Overfitting, Bias and Variance

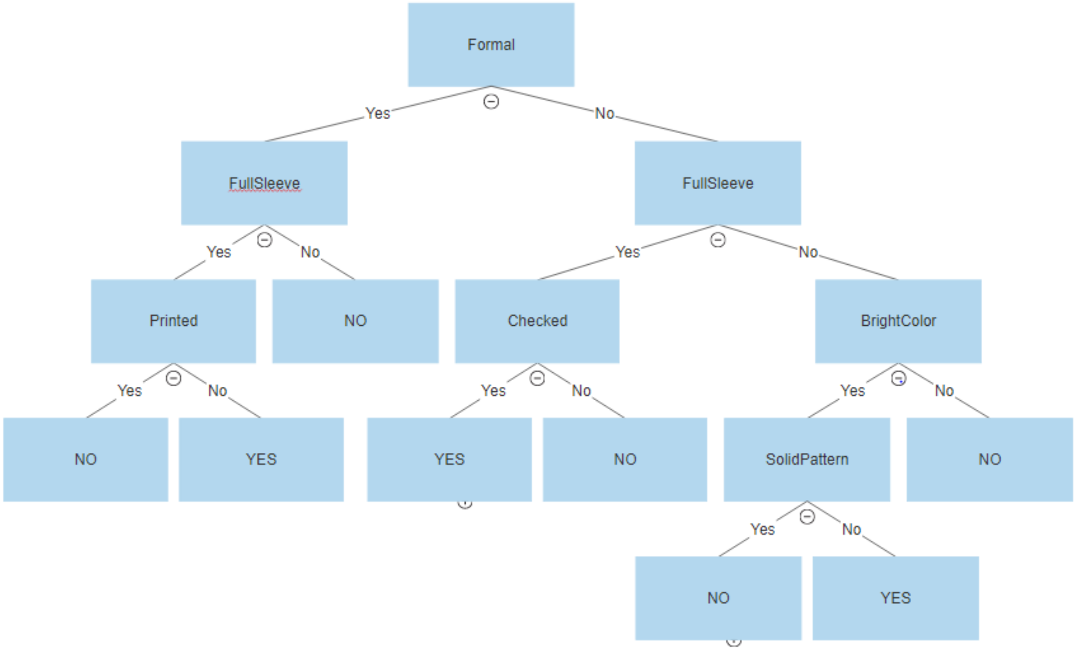
Machine Learning Unit 9

Sudeshna Sarkar

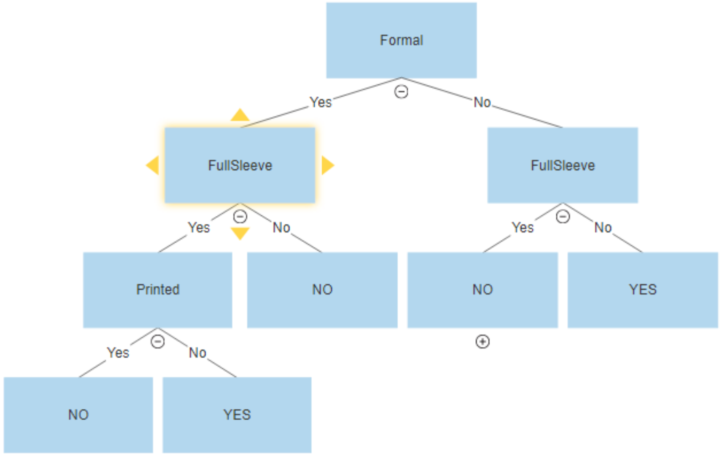
Centre of Excellence in Artificial Intelligence

Indian Institute of Technology Kharagpur

Which Decision Tree?



Training Error = 0.05
Test Error = 0.2



Training Error = 0.1
Test Error = 0.15

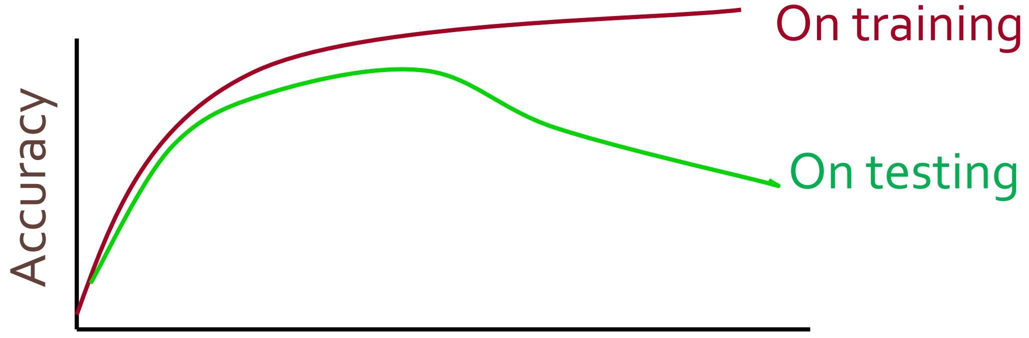
Overfitting

Overfitting :

