Lecture 35

Nearest Neighbor
Announcements
Nearest Neighbor

How to classify a new individual:

- Find their nearest neighbor: the individual closest to them in the data set
- Assign the new individual the same label as that nearest neighbor

(Demo)
Nearest Neighbor recap

How to classify a new individual:

- Find their **nearest neighbor**: the individual closest to them in the data set
  - (We put data in standard units because scale of one attribute was so different than the other attribute--you will **not** need to do that on your proj3)
  - Compute table of distances from that individual to all other individuals
  - Sort by distance, so that closest is in the first row
- Assign the new individual the **same** label as that nearest neighbor
Decision Boundary

- Partition between the two classes
- Computer figured out that boundary, instead of humans having to “hard code” it: machine learning
Evaluating a Classifier

How do we evaluate whether classifier is doing a good job on all those points where we have no data?
Train vs. Test

All data

Attributes | Labels

Training data

Attributes | Labels

Test data

Attributes | Labels

random partition
Train vs. Test

- Use **training data to create** the classifier
- Use **test data to evaluate** the finished classifier

- **Never** allow classifier to see test data until the very end: think of classifier as a cheater who would be happy to just memorize the answers

(Demo)
Multiple Neighbors

● If data are noisy, asking just the closest neighbor might not be ideal for accuracy
● Instead, ask the $k$ closest neighbors, and take the majority label

(Demo)
We’ve used 2 attributes so far. But nothing special about 2, just have to compute distances in higher dimensional spaces.

(Demo)