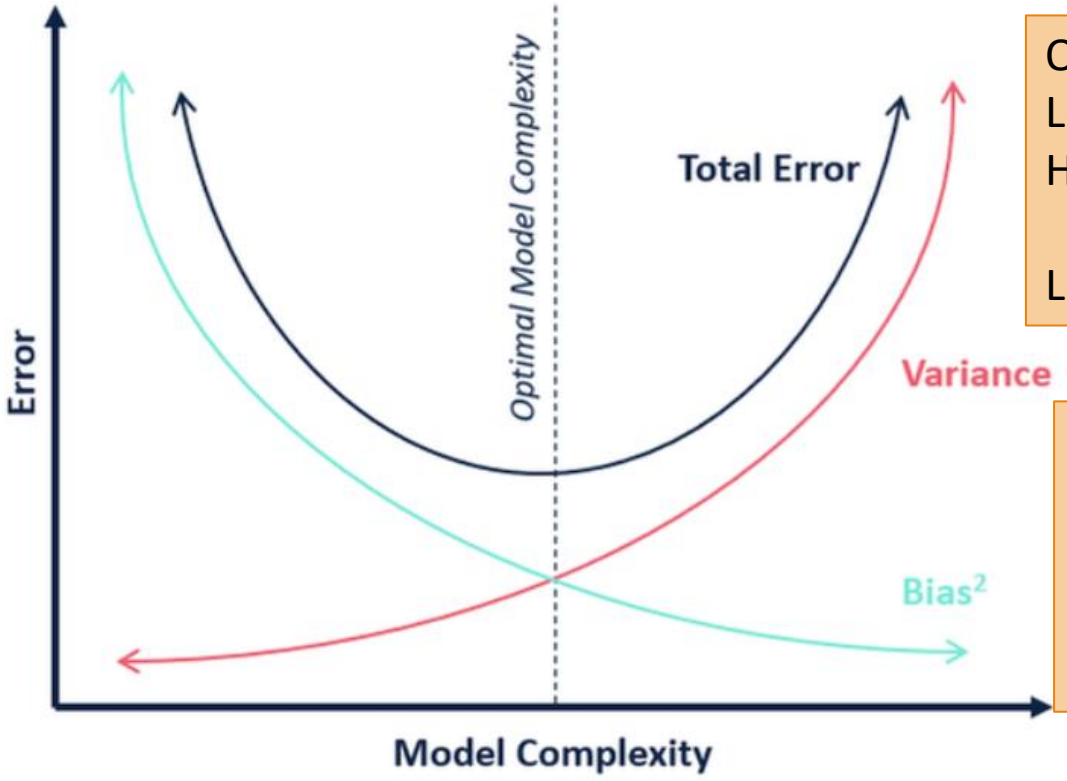
Feature Selection Backward Elimination

UTKARSH GAIKWAD

CLASS STARTING SHARP AT 8:05 PM

Need for feature selection

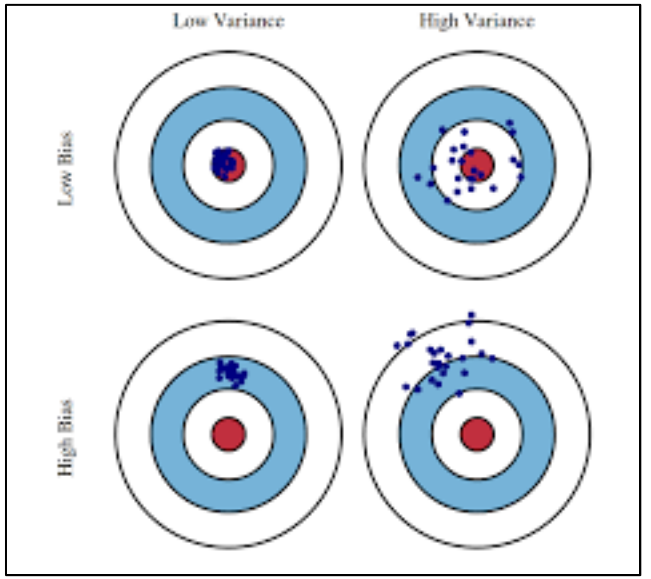


Overfit
Low Training Error
High Testing Error

Low Bias , High Variance

Underfitting
High Training Error
High Testing Error

High Bias , Low Variance

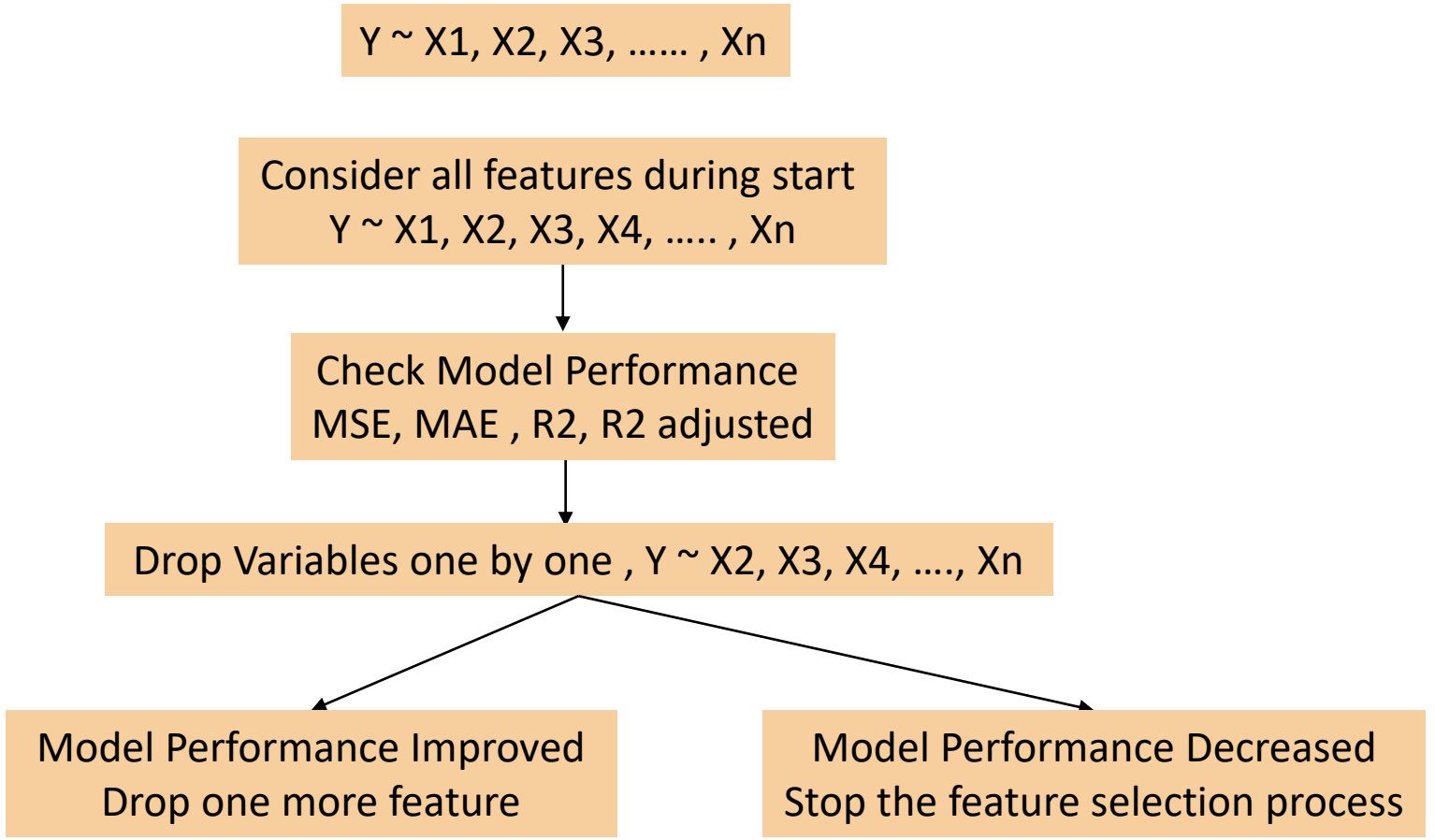