- Fit the training data too well
- But fail to generalize to new examples

Causes

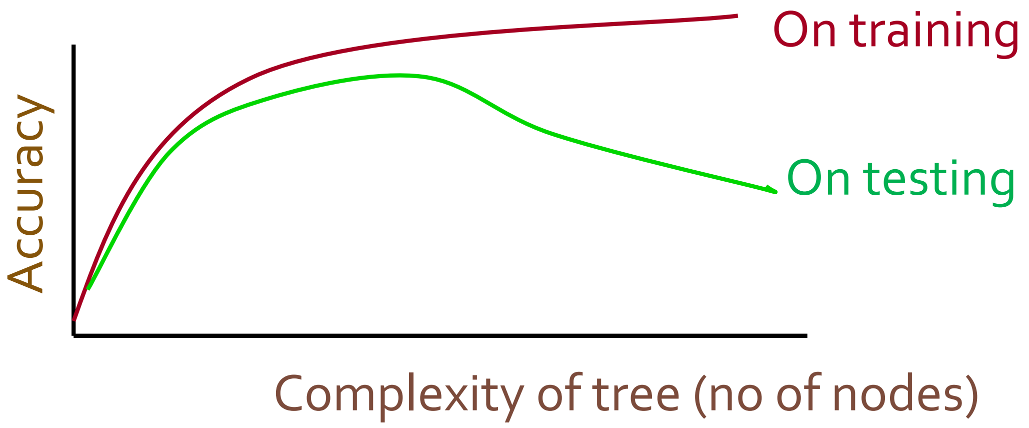
- Noise
- Irrelevant Features
- Insufficient Data



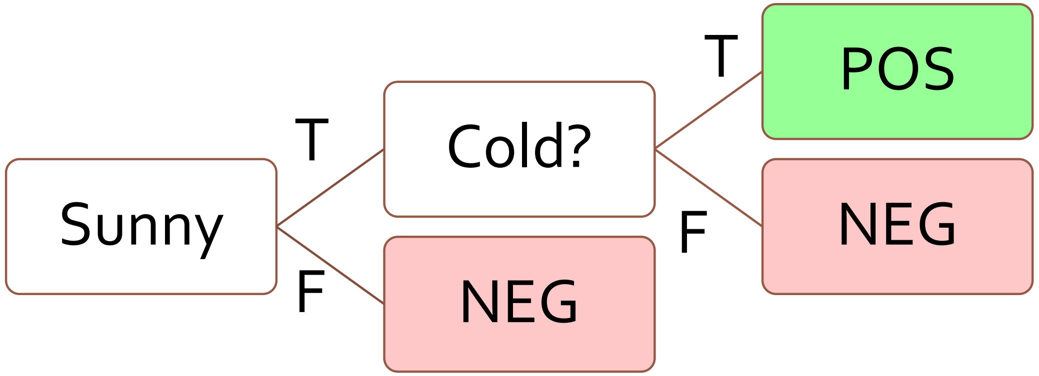
Complexity of tree (no of nodes)

