First Steps in Machine Learning: Misfit Models and Fixes

Dr. Mark Fenner

Author, Educator, Consultant

Pearson

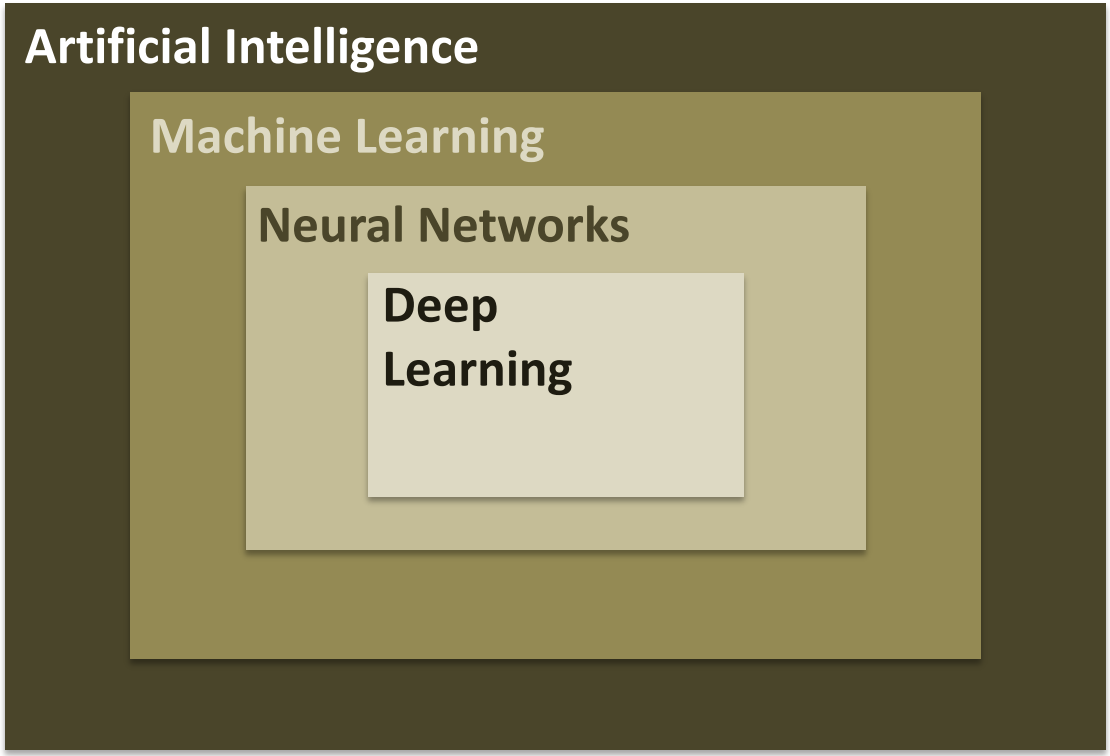
Part I: Introduction, Terminology, and the Machine Learning Process

Audience Poll

What is your experience level with training and testing models in machine learning? (Choose one.)

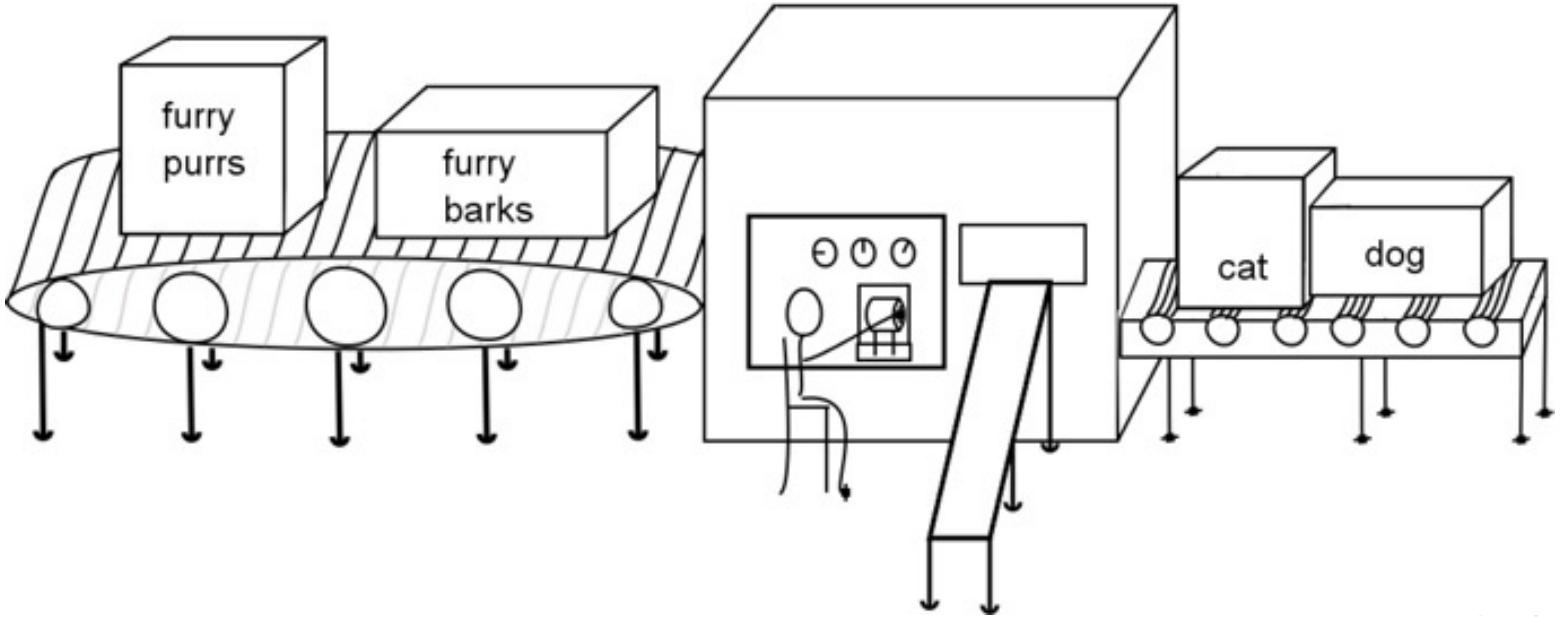
- Well-versed
- Some familiarity
- Never ever