More the numbers of features more complex the model

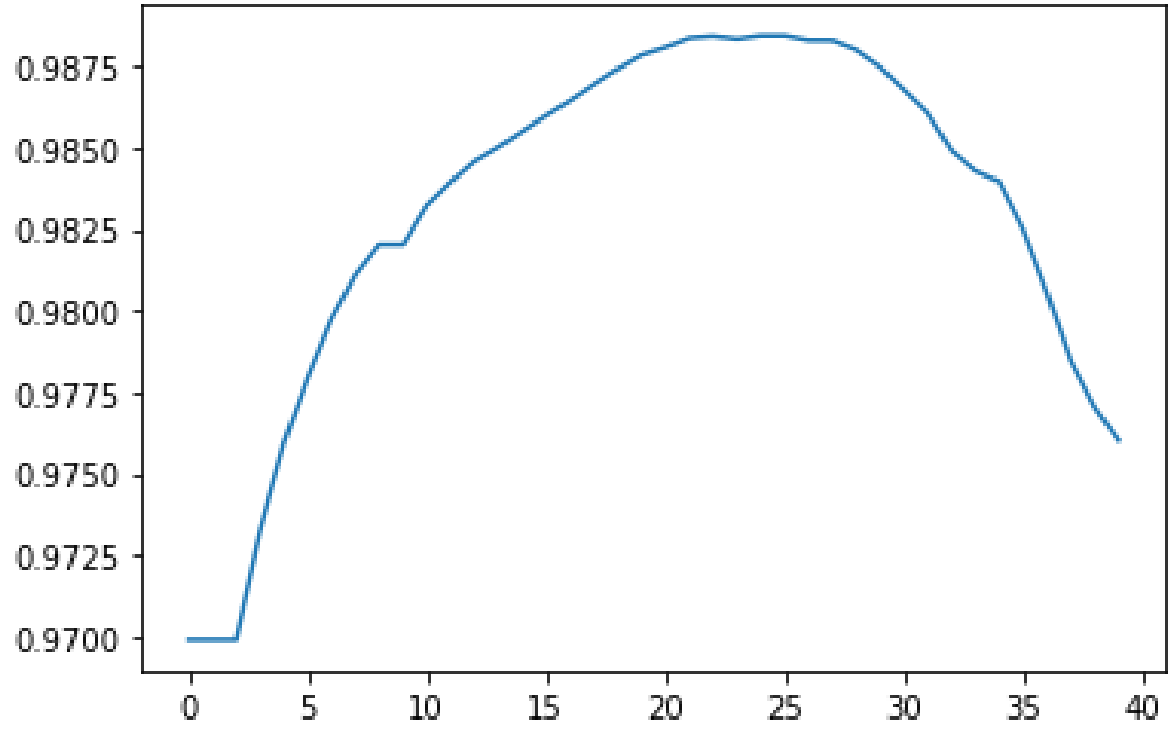
Underfit and Overfit

Overfitting Complex Model 100,200				
Train MSE	0		Train R2	1
Test MSE	12300000		Test R2	0.6
Underfitting Model, Simple 1-2 features				
Train MSE	1234131		Train R2	0.7
Test MSE	1241334		Test R2	0.6

Backward Elimination process



R2 Adjusted



$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where

R^2 Sample R-Squared

N Total Sample Size

p Number of independent variable

R2 Adj Calculation

N	100
p	10
R2	0.95

Num	4.95
Den	89
R2_adj	0.9444

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where

R^2 Sample R-Squared

N Total Sample Size

p Number of independent variable

Thank you

PING ME ON SKYPE FOR ANY QUERIES