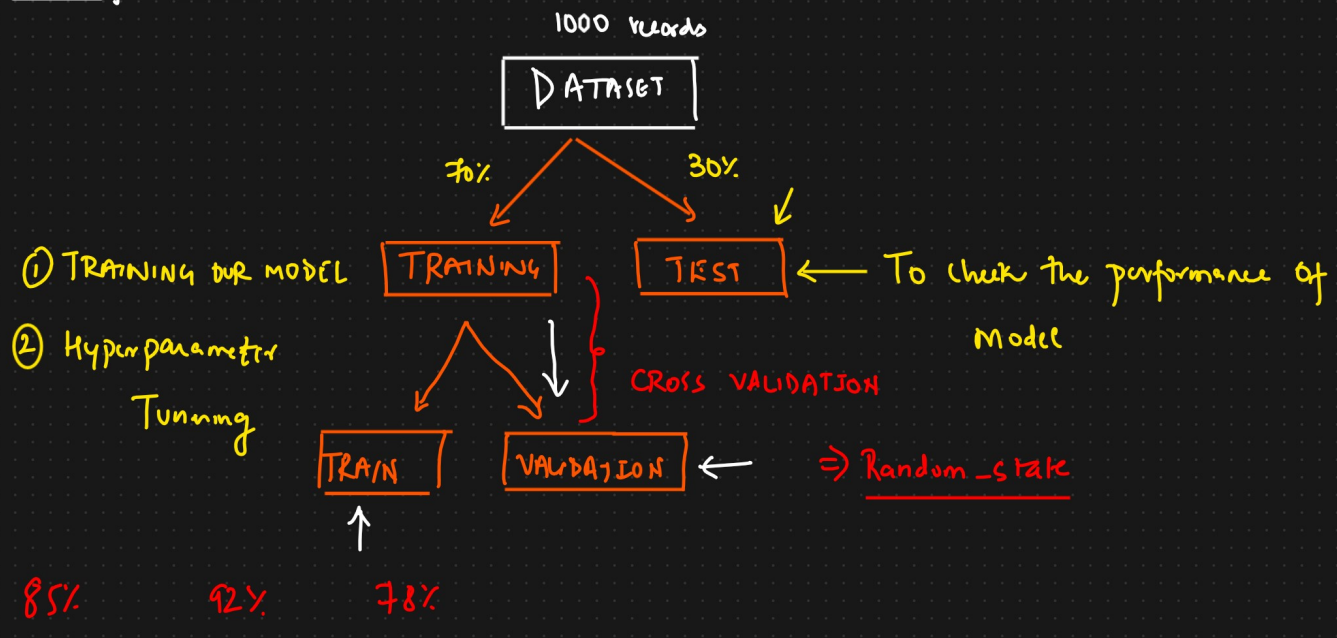
$$\text{Cost fn} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 + \boxed{\lambda_1 \sum_{i=1}^{m} (\text{slope})^2} + \boxed{\lambda_2 \sum_{i=1}^{m} |\text{slope}|}$$

⇓ Reduce Overfitting

⇓ Feature Selection

{ Hyperparameter Tuning the Linear Regression }

Types of CROSS VALIDATION

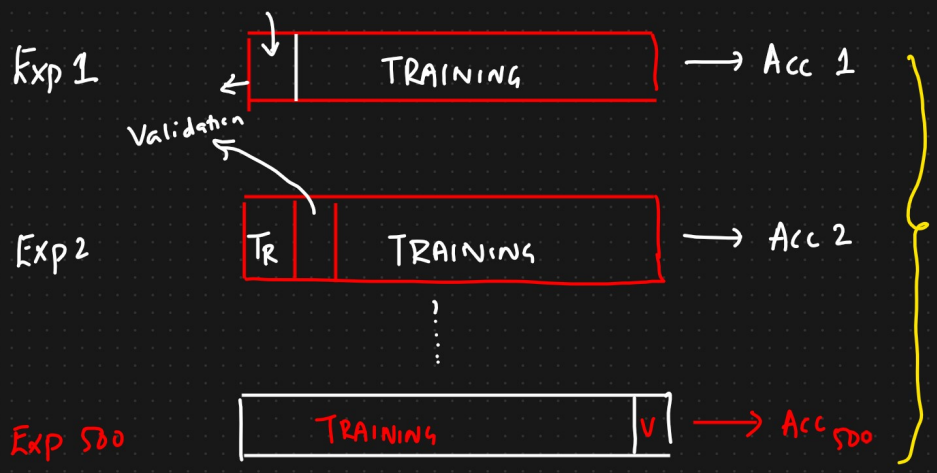


① leave One Out CV (LOOCV)