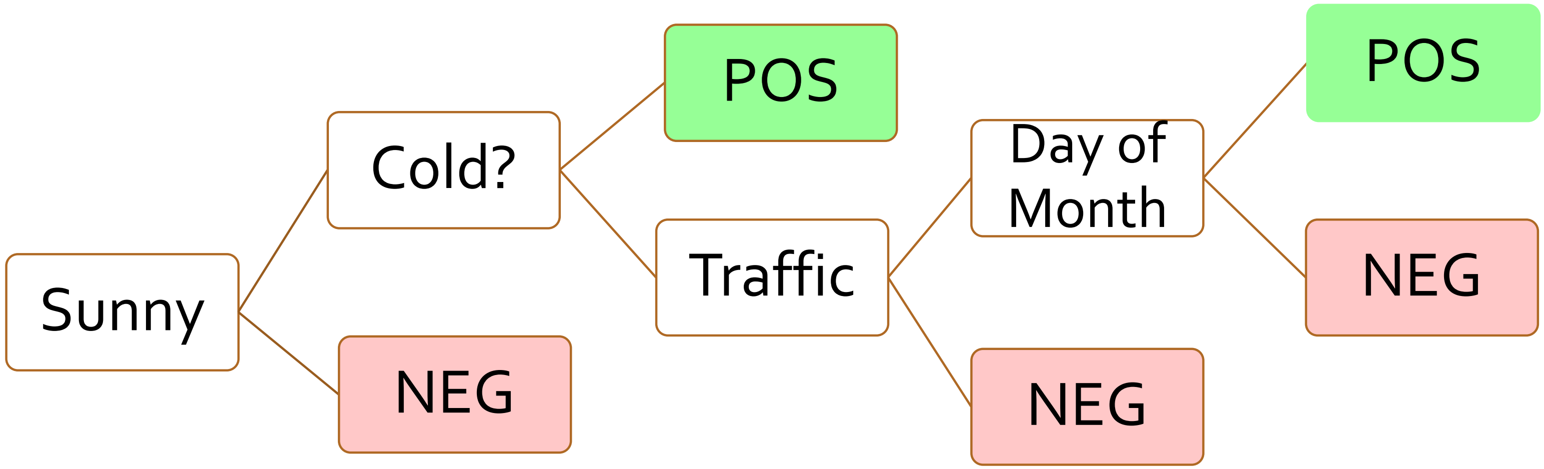
Overfitting

A hypothesis h is said to overfit the training data if there is another hypothesis h' such that h has smaller error than h' on the training data but h has larger error on the test data than h' .



