

# Nearest Neighbor Algorithm

---

Poor man's Neural Net

## Goals

- What is Nearest Neighbor Algorithm
- How to Learn
- How to Use
- When to Use
- Decision Boundaries: Voronoi
- Evaluation: Cross-validation

# Basics

- Basic Assumption
- Similarity Hypothesis (Bias):  
Near Examples  $\rightarrow$  Same class.
- Example Representation:
- list of numbers or nominals + label  
tall, male, muscular, 185, athlete
- Learning
- store all the instances
- Classification:  
Given unknown instance  $x$ ,  
find closest instance  $y$ .  
Guess class of  $y$  is class of  $x$ .
- Batch or Incremental
- Understanding: divides space into convex regions

# Decision Boundaries

- Voronoi Diagram (graphics)
- And
- XOR
- $X > Y$
- General

# Evaluation

## K-fold Cross-validation

- Divide data into 10 (k) pieces
- Learning on 9 pieces
- test on left-out piece  
compute generalization accuracy
- Repeat 10 (k) times
- Average generalization accuracy.

## Extensions/Variations

- Better metric  
statistics, knowledge, ...
- Throw out Noisy examples
- Throw out bad attributes
- Multiple votes
- Regression (guessing a real value)  
method: use average of nearest neighbors.

## Performance

- Sometimes Best algorithm
- Best on Letter recognition (20,000 examples)
- Predicts Secondary Protein Structure
- Molecular Biologist Predict Protein function.