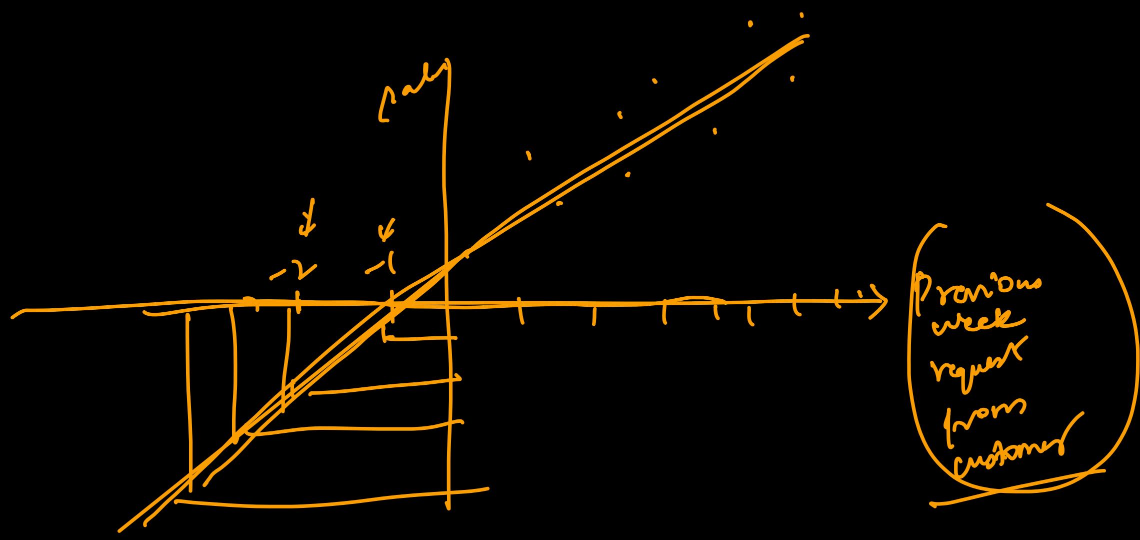
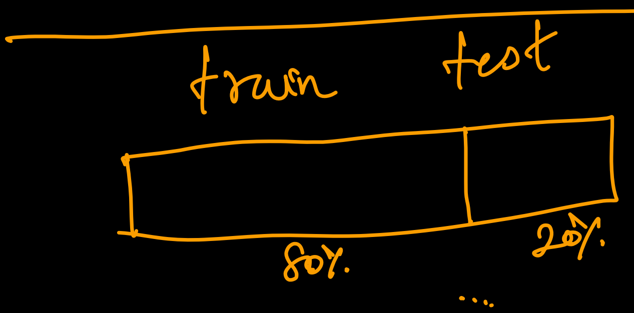
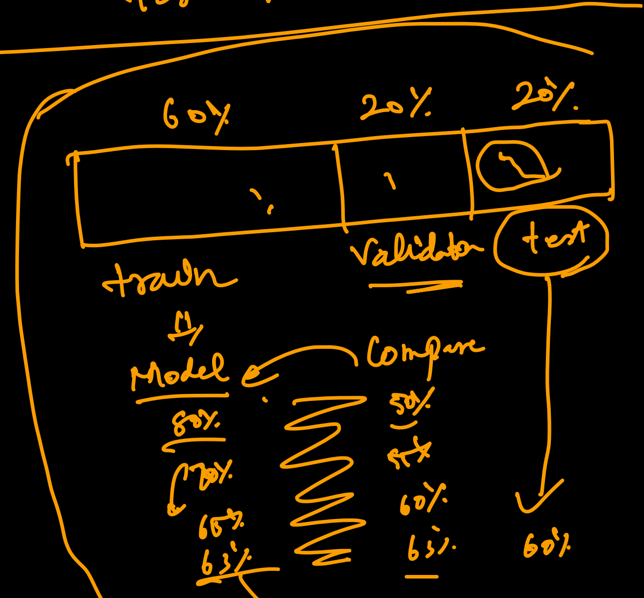


Overfitting

1. Quality of Data
2. Complex model
3. lot of adjustment done in tuning the result in train
4. Building the model without test validation



