1 Conceptual review: GMMs, K-means, and K-mediods

Recap the steps of the k-means algorithm:
1: Fixing the cluster centers, assign points to nearest clusters
2: Given the point assignments, re-estimate cluster centers
Termination: No points change clusters in next iteration

1. What is a key difference between mixture modeling and k-means?

2. What is K-medoids, and how is it different from k-means? Discuss their pros and cons.

3. Show why the point s that minimizes $\sum_i \|s - x_i\|_1$ is the median of all points.

2 Expectation Maximization for GMMs

Let $\lambda = \mu_1, \mu_2, ..., \mu_k, \Sigma_1, ..., \Sigma_k, p_1, ...p_k$.
The log likelihood is $\ell(\lambda \mid x_1, x_2, ..., x_m = \sum_{i=1}^{m}(x_i \mid \lambda) = \sum_i \log \sum_{j=1}^{k} p_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)$.

1. Calculate the posterior probability $P(y = j \mid x_i, \lambda)$

2. Apply MLE to update the parameter μ_j

3 Conceptual review: Hierarchical Clustering

Recap the steps of the Bottom-Up Agglomerative Clustering algorithm:

Start: each object is in a separate cluster

Repeat:

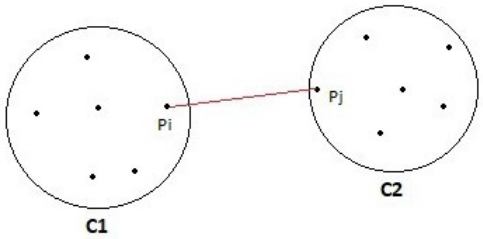
1: Join the most similar pair of clusters

2: Update the similarity of the new cluster to the other clusters

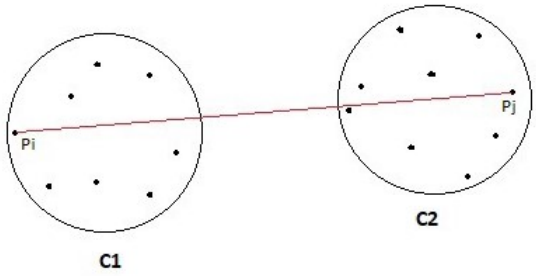
Termination: Only one cluster remains

There are multiple algorithms used to define similarities between two clusters:

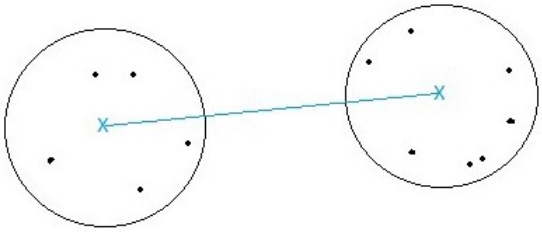
- Single-linkage: Uses similarity between their closest members (nearest neighbor)



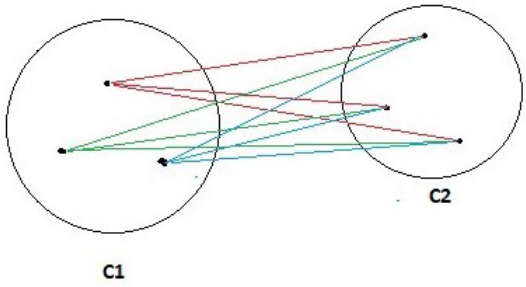
- Complete-linkage: Uses similarity between their furthest members (furthest neighbor)



- Centroid: Uses similarity between the centers of gravity

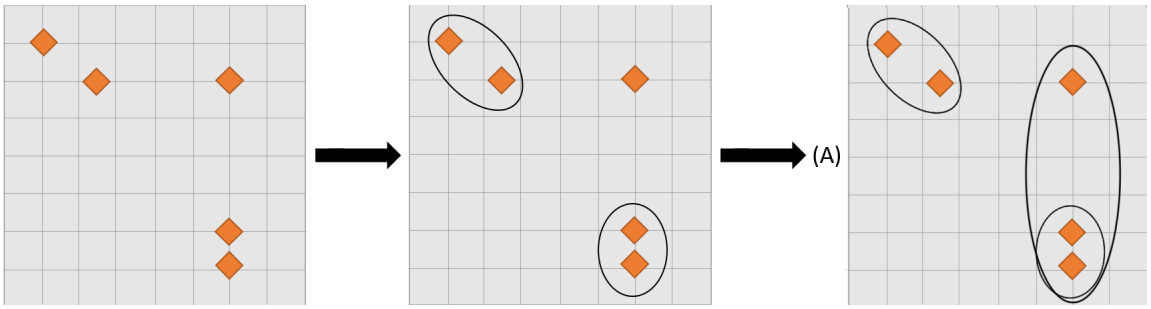


- Average-linkage: Uses average similarity of all cross-cluster pairs



4 Hierarchical Clustering Example

Consider the five data points shown below (placed at the center of the diamond representation). Assume the clustering shown in the middle figure was generated at some iteration of hierarchical clustering.

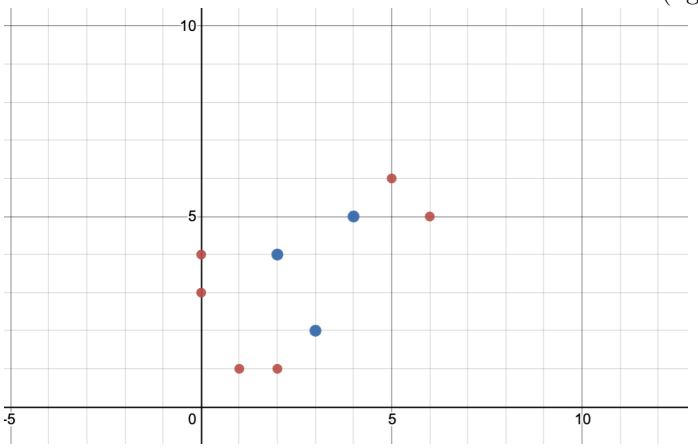


1. Given only the middle figure, we can assume that the clustering technique is [single, complete, average, not enough info to tell]

2. Assume that the clustering in (A) (rightmost figure) was a result of an extra iteration. The clustering algorithm used is [single, complete, average, not enough info to tell].

5 Expectation Maximization Example

Consider the set of points $\{(2,1),(1,1),(6,5),(5,6),(0,4),(0,3)\}$. Consider the initializations of cluster centers $\{(3,2),(2,4),(4,5)\}$. Assume that the covariances for the clusters are given by the identity matrix, and the initial class distribution is the uniform distribution. (figure below)



Perform one iteration of expectation maximization on this dataset.