Terminology within Artificial Intelligence



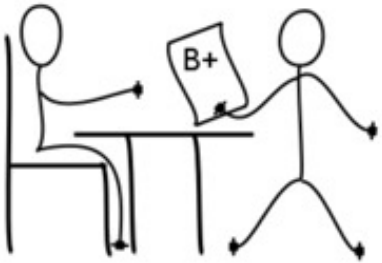
To Measure is Human

Training with Input Features and Output Targets



School Work: Training, Testing, and Evaluating

studying

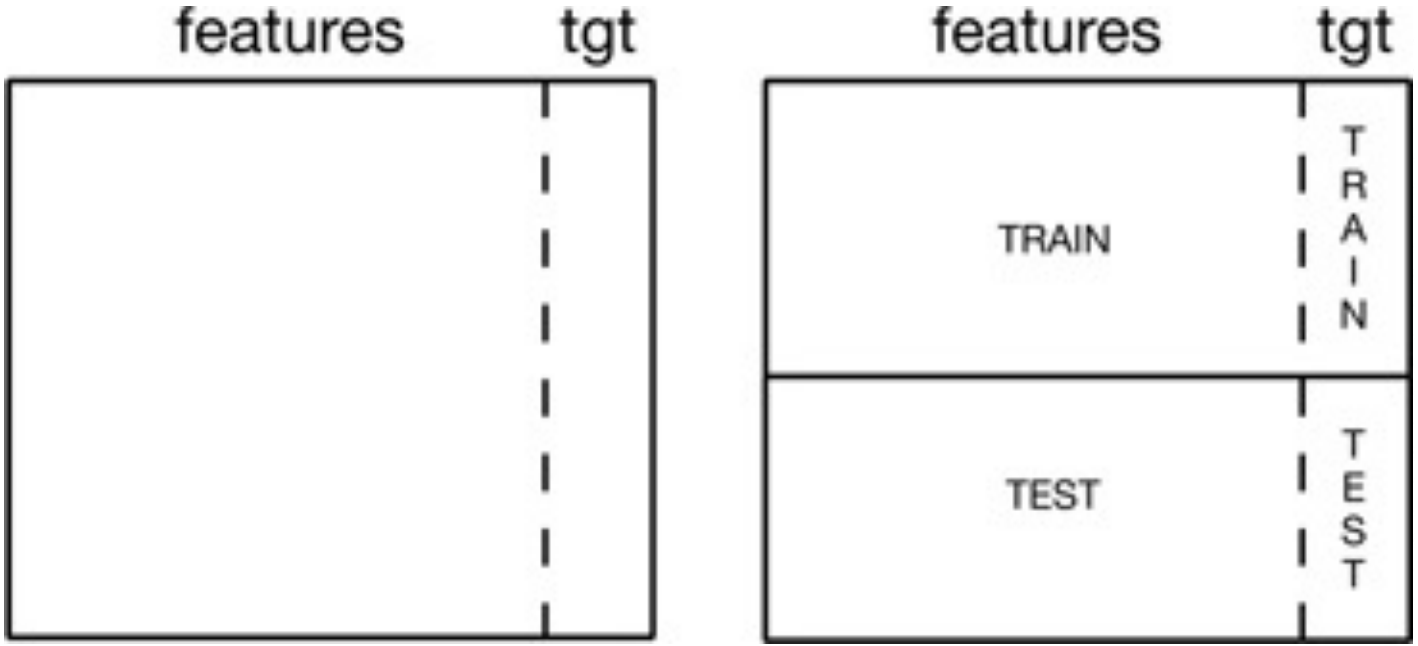


evaluation

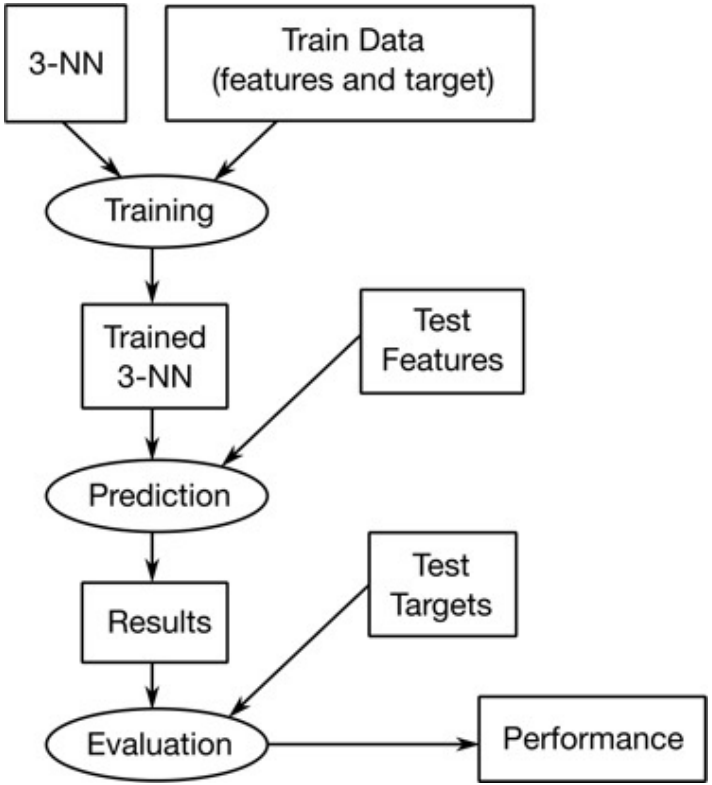
training

testing

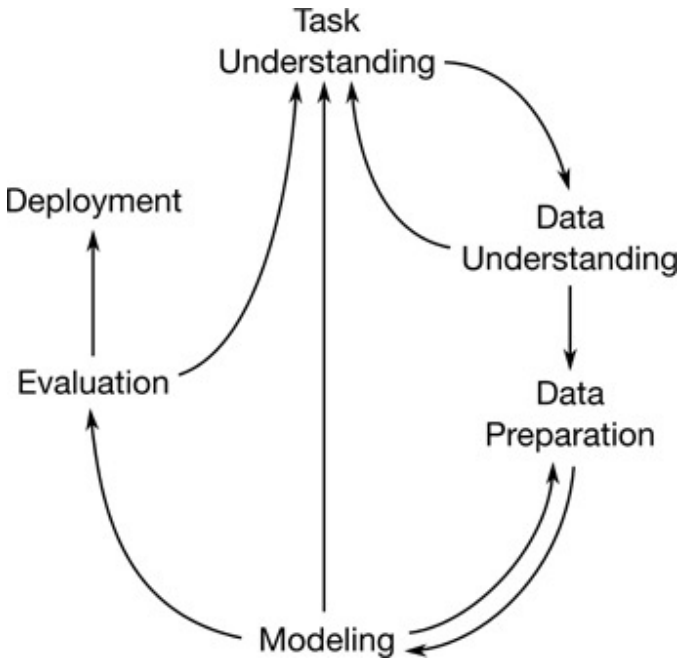
Training and Test Sets



Training, Testing, and Evaluation for 3-Nearest Neighbor



A High-level View of Machine Learning



To the Code Demo!

Exercise

Time

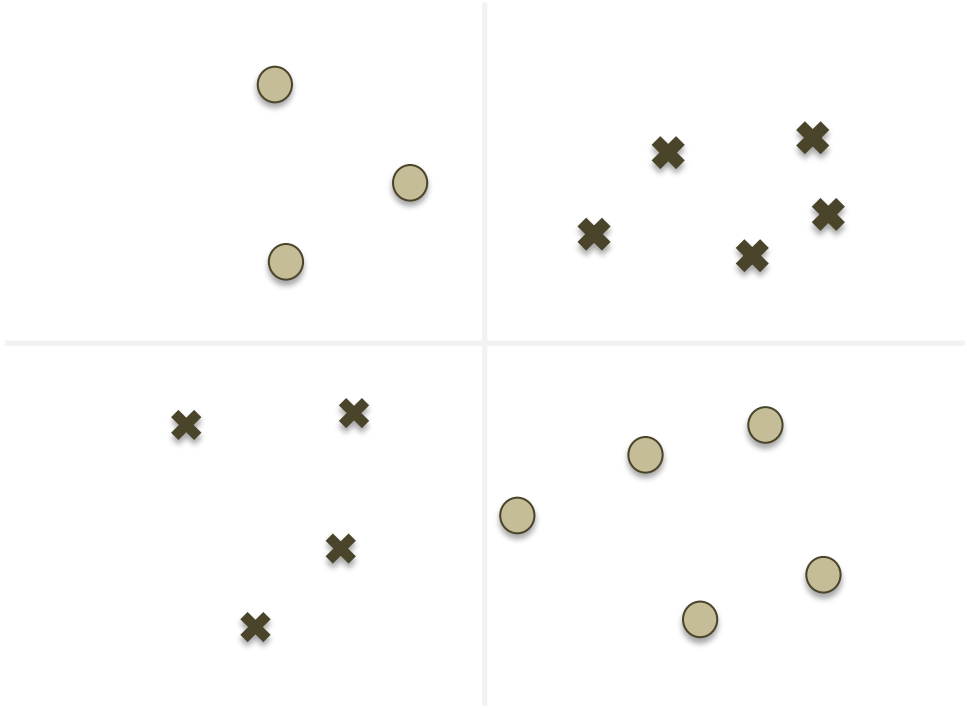
Part II: k-Nearest Neighbors and Decision Trees

Audience Poll

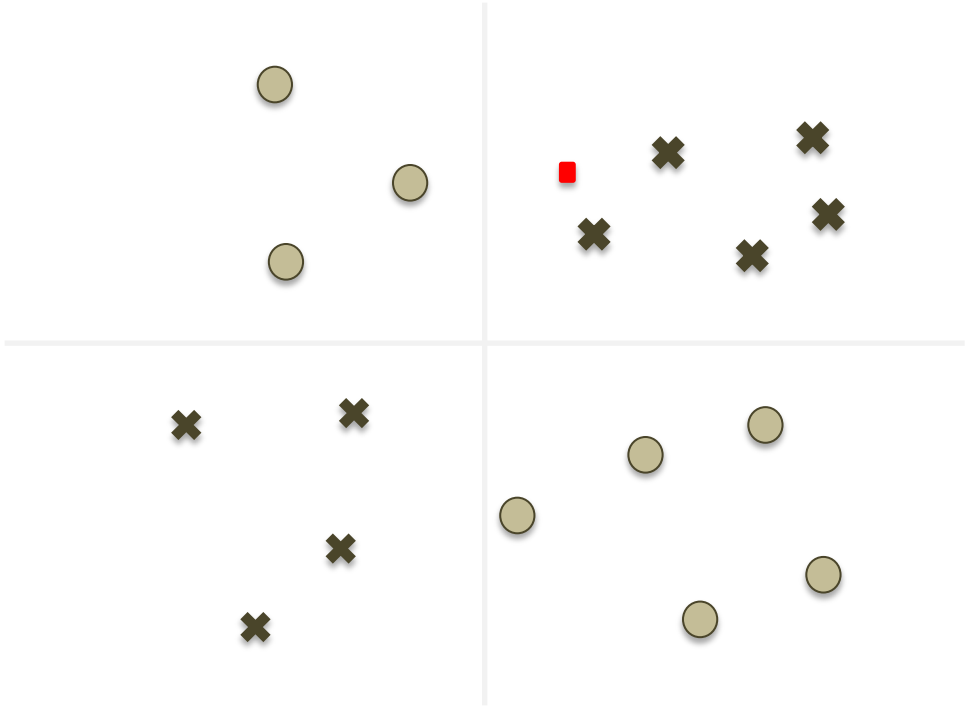
What is your experience level with models like decision trees and nearest neighbors? (Choose one.)