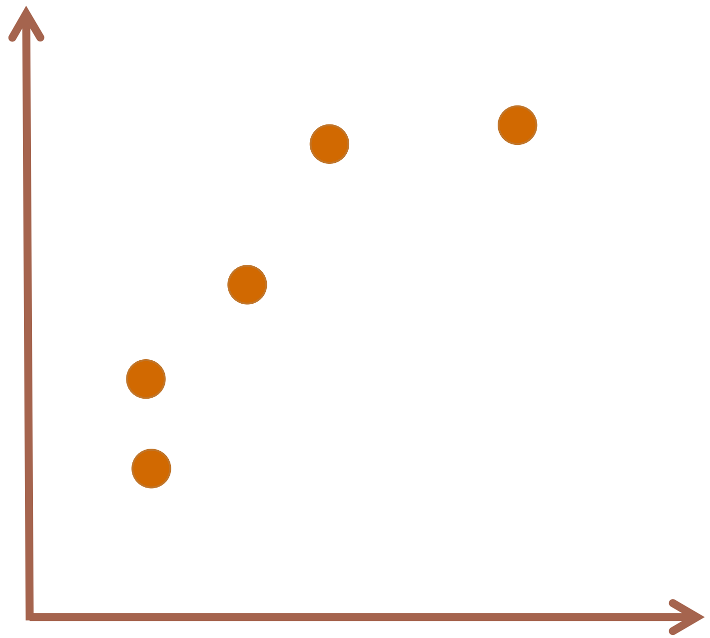
Overfitting with noisy data



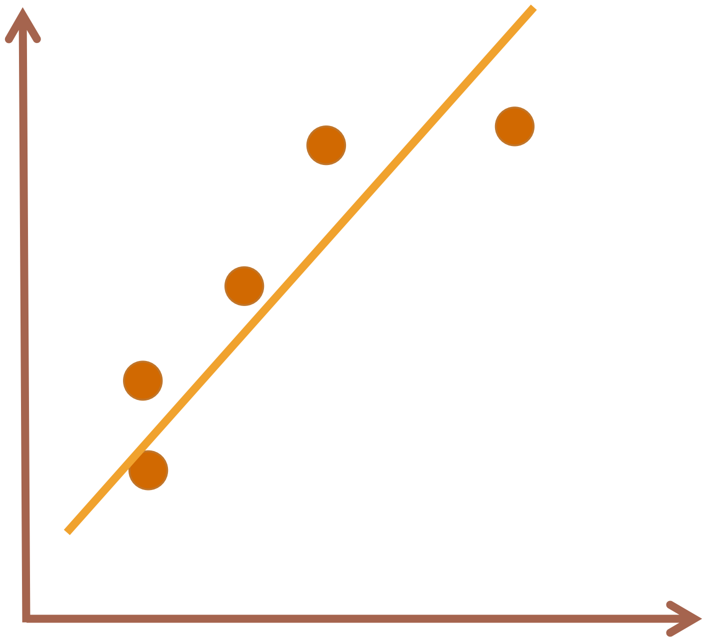


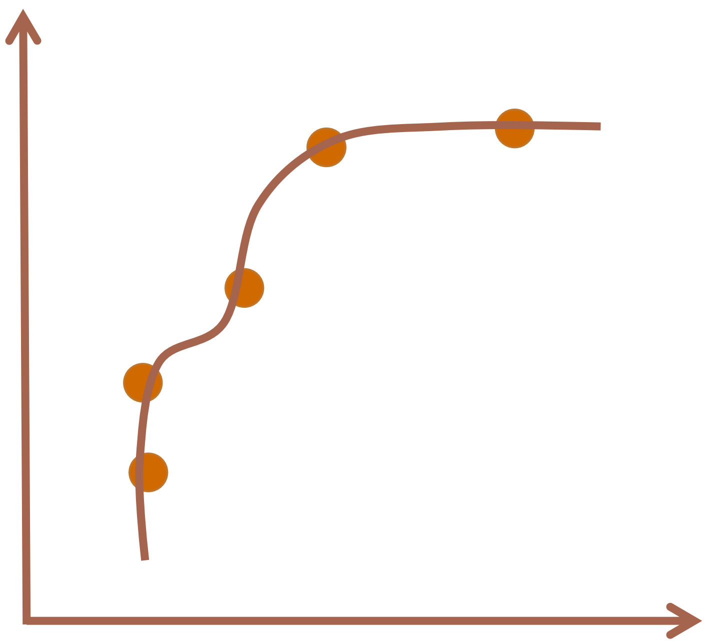
Overfitting results in decision trees that are more complex than necessary

