CLUSTERING / GAUSSIAN MIXTURE MODEL
CLUSTERING

• Group a collection of points into clusters

• In “supervised methods”, the outcome (or response) is based on various predictors.

• In clustering, we want to extract patterns on variables without analyzing a specific response variable.

• This is a form of “unsupervised learning”
CLUSTERING

• The points in each cluster are closer to one another and far from the points in other clusters.
DATA POINTS

• Each of the data points belong to some n-dimensional space.
Given measurements $x_{ij}$ for $i = 1, \ldots, N$ observations over $j = 1, \ldots, p$ predictors.

Define dissimilarity, $d_j(x_{ij}, x'_{ij})$

- We can define dissimilarity between objects as

$$d(x_i, x'_i) = \sum_{j=1}^{p} d_j(x_{ij}, x'_{ij})$$

- The most common distance measure is squared distance

$$d_j(x_{ij}, x'_{ij}) = (x_{ij} - x'_{ij})^2$$
DISSIMILARITY MEASUREMENTS

• Absolute difference

\[ d_j(x_{ij}, x'_{ij}) = |x_{ij} - x'_{ij}| \]

• For categorical variables, we could set

\[ d_j(x_{ij}, x'_{ij}) = 0 \text{ if } x_{ij} = x'_{ij} \]

\[ 1 \text{ otherwise} \]
K-MEANS CLUSTERING

• A commonly used algorithm to perform clustering

• Assumptions:
  • Euclidean distance,

\[ d(x_i, x_i') = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 = ||x_i - x_i'||^2 \]

• K-means partitions observations into K clusters, with K provided as a parameter.
K-MEANS CLUSTERING

• Given some clustering or partition, \( C \), the cluster assignment of observation, \( x_i \) to cluster \( k \in \{1, \ldots, K\} \) is denoted as \( C(i) = k \).

• K-means seeks to minimize a clustering criterion measuring dissimilarity of observations assigned to each cluster.
K-MEANS OBJECTIVE FUNCTION

• We want to minimize within-cluster dissimilarity.

\[ W = \sum_{k=1}^{K} \sum_{i=1}^{N} ||x_{ik} - \bar{x}_k||^2 \]

where \( \bar{x}_k \) is the centroid of the cluster \( k \)

• The criteria to minimize is the total distance given by each observation to the mean(centroid) of the cluster to which the observation is assigned.
K-MEANS - ITERATIVE ALGORITHM

1. Initialize by choosing K observations as centroids.
   \[ m_1, m_2, \ldots, m_k \]

2. Assign each observation i to the cluster with the nearest centroid, i.e,
   \[ \min_{1 \leq k \leq K} ||x_i - m_k||^2 \]

3. Update centroids \[ m_k = \bar{x}_k \]

4. Iterate steps 2 and 3 until convergence.
K-MEANS - ITERATIVE ALGORITHM
- CLUSTERS
K-MEANS - ITERATIVE ALGORITHM
- CLUSTERS
Example

K-means clustering applied to the color vectors of pixels in RGB color-space

Fig. 3 [1]
Image Segmentation by K-Means

- Select a value of K
- Select a feature vector for every pixel (color, texture, position, or combination of these etc.)
- Define a similarity measure between feature vectors (Usually Euclidean Distance).
- Apply K-Means Algorithm.
- Apply Connected Components Algorithm.
- Merge any components of size less than some threshold to an adjacent component that is most similar to it.

* From Marc Pollefeys COMP 256 2003
Results of K-Means Clustering:

K-means clustering using intensity alone and color alone

* From Marc Pollefeys COMP 256 2003
K means: Challenges

• Will converge
• But not to the global minimum of objective function
• Variations: search for appropriate number of clusters by applying k-means with different k and comparing the results
K means: Challenges
K means: Challenges
K means: Challenges
Choosing $k$
GAUSSIAN MIXTURE MODEL
Notation: Normal distribution 1D case

$N(\mu, \sigma)$ is a 1D normal (Gaussian) distribution with mean $\mu$ and standard deviation $\sigma$ (so the variance is $\sigma^2$.)
Notation: Normal distribution 1D case

$N(\mu, \sigma)$ is a 1D normal (Gaussian) distribution with mean $\mu$ and standard deviation $\sigma$ (so the variance is $\sigma^2$).

$$N(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$
Multivariate Normal distribution

\[ 
\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} 
\]

- \( x \) is a D dimensional vector
- \( \mu \) is a D-dimensional mean vector
- \( \Sigma \) is a D x D covariance matrix
Surface Plot

\[ N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \]
Uni-modal dataset
Multi-modal dataset
Multi-modal dataset
Multi-modal dataset

\[ \mathcal{N}(x | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} \]
Multi-modal dataset
Gaussian Mixtures Model

A linear combination of Gaussian distributions forms a superposition

Formulated as a probabilistic model known as mixture distribution
Gaussian Mixtures Model
Gaussian Mixtures Model

\[
p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x | \mu_k, \Sigma_k)
\]
Gaussian Mixtures Model
Gaussian Mixtures Model

• We have a linear combination of several Gaussians

• Each Gaussian is a cluster, one of K clusters

• Each cluster has a mean and covariance

• Mixing probability,
Gaussian Mixtures Model

Parameters - $\mu, \Sigma, \pi$

\[
\sum_{k=1}^{K} \pi_k = 1 \quad ; \quad 0 \leq \pi_k \leq 1
\]

\[
p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)
\]

\[
\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}
\]

$x$ is a $D$ dimensional vector

$\mu$ is a $D$-dimensinal mean vector

$\Sigma$ is a $D \times D$ covariance matrix
Maximum Likelihood Estimate

\[
N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}
\]

\[
\ln N(x | \mu, \Sigma) = -\frac{D}{2} \ln 2\pi - \frac{1}{2} \ln \Sigma - \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)
\]

Once Optimal values of the parameters are found,

the solution will correspond to the Maximum Likelihood Estimate (MLE)