- Well-versed
- Some familiarity
- Never ever

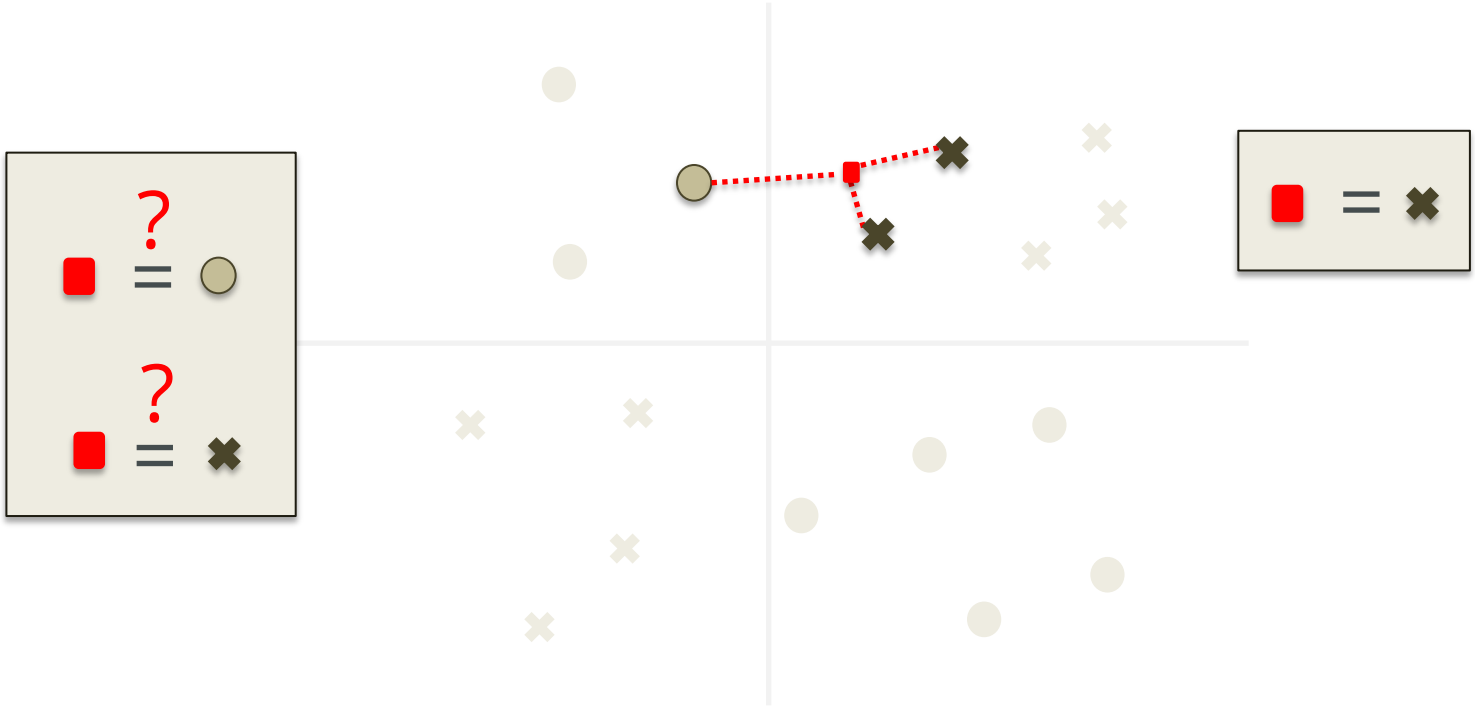
Simple Classification Data



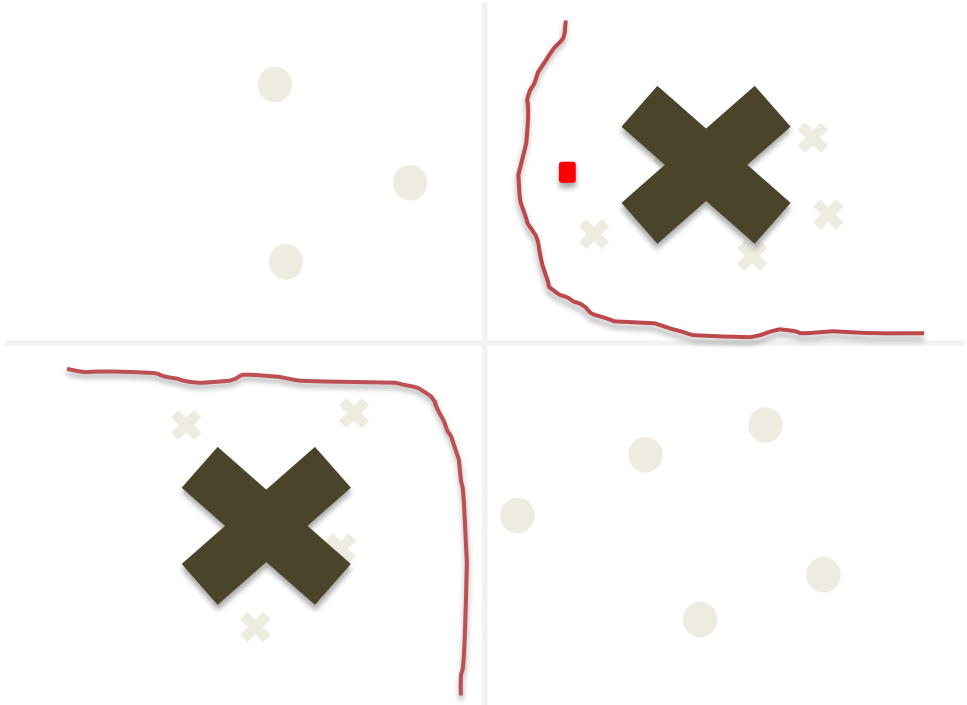
3-Nearest Neighbor (1 of 3)



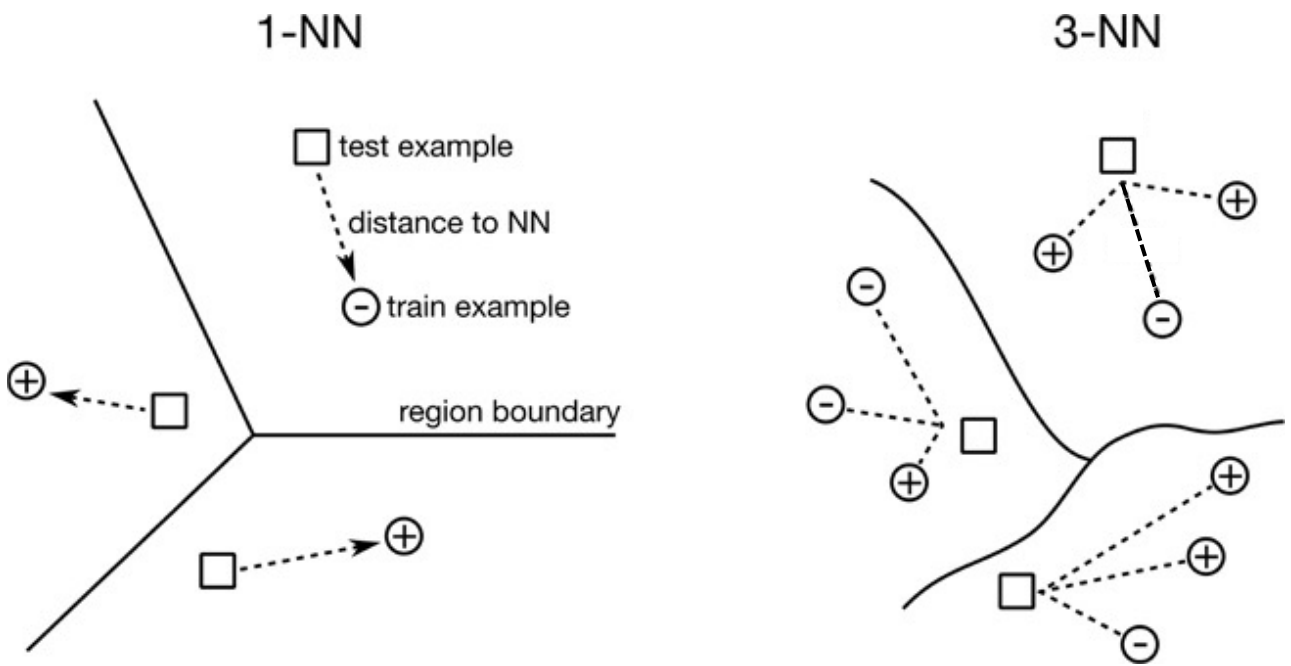
3-Nearest Neighbor (2 of 3)