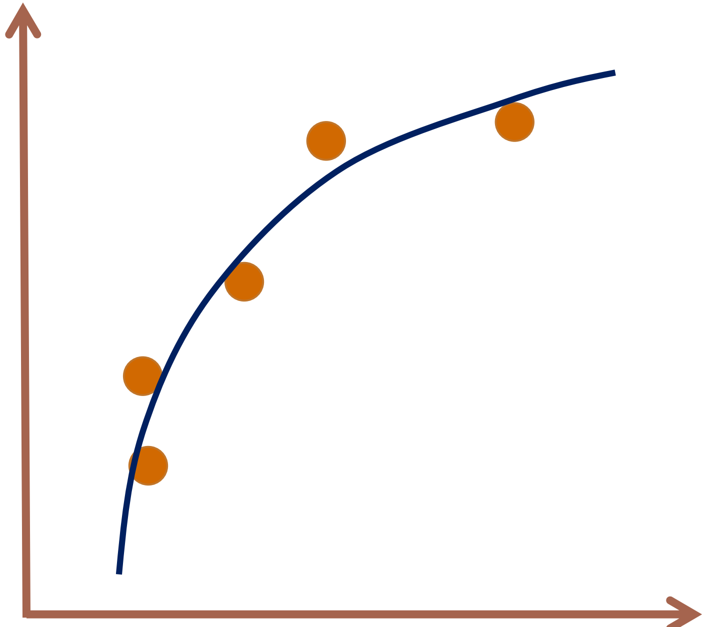
Regression



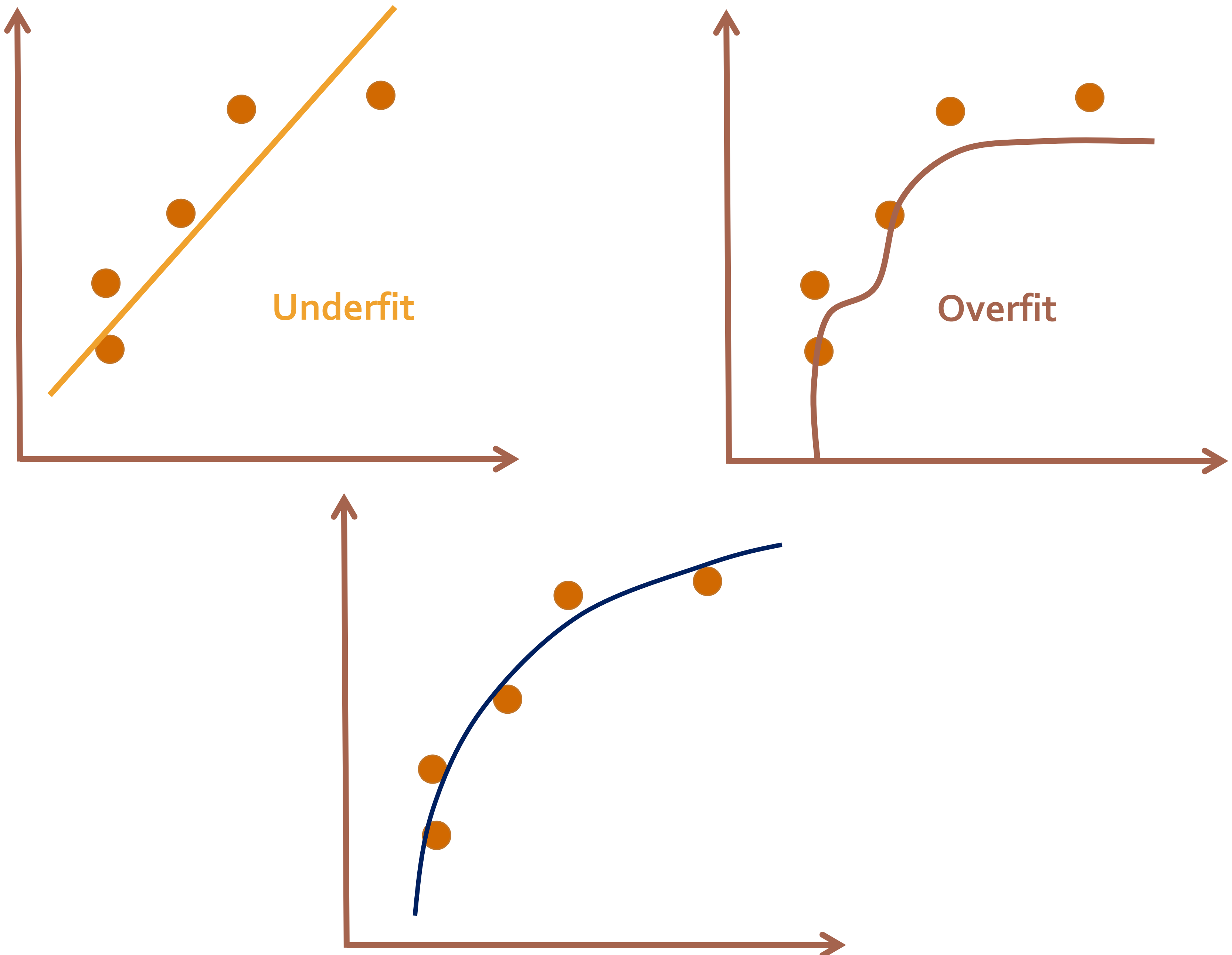
Regression







Regression



Regularization

- In a linear regression model overfitting is characterized by large weights

- Penalize large weights in Linear Regression
  - L2-Regularization or Ridge Regression
  - L1-Regularization

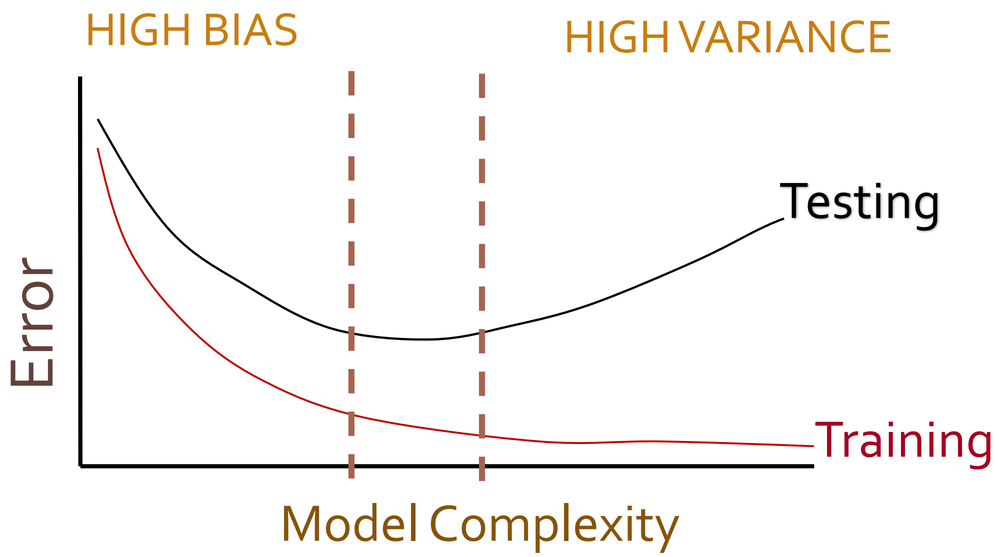
Overfitting vs Underfitting

Underfitting

- Not able to capture the concept
  - Features don't capture concept
  - Model is not powerful.