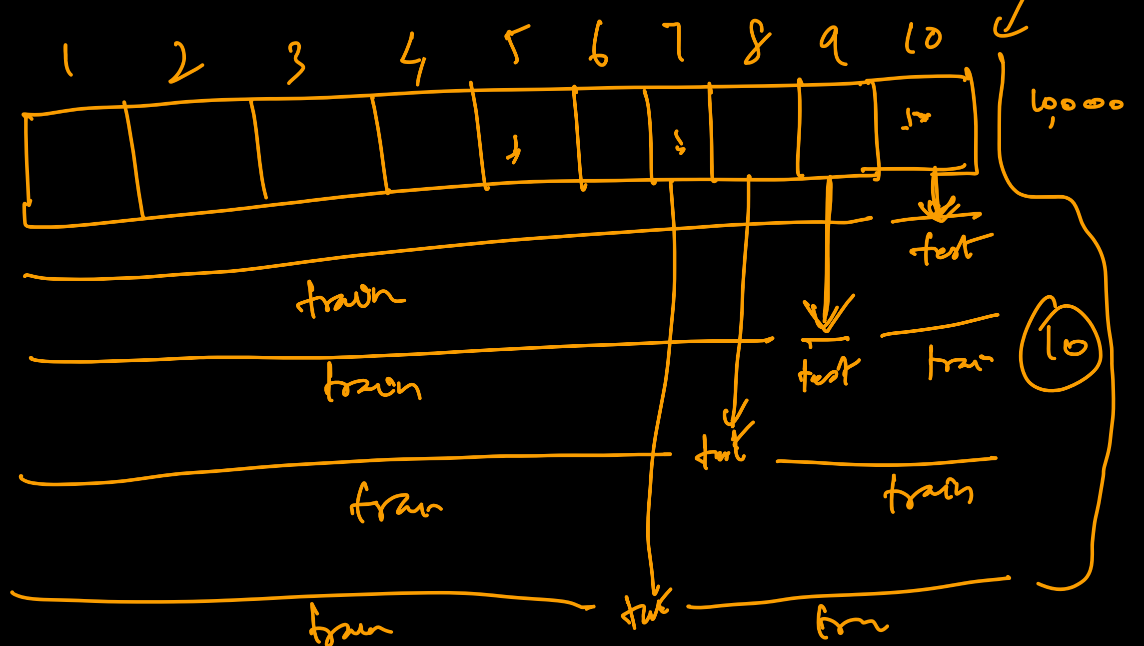


K fold Validation

Cross validation

10 K-fold

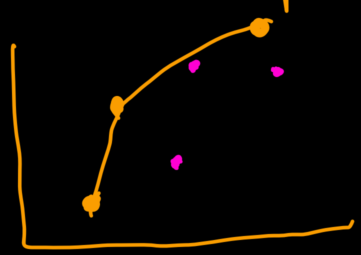
5



High Variance

Overfitting

Model is Complex

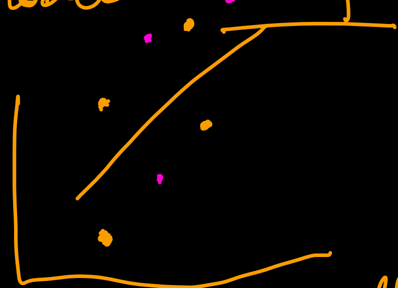


If there is a small change in train data there will be a big shift in the model

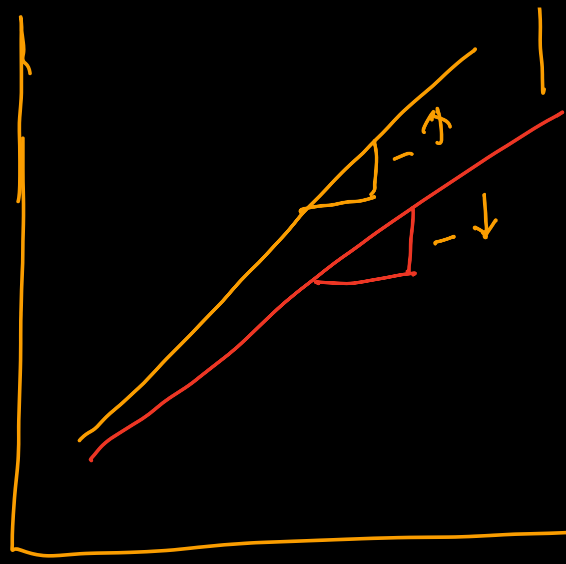
High Bias

Underfitted

Model is simple



If there is a small chang in train data there won't be a big change in the model

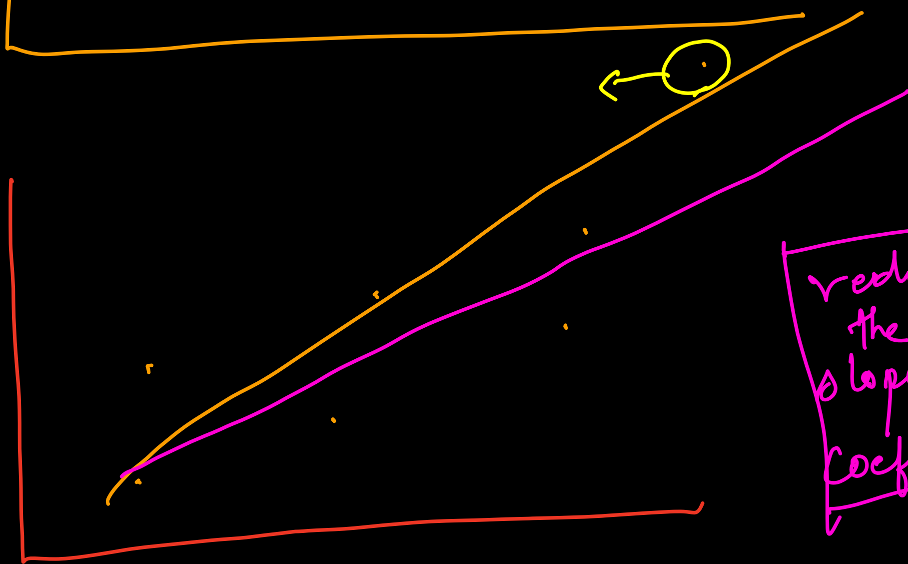


overfit

$m = \uparrow$

$m = \downarrow$

$mx + b$