3-Nearest Neighbor (3 of 3)

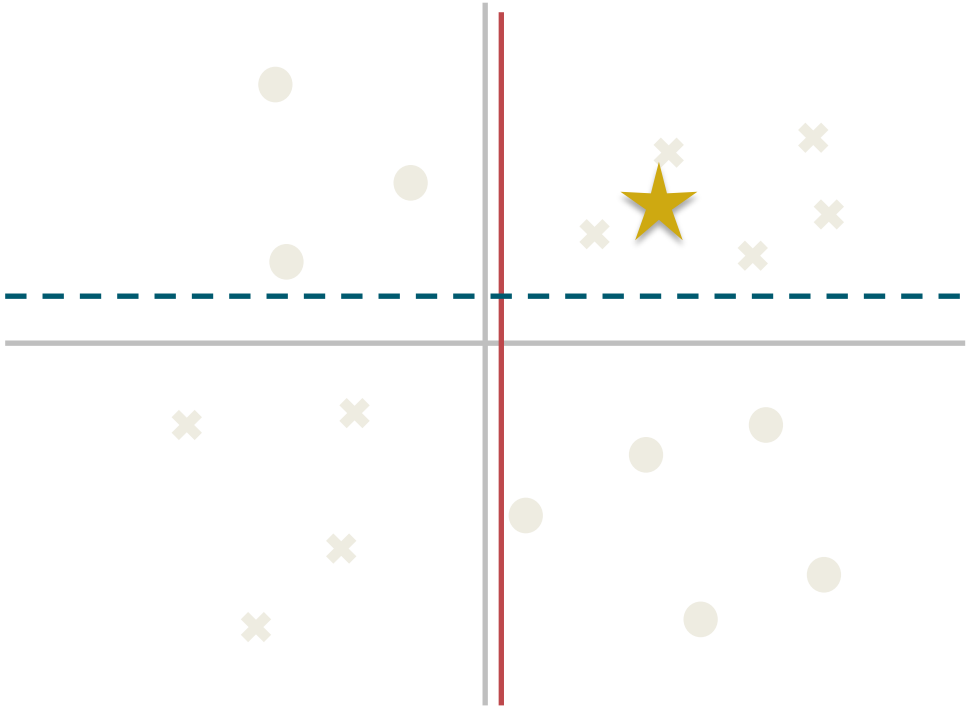


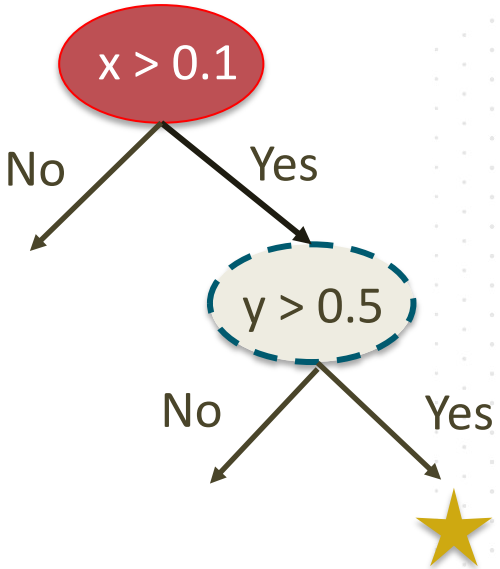
Memorization in k-Nearest Neighbors



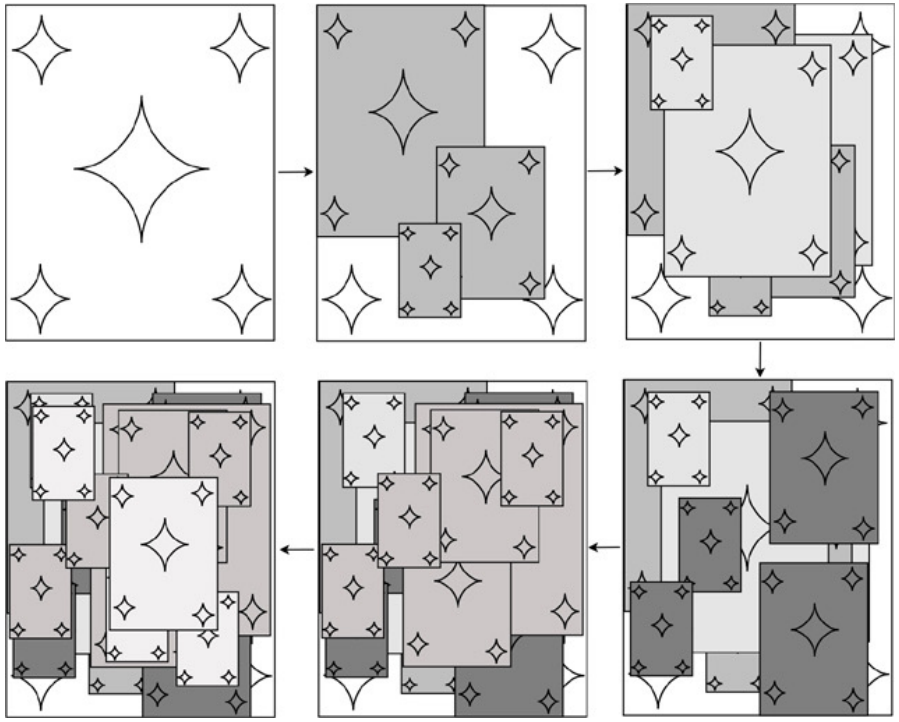
Decision Tree

DT 1 (with data)



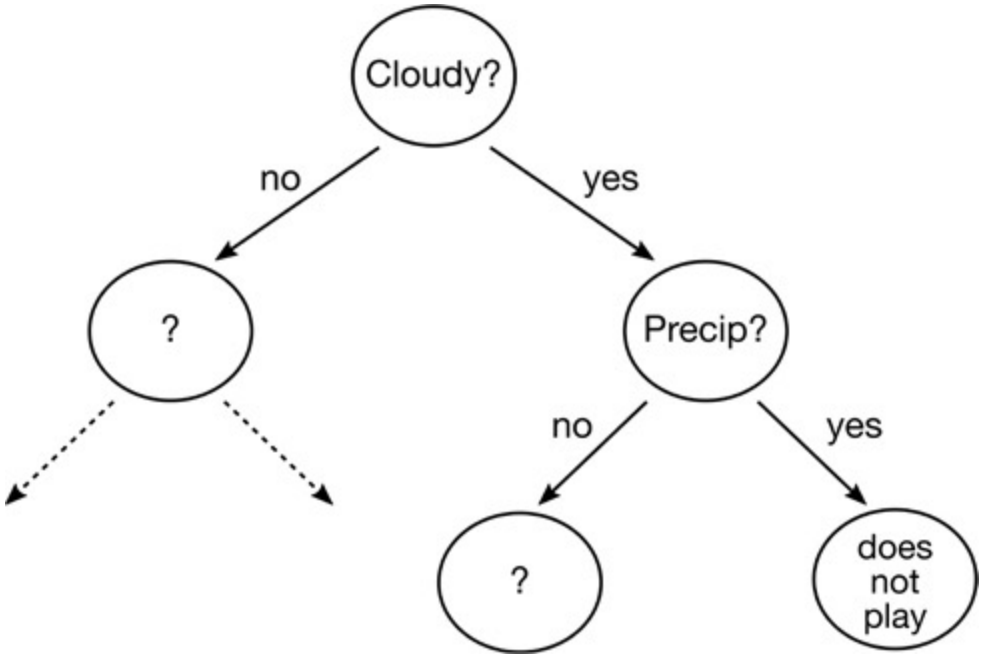


Overlapping Rectangles (1 of 2)

