Statistical learning

Chapter 20a-1, Section 3
Unsupervised Clustering

Clustering - group data into coherent groups, in the absence of group labels.

K-means clustering
   0. Start with K arbitrary cluster centers (means)
   1. Classify each pattern according to closest mean
   2. Recalculate the means based on classification
   3. Go to step 1 and repeat until convergence (or n iterations)

EXERCISE: Go through example on the board

Learning mixture of Gaussians:
   - Similar to K-means, but more general
Learning mixture of Gaussians

See Figure 20.8

Algorithm:

0. Start with some initial estimate of Gaussian parameters:
   \( \{\mu_1 \cdots \mu_k\}, \{\Sigma_1 \cdots \Sigma_k\} \)

1. E-step: For each data point \( x_j \), compute probabilities
   \[
   p_{ij} = P(C = i|x_j) = \alpha P(x_j|C = i)P(C = i)
   \]
   \[
   p_i = \sum_j p_{ij}
   \]

2. M-step: Compute the new mean, covariance and component weights:
   \[
   \mu_i = \sum_j p_{ij}x_j/p_i
   \]
   \[
   \Sigma_i = \sum_j p_{ij}(x_j - \mu_i)(x_j - \mu_i)^T/p_i
   \]
   \[
   w_i = p_i
   \]

3. Go to step 1 and repeat until convergence (or n iterations)
Nearest-Neighbor Density Estimation

Density Estimation - use a non-parametric neighborhood average to estimate the probability density at a given point.

Fixed kernel and K-NN methods. (See Figure 20.12 and 20.13)

EXERCISE: Draw example on board.
Nearest-Neighbor Classification

K-NN classification
Classify new examples as the majority of their neighbors.

Distance metrics
- Euclidean
- Scaled axes (Scale each feature based on variance)
- Mahalanobis (scale each feature based on covariance)
- Hamming distance (Discrete distance = num different features)

Problems
- Memory requirements
- Speed (Kd-trees, voxel grids used)
- High dimensional data

EXERCISE: Draw example on board
Kernel-Based Methods

Kernel-method

Each training instance has its own Kernel Function.
Density at each point is just sum of all the kernels:

\[ P(x) = (1/N) \sum_1^N K(x, x_i) \]

Benefits

Can be smoother than KNN
Allows weighted voting

Drawbacks

Simple method requires us to look at ALL the training set
If kernels are fixed size, they don’t adjust to density changes.

See figure 20.14.