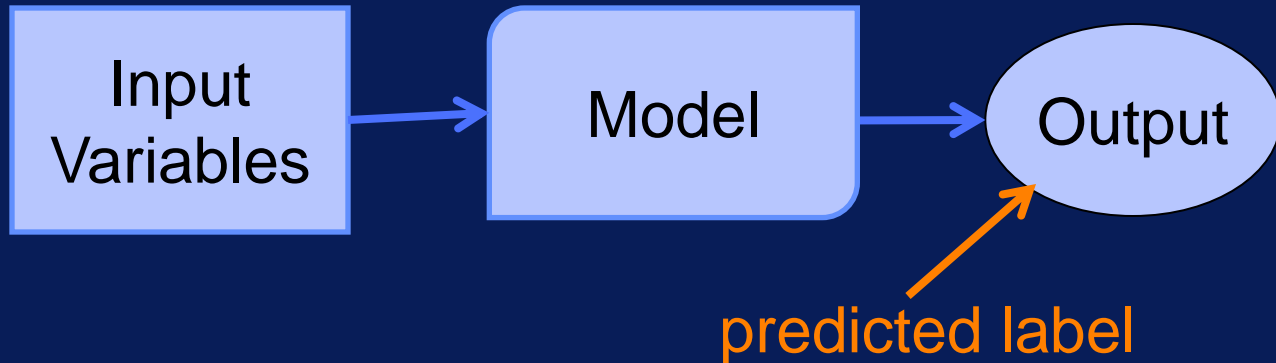# Generalization & Overfitting
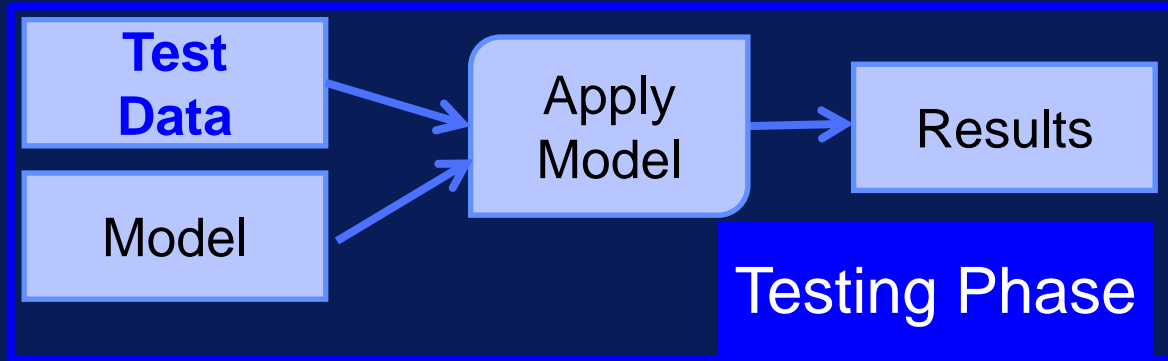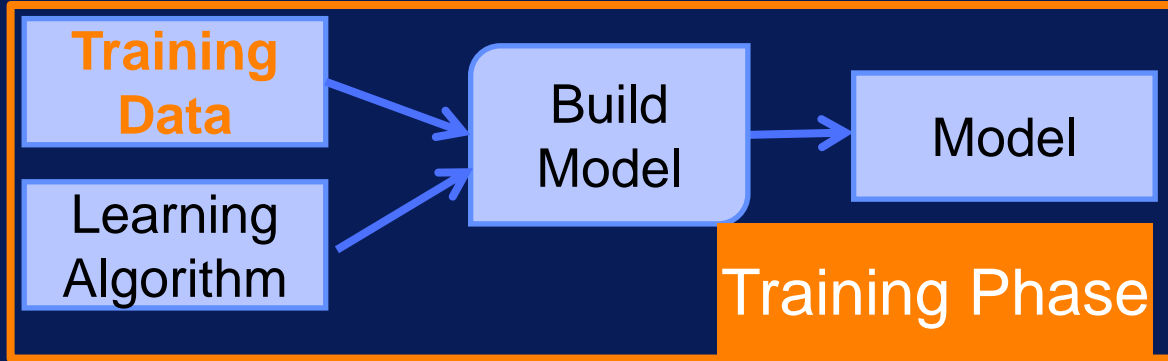
After this video you will be able to..