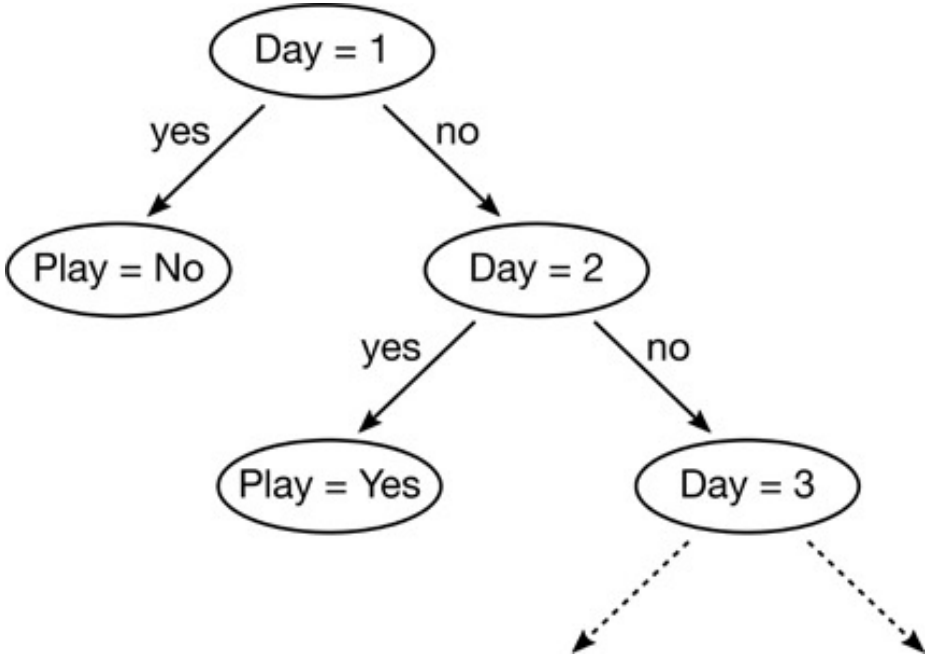


Overlapping Rectangles (2 of 2)

Decisions about Tennis

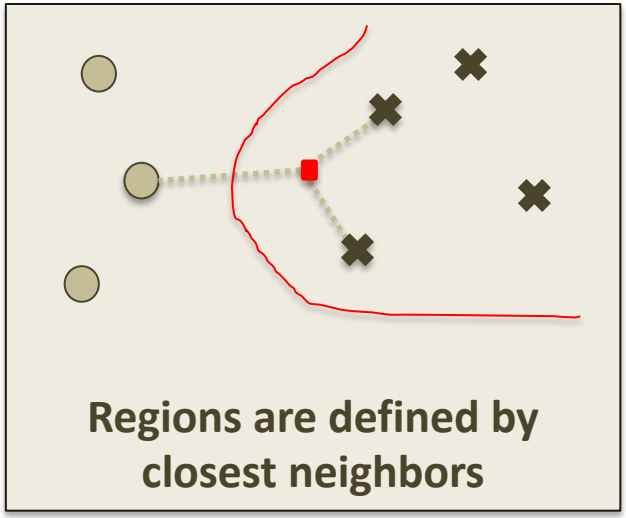


Memorization in Decision Trees



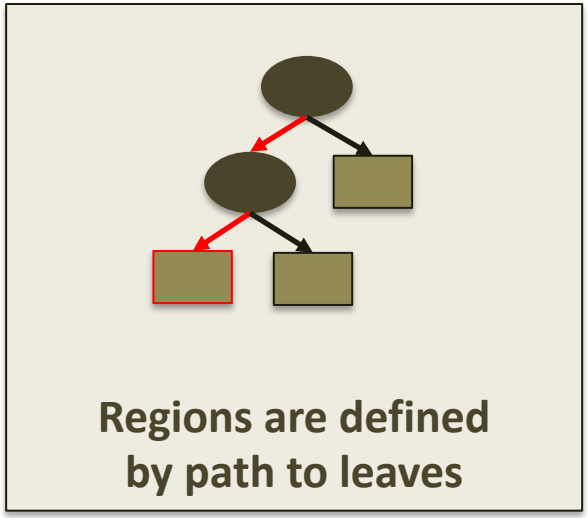
Defining Regions: k-Nearest Neighbor and Decision Trees

k-NN



Regions are defined by closest neighbors

Tree



Regions are defined by path to leaves

Classification = majority vote
Regression = measure of center

To the Code Demo!

Exercise

Time

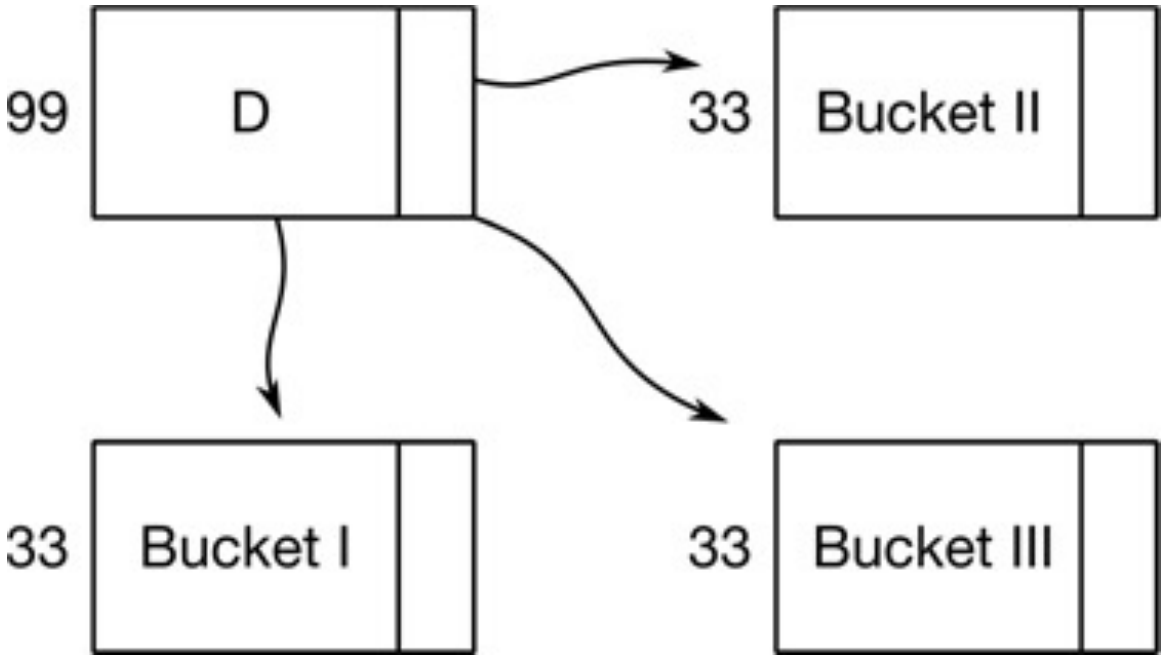
Part III: Measuring Models with Cross-Validation and Fixing Misfits

Audience Poll

What is your experience level with cross-validation and improving model performance? (Choose one.)