② leave p out CV

p=10 p=20 p=30

TRAINING → 500 Records ↑↑ Complexity of Training Model

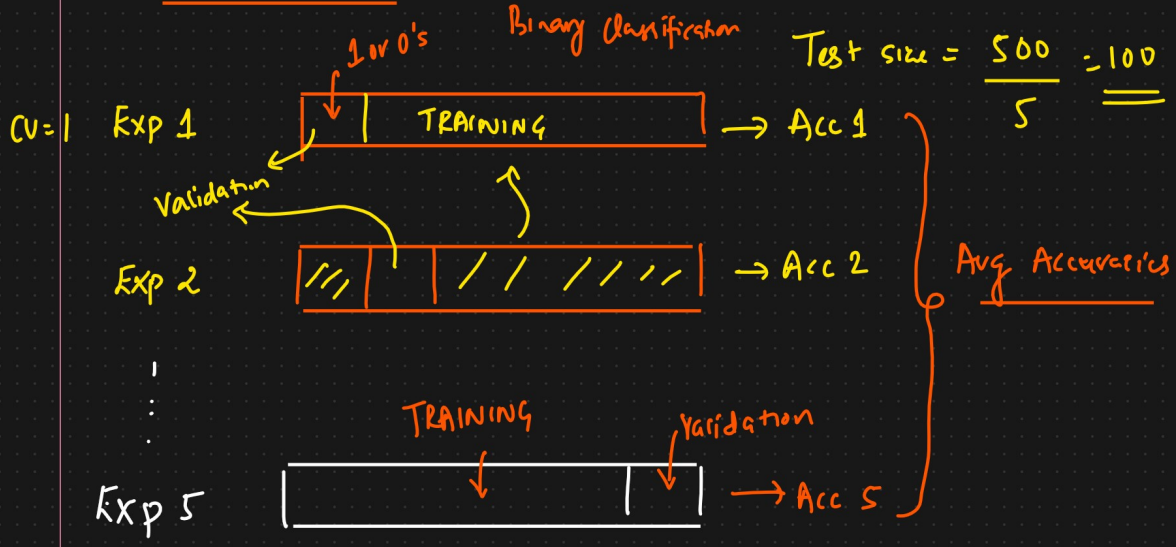


① Overfitting → TRAINING ↑↑ Acc → New Test Data → Acc ↓↓

Validation Acc ↓↓

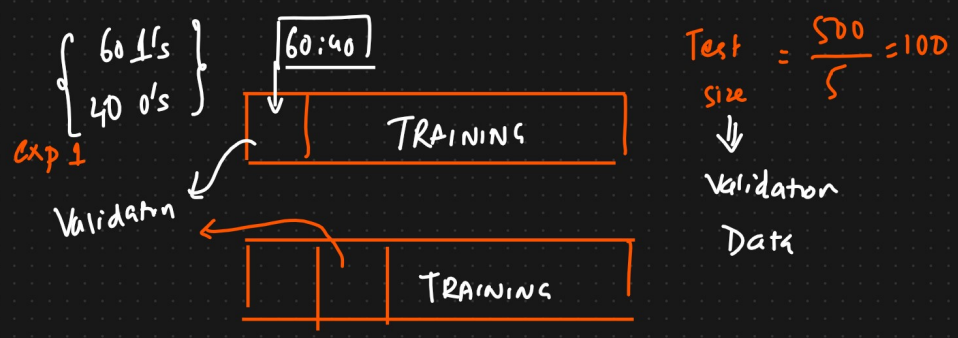
③ K Fold CV

K=5 n=500



④ Stratified K Fold CV

K=5



⑤ Time Series CV

Reviews

Product Sentiment Analysis

Time

JAN → DEC

TRAINING | Validation

DAY 1 DAY 2 DAY 3 DAY 4 . . . DAY N

Time Series Application