- Define what generalization is
- Describe how overfitting is related to generalization
- Explain why overfitting should be avoided

Errors in Classification

- Success: Output = Target ← true label
- Error: Output != Target
- Error rate = Error = Misclassification Error
  - # errors / # samples = % error

Input Variables → Model → Output

predicted label

Training vs. Testing Phases

Training Data → Build Model → Model

Learning Algorithm → Build Model

Training Phase

Test Data → Apply Model → Results

Model → Apply Model

Testing Phase

Errors in Classification

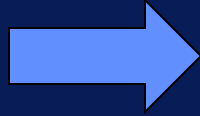Error on Training Data

Training Error

Error on Test Data

Test Error

Test error indicates how well model will perform on new data!
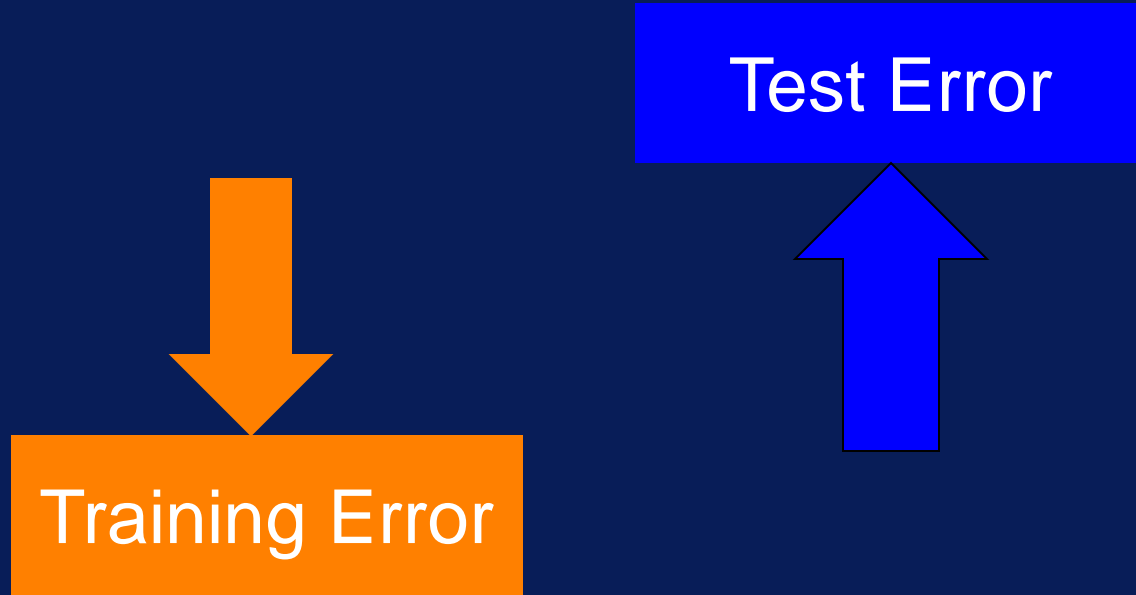
Generalization

Performs well on new data → *Good Generalization*
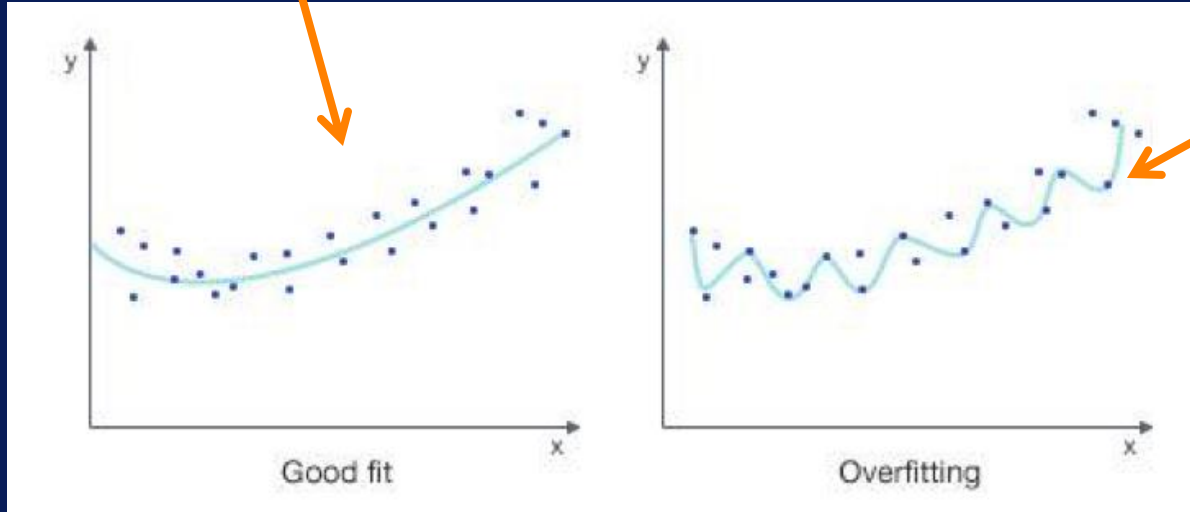