- Well-versed
- Some familiarity
- Never ever

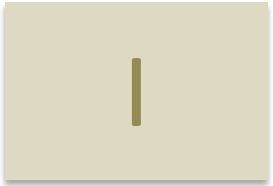
Cross-Validation Splits the Data



Setting Up Cross-Validation

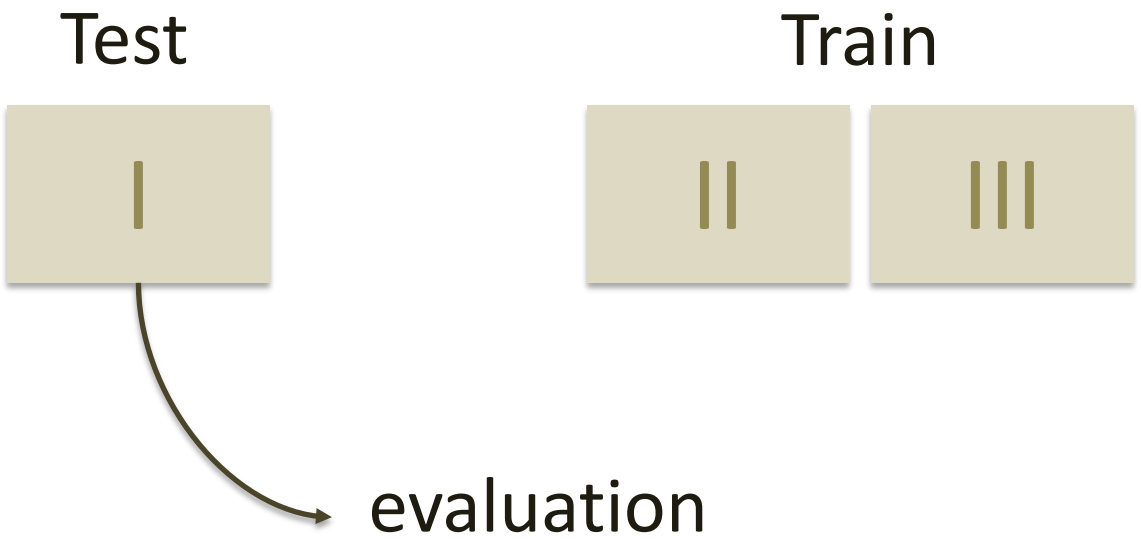
Cross-Validation Setup (1 of 3)

Test

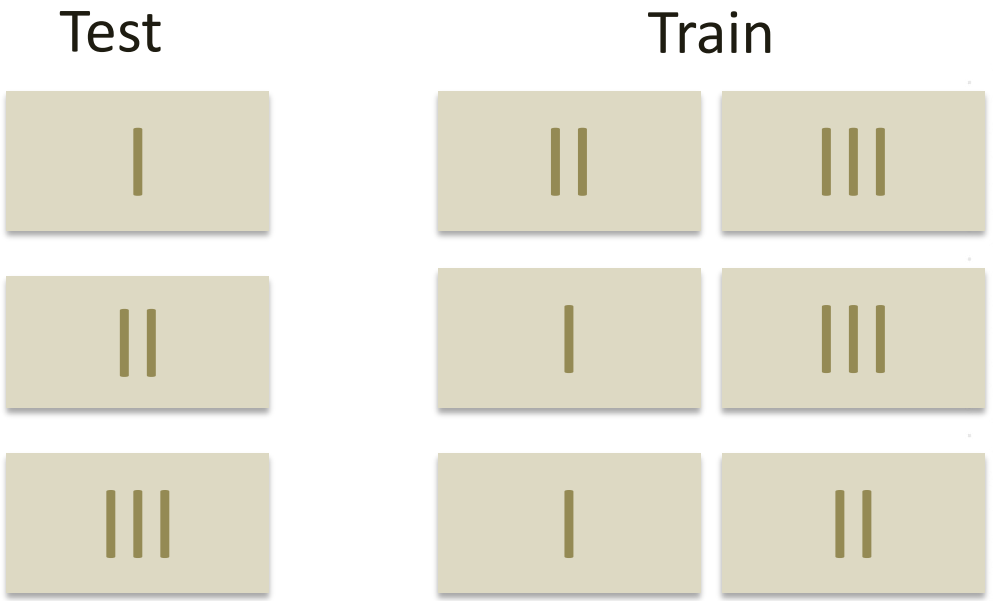


Cross-Validation Setup (2 of 3)

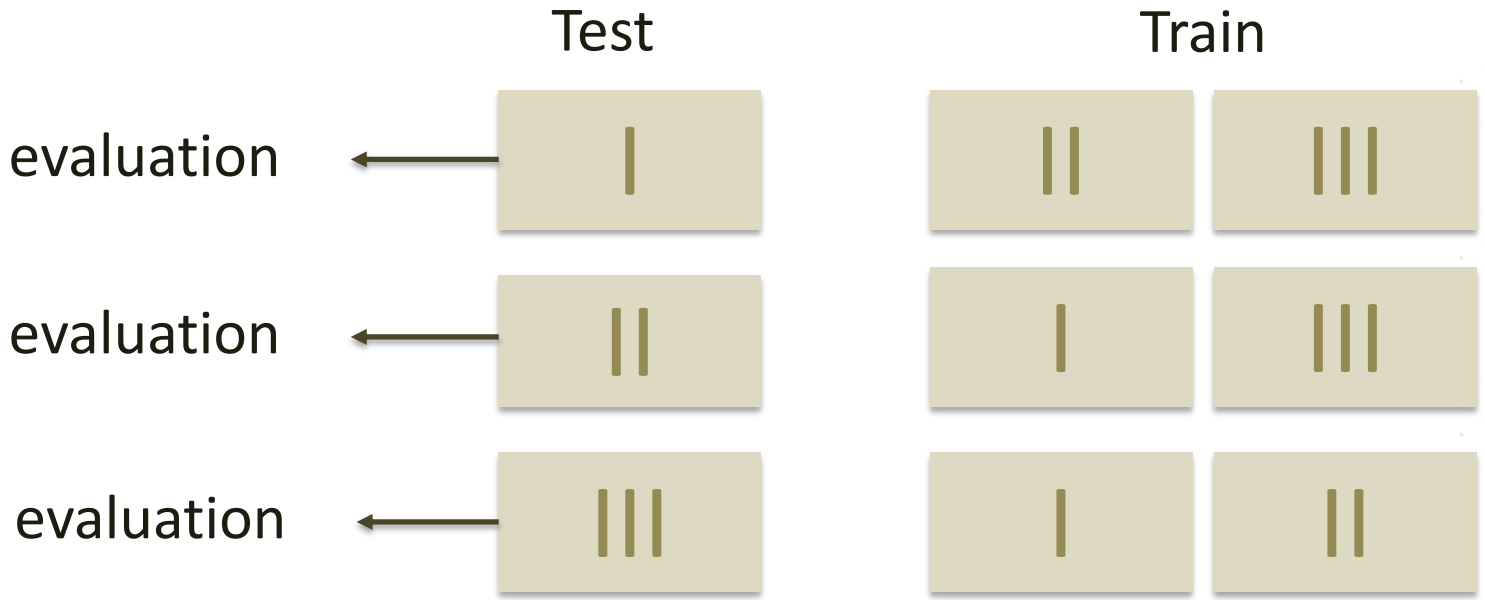
Cross-Validation Setup (3 of 3)



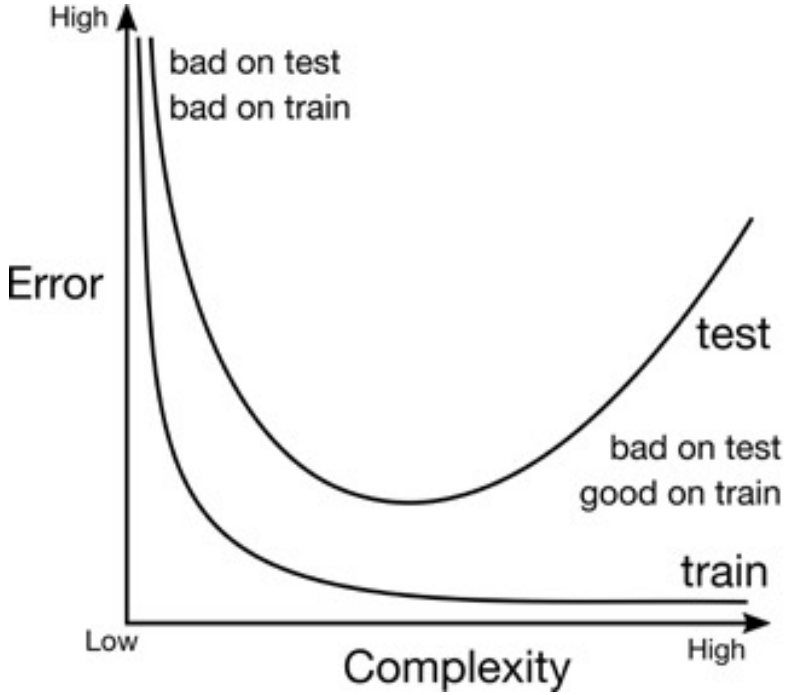
Cross-Validation (1 of 2)

