2 handouts: slides, asst 11
Unsupervised Learning

Unsuperv. Learning
- Overview
- Bottom-Up
- RANSAC
- $k$-Means
- An Algorithm
- EM
- Basic Clustering
- Summary
- EOLQs

Unsupervised Learning
modeling = predicting = understanding
clustering
finding ‘structure’ in data
explain the data all-at-once vs piece-by-piece?

repeat

try to explain a minimal amount of the data
see if model fits a decent amount of the data
if so, remove explained data from the set
until hard to find a decent model or not enough data left
Random Sample Consensus (RANSAC)

repeat
  repeat $n$ times
    select $k$ points
    fit model
    find inliers
    if best model has enough inliers
      re-fit model to all inliers
      record model
      re-find inliers
      remove inliers from data
  until best model not good enough or not enough data left
Naive Bayes model: choose class, generate attributes independently

mixture model: choose class, generate data

\[
P(x|\theta) = \sum_k P(C = k|\theta_k)P(x|C = k, \theta_k)
\]

eg, for mixture of Gaussians,

\[
P(x|C = k, \mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi \sigma_k^2}} \exp \left( -\frac{(x - \mu_k)^2}{2\sigma_k^2} \right)
\]
Means represent the center of a cluster/class
Values for the means are the model
Model changes based on the classes assigned to the data

init the $k$ means somehow
repeat until cluster assignments do not change:
  Assign each data point to the mean nearest to it
  Calculate new means for the data assigned to each cluster
Means represent the center of a cluster/class
Values for the means are the model
Model changes based on the classes assigned to the data

init the $k$ means somehow
repeat until cluster assignments do not change:
  Assign each data point to the mean nearest to it
  Calculate new means for the data assigned to each cluster

Example
Means represent the center of a cluster/class
Values for the means are the model
Model changes based on the classes assigned to the data

init the $k$ means somehow
repeat until cluster assignments do not change:
  Assign each data point to the mean nearest to it
  Calculate new means for the data assigned to each cluster

Example

Is the classification optimal?
What is it optimizing?
model parameters $\theta$ (eg, $\mu, \sigma^2, P(C = k)$) 
observed variables $x_j$ 
hidden variables $C_j$

init the $\theta_k$ somehow
repeat until done: 

E: compute expected values of hidden vars: $P(C_j = k|x_j, \theta_k)$ 
    eg by $\alpha P(C = k)P(x_j|C = k, \theta_k)$ 

M: maximize data likelihood using current estimates: 
    $\theta_k$, with each $x_j$ weighted by $P(C_j = k|x_j)$, eg by
model parameters $\theta$ (eg, $\mu, \sigma^2, P(C = k)$)
observed variables $x_j$
hidden variables $C_j$

init the $\theta_k$ somehow
repeat until done:
  E: compute expected values of hidden vars: $P(C_j = k|x_j, \theta_k)$
    eg by $\alpha P(C = k)P(x_j|C = k, \theta_k)$
  M: maximize data likelihood using current estimates:
    $\theta_k$, with each $x_j$ weighted by $P(C_j = k|x_j)$, eg by

$$\theta \leftarrow \arg\max_{\theta} \sum_z P(Z = z|x, \theta)P(x, Z = z|\theta)$$

greedy increase of data likelihood
Expectation-Maximization

Features

- Probabilistic clustering
- Explicit model
- Locally optimal

Issues

- Number of classes (means, Gaussians, etc.)
- Local maxima
dendrogram
$O(n^2)$ vs $O(kn)$
AutoClass
supervised learning: learning a function or a density
unsupervised learning: explaining data
reinforcement learning: learning how to act
What question didn’t you get to ask today?
What’s still confusing?
What would you like to hear more about?

Please write down your most pressing question about AI and put it in the box on your way out.

Thanks!