CS 440: Introduction to Artificial Intelligence

Lecture 17

Matthew Stone

October 28, 2015

Recap—Kinds of Learning

So far: discrete

- Categorization problems
- World is in one of discrete set of states
- Need to infer state from sensor values

Alternative: continuous

- True state of the world is a quantitative value
- Still want to infer true state from sensor values

Nearest Neighbor Predictions—Recap

- find nearest neighbors—most similar **users**
- get their predictions
- multiply by similarity
- sum up
- normalize by total similarity

Table 2-2. Creating recommendations for Toby

Critic	Similarity	Night	S.xNight	Lady	S.xLady	Luck	S.xLuck
Rose	0.99	3.0	2.97	2.5	2.48	3.0	2.97
Seymour	0.38	3.0	1.14	3.0	1.14	1.5	0.57
Puig	0.89	4.5	4.02			3.0	2.68
LaSalle	0.92	3.0	2.77	3.0	2.77	2.0	1.85
Matthews	0.66	3.0	1.99	3.0	1.99		
Total			12.89		8.38		8.07
Sim. Sum			3.84		2.95		3.18
Total/Sim. Sum			3.35		2.83		2.53

Linear Regression

Mathematical approach for predicting one attribute from another

- Assumes output y is a linear function of input x

$$y = mx + b$$

- Minimize least squared error
- Correlation measures how good linear fit is
  - 1 for line upward
  - 0 for no trend
  - -1 for line downward
  - note: undefined for flat line

Demo for Understanding Regression and Correlation

`http://www.ruf.rice.edu/~lane/stat_sim/reg_by_eye/`

Prediction by Correlation

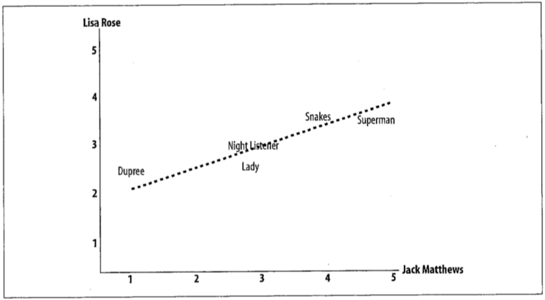

Figure 2-3. Two critics with a high correlation score

Nearest Neighbor Predictions—Twist

- find nearest neighbors—most similar **users**
- get their predictions—use a linear model here!
- multiply by similarity
- sum up
- normalize by total similarity

Table 2-2. Creating recommendations for Toby

Critic	Similarity	Night	S.xNight	Lady	S.xLady	Luck	S.xLuck
Rose	0.99	3.0	2.97	2.5	2.48	3.0	2.97
Seymour	0.38	3.0	1.14	3.0	1.14	1.5	0.57
Puig	0.89	4.5	4.02			3.0	2.68
LaSalle	0.92	3.0	2.77	3.0	2.77	2.0	1.85
Matthews	0.66	3.0	1.99	3.0	1.99		
Total			12.89		8.38		8.07
Sim. Sum			3.84		2.95		3.18
Total/Sim. Sum			3.35		2.83		2.53

Learning Overview

- So far focused on supervised learning
- Also have unsupervised learning
  - No labeled examples
  - Have to infer categories without labels
  - Also called "clustering"

K means

Iterative algorithm

- Assign each point to best cluster
- Move cluster to match points assigned to it

K means

Demo: http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/AppletKM.html

K means analysis

- Attempts to minimize average distance to cluster center
- Greedy search—May find local minima
- Hard to know how many clusters you need
- Tends to make equal-sized clusters

Probabilistic clustering

Probabilistic idea

- Data comes from two underlying categories
- You can't see categories, but
- Categories are reflected in measured similarities

Overall problem

- Simultaneously guess what clusters there are
- Guess which data point goes to which cluster

Iterative algorithm

Find posterior for each point based on current clusters

- use categorization model
- commonly: Gaussian (normal) distribution

Estimate new cluster centers

- Points contribute to all clusters
- Contribution depends on posterior point belongs to cluster

K means

Demo: `http://jormungand.net/projects/misc/em/`