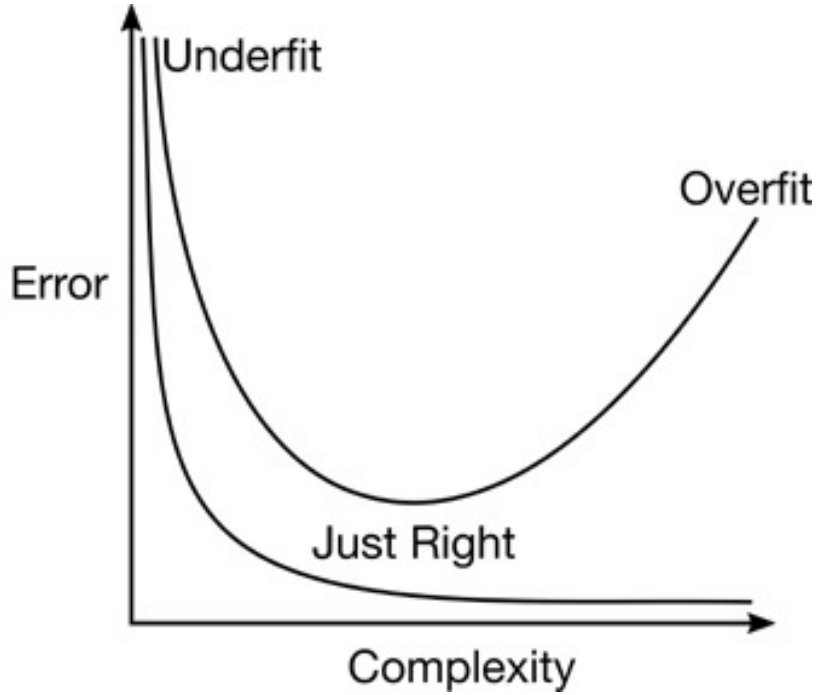


Cross-Validation (2 of 2)