reduce the slope, the coefficient

1. Ridge Regression
2. Lasso Regression
3. Elastic Net Regression

Ridge Regression (L2)

Cost

$$MSE = \frac{\sum (y - \hat{y})^2}{n}$$

$$Cost = \frac{\sum (y - (mx + b))^2}{n}$$

Objective = find m and b where cost is minimum

Started with $\boxed{\text{Random number}}$

Penalize my cost function whenever you chose big Random number

$$Cost = \frac{\sum (y - (\beta_0 + \beta_1 x))^2}{n} + \lambda \sum_{i=0}^{n} \beta^2$$

Make the Slope of Unsignificant Column near to Zero

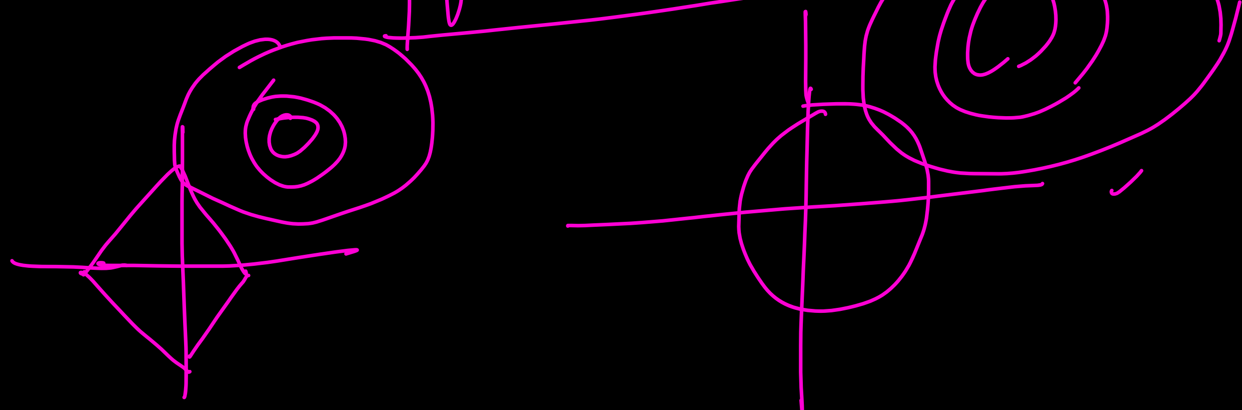
Lasso Regression (L1)