Overfitting

- Fitting the data too well



Bias

BIAS

- Error caused because the model can not represent the concept
- Bias is the expected difference between the model prediction and the true y's.
- Higher Bias:
  - Decision tree of lower depth
  - Linear functions
  - Important features missing

VARIANCE

- Error caused because the learned model reacts to small changes (noise) in the training data
- High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs
- Higher Variance
  - Decision tree with large no of nodes
  - High degree polynomials
  - Many features

Bias and Variance

BIAS

- if we train models $f_D(X)$ on many training sets D, bias is the expected difference between their predictions and the true y's.

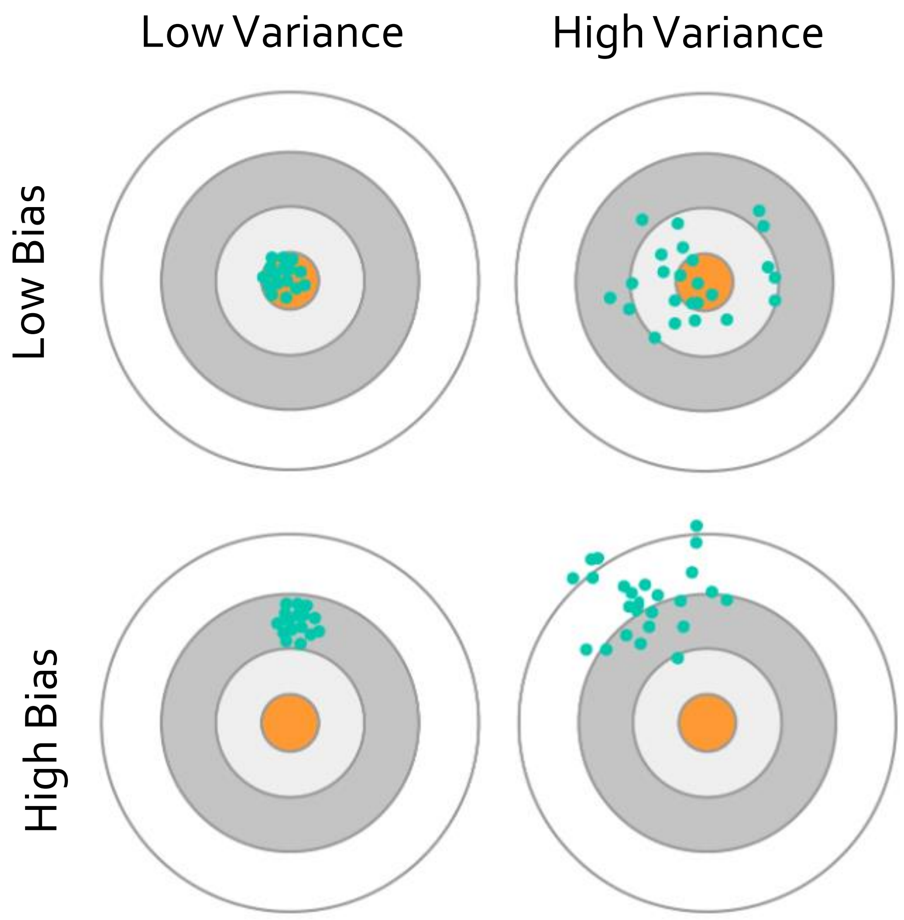
$$Bias = \mathrm{E}[f_D(X) - y]$$

VARIANCE

- if we train models $f_D(X)$ on many training sets D, variance is the variance of the estimates:

$$Variance = \mathrm{E}\left[\left(f_D(X) - \bar{f}(X)\right)^2\right]$$

Bias and Variance



Bias and Variance Tradeoff

There is usually a bias-variance tradeoff caused by model complexity.

Complex models often have lower bias, but higher variance.

Simple models often have higher bias, but lower variance.