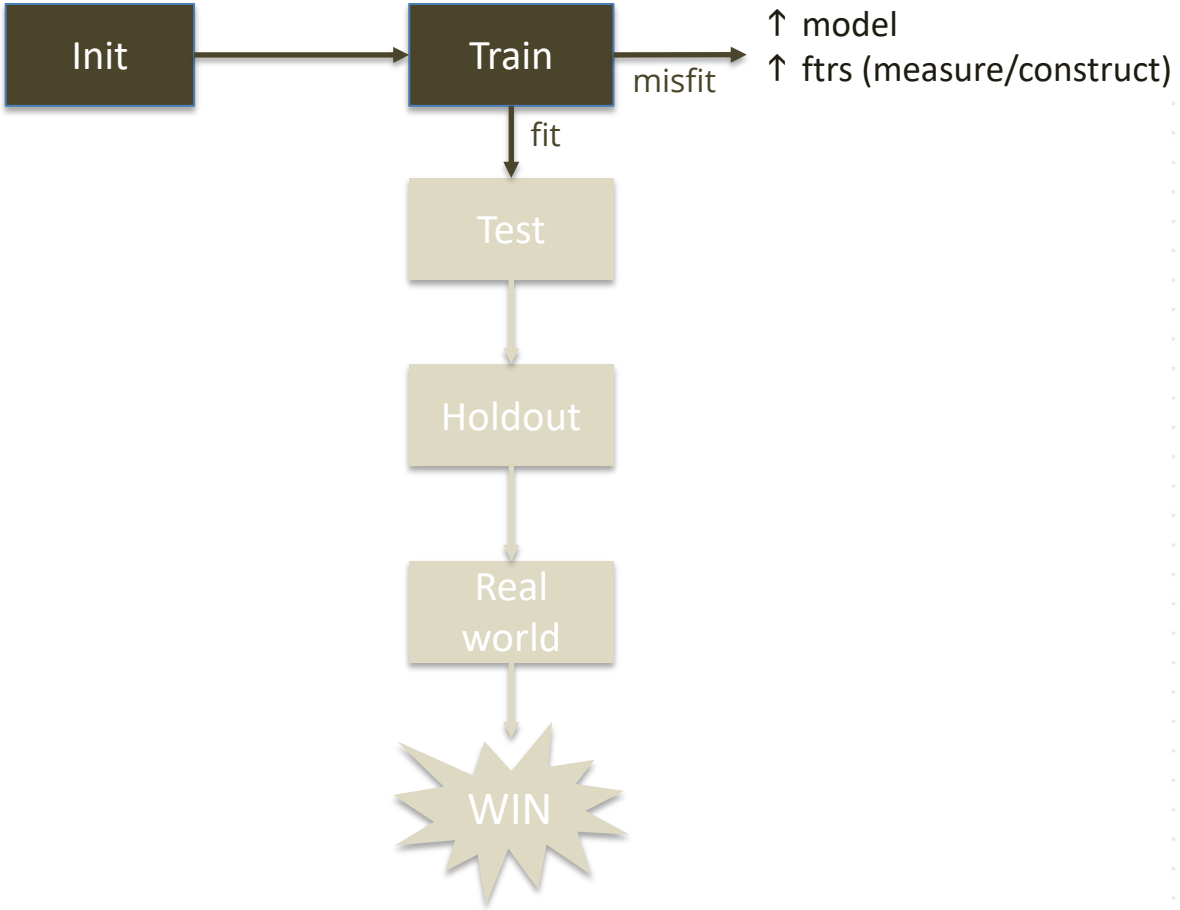


Overfitting and Underfitting

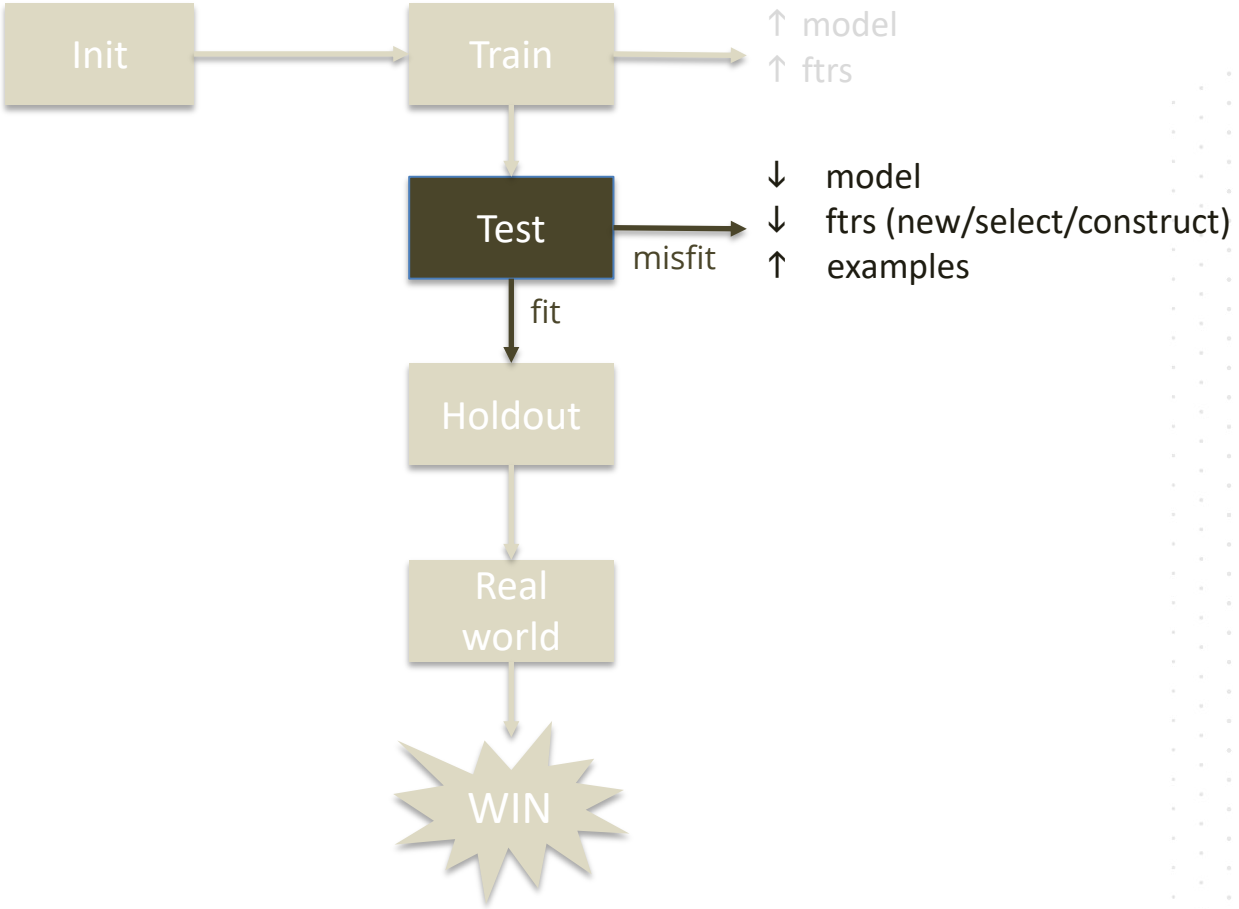




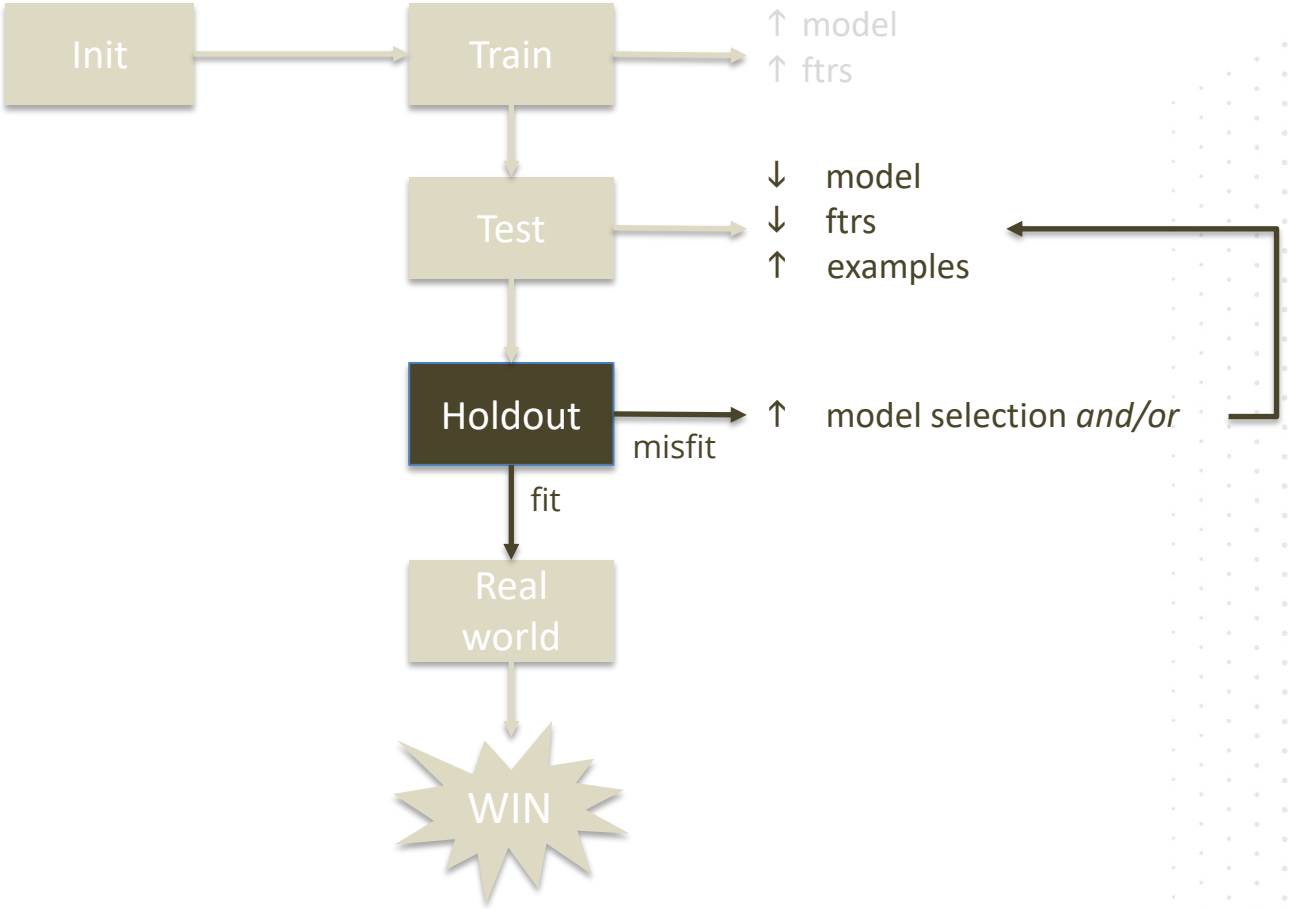
Fit and Misfits (1 of 4)



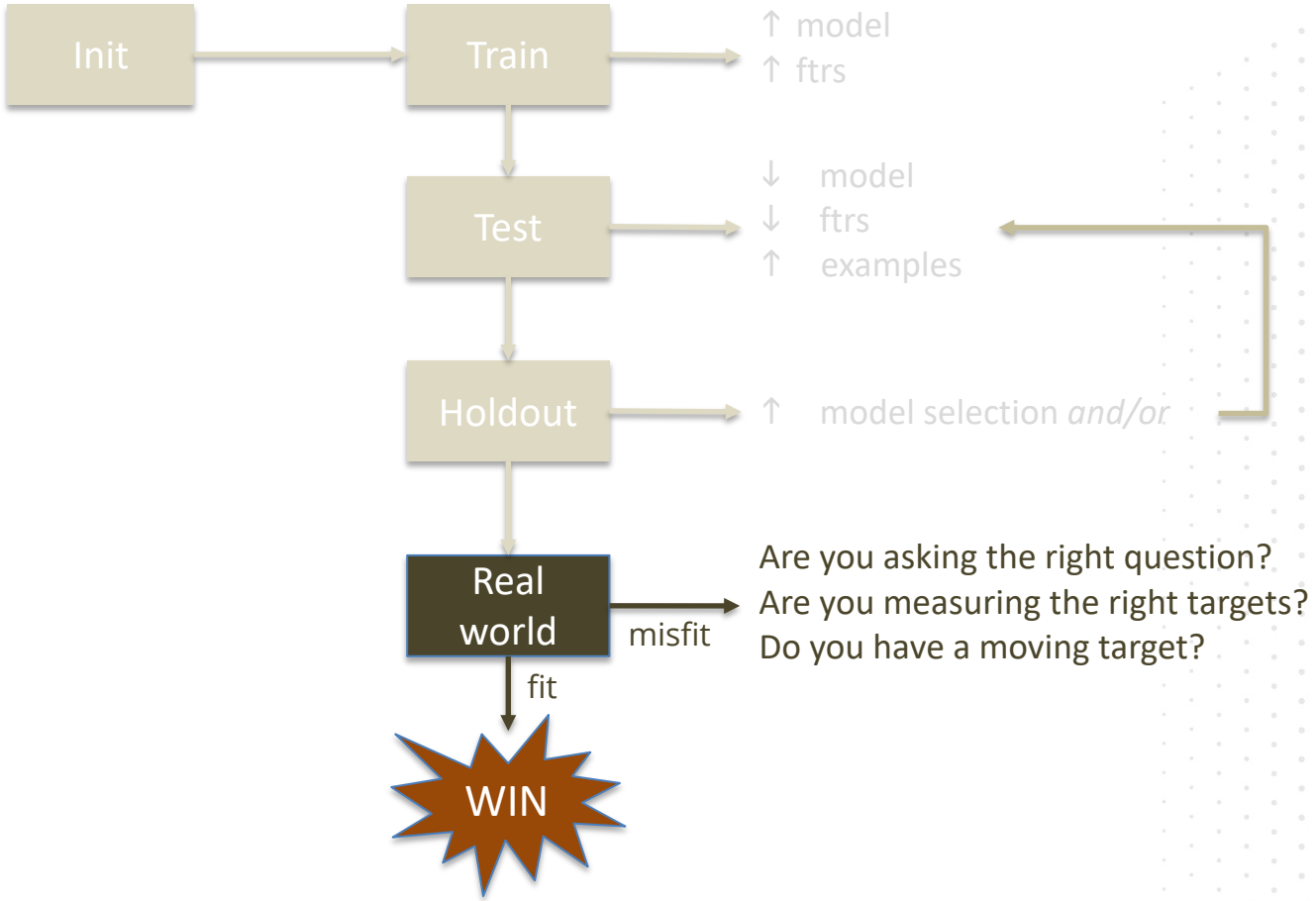
Fit and Misfits (2 of 4)



Fit and Misfits (3 of 4)



Fit and Misfits (4 of 4)



To the Code Demo!

Exercise

Time

Review of Topics

- Machine learning terminology and process
- Training and testing
- k-Nearest neighbors and decision trees
- Cross-validation
- Overfitting and underfitting
- Fixing misfit models