handout: slides
Unsupervised Learning
modeling = predicting = understanding
clustering
finding ‘structure’ in data
Bottom-Up Unsupervised Learning

explain the data all-at-once vs piece-by-piece?

repeat

make a model to explain a minimal amount of data
check how much of the total data the model explains
repeat until model fits a decent amount of the data
when found, remove explained data from the set
until hard to find a decent model or not enough data left
given data, find a set of explanatory models:

repeat
  repeat many times
    randomly pick minimum data to fit model
    find inliers
    repeat until no change
      fit model to inliers
      find new inliers
    if best model has enough inliers
      record model
      remove inliers from data
  until best model not good enough or not enough data left
Break

- Wed May 8: 9-noon: project presentations
- Mon May 13 3pm: final paper
two hardcopies of paper, one hardcopy of source PDF and tarball via email
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,

\[ P(x|C = k, \mu_k, \sigma^2_k) = \frac{1}{\sqrt{2\pi\sigma^2_k}} \exp \left( -\frac{(x - \mu_k)^2}{2\sigma^2_k} \right) \]
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

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

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
EOLQs

- 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!