Test Error = Generalization Error

Overfitting

Test Error

Training Error

Overfitting

Model is fitting to structure of data

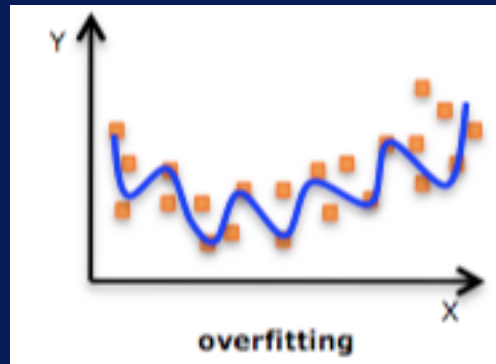Model is fitting to noise in data

Good fit

Overfitting

Overfitting & Generalization
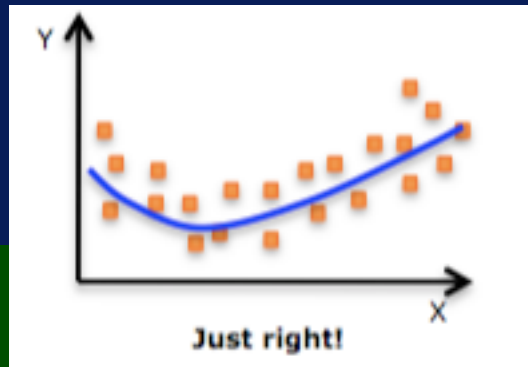
Overfitting ➡ Poor Generalization

Overfitting & Underfitting
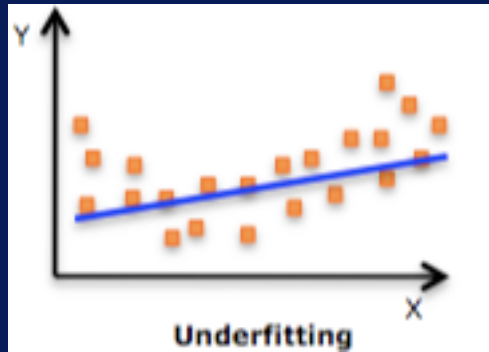
Underfitting

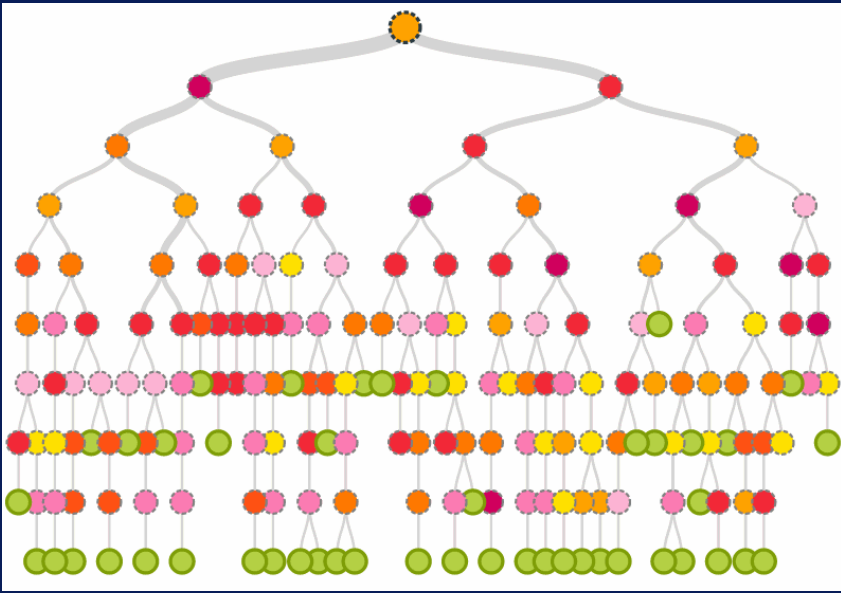Overfitting
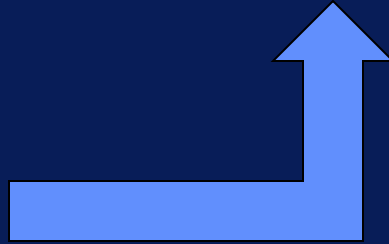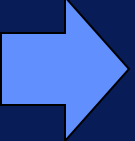
Just Right

What Causes Overfitting?

Overly complex model

Overfitting

Generalization & Overfitting

Ove~~rfi~~tting ➡ Good Generalization