Least Absolute and $\boxed{\text{Selection}}$ Operation:

$$\text{Cost} = \sum^{n} \left(y - (\beta_0 + \beta_1 x)\right)^2 + \lambda \sum_{i=0}^{n} |\beta|$$

Unsignificant column will become Zero

$\boxed{\text{feature Selection}}$



Salary = 5000 + 2.5 Experiance + 1.0 Education + 0.02 phone number

Salary = 4000 + 2.45 Experience + 0.8 educatn + 0 (phone number)

Shrinks the less impormant features coefficient

Elastic Net

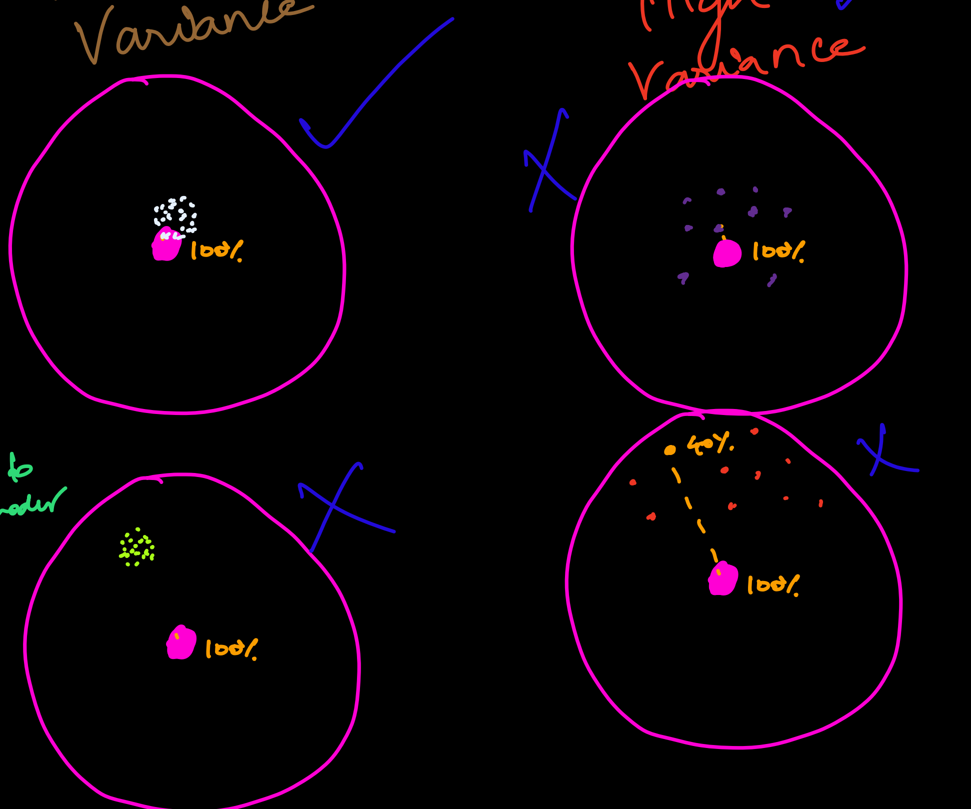
$$\text{Cost} = \sum_{i=1}^{n} \left(y - (\beta_0 + \beta_1 x)\right)^2 + \lambda\left(\sum_{i=0}^{n} \beta^2 + (\alpha)\sum |\beta|\right)$$

Regularization

Low Variance

High Variance

Low Bias



more information to increase model complexity

make

High bias