EM

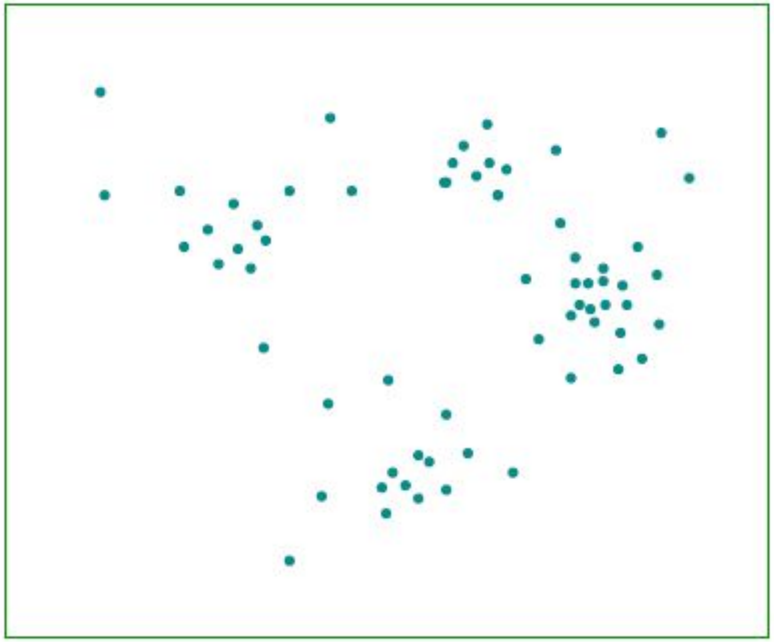
MGTF 495

Mixtures of Gaussians

Idea: model each cluster by a Gaussian:

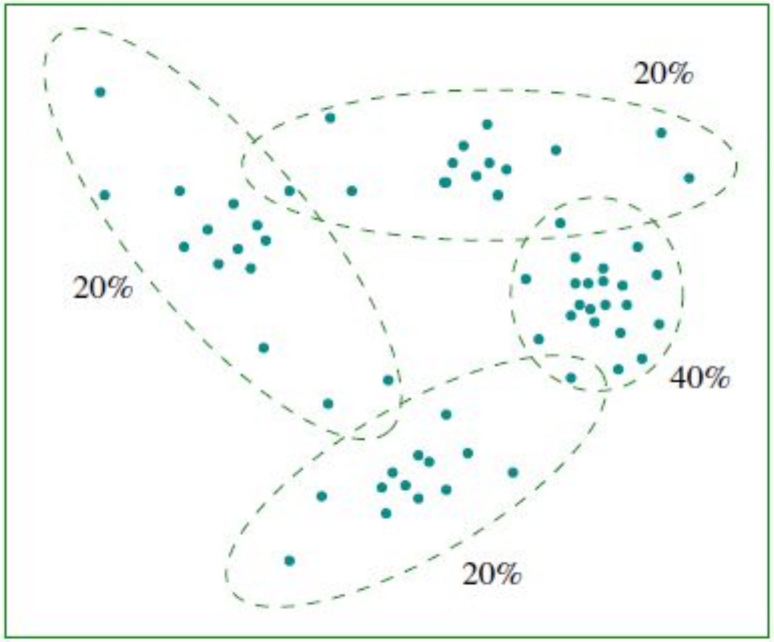
Mixtures of Gaussians

Idea: model each cluster by a Gaussian:



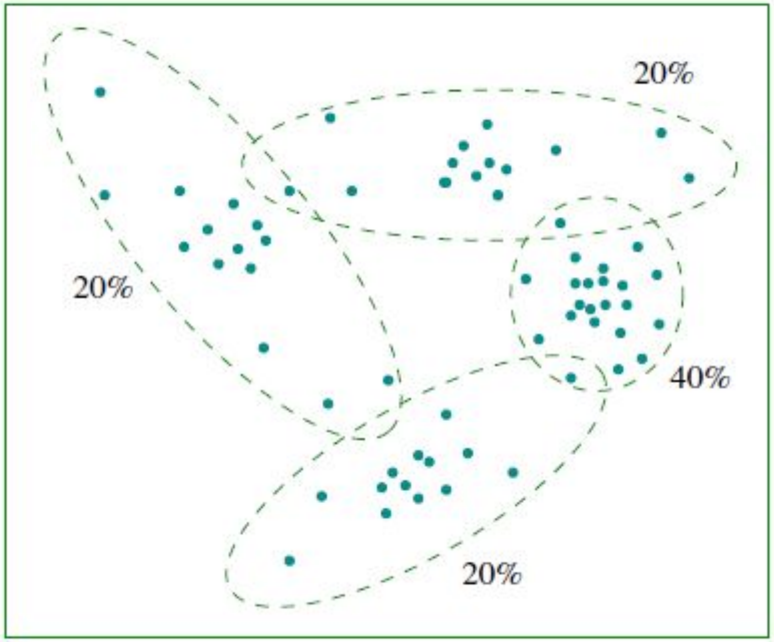
Mixtures of Gaussians

Idea: model each cluster by a Gaussian:



Mixtures of Gaussians

Idea: model each cluster by a Gaussian:

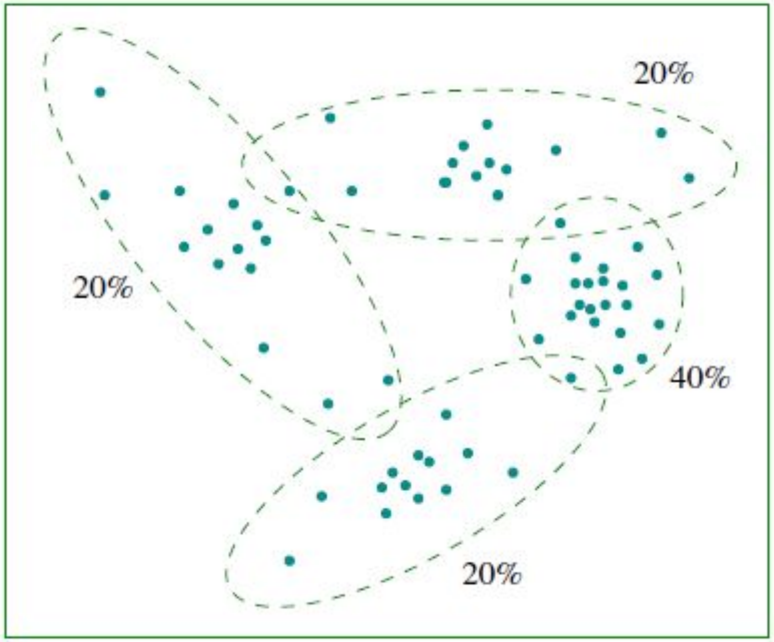


Each of the k clusters is specified by:

- a Gaussian distribution $P_j = N(\mu_j, \Sigma_j)$
- a mixing weight π_j

Mixtures of Gaussians

Idea: model each cluster by a Gaussian:



Each of the k clusters is specified by:

- a Gaussian distribution $P_j = N(\mu_j, \Sigma_j)$
- a mixing weight π_j

Overall distribution over $\mathbb{R}^p$: a **mixture of Gaussians**

$$\Pr(x) = \pi_1 P_1(x) + \cdots + \pi_k P_k(x)$$

The clustering task

Given data $x_1, \ldots, x_n \in \mathbb{R}^P$, find the maximum-likelihood mixture of Gaussians: that is, find parameters

- $\pi_1, \ldots, \pi_k \geq 0$ summing to one
- $\mu_1, \ldots, \mu_k \in \mathbb{R}^p$
- $\Sigma_1, \ldots, \Sigma_k \in \mathbb{R}^{p \times p}$

to maximize

$$\Pr(\text{data} \mid \pi_1 P_1 + \cdots + \pi_k P_k) = \prod_{i=1}^{n} (P(x_i | \pi_1 P_1 + .. + \pi_k P_k))$$

$$= \prod_{i=1}^{n} (\sum_{j=1}^{n} P(x_i, z_j | \pi_1 P_1 + .. + \pi_k P_k)) = \prod_{i=1}^{n} (\sum_{j=1}^{n} \pi_j P(x_i | z_j))$$

$$= \prod_{i=1}^{n} \left(\sum_{j=1}^{k} \pi_j P_j(x_i) \right)$$

$$= \prod_{i=1}^{n} \left(\sum_{j=1}^{k} \frac{\pi_j}{(2\pi)^{p/2} |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right) \right)$$

where P_j is the distribution of the jth cluster, $N(\mu_j, \Sigma_j)$.

The EM algorithm

1. Initialize $\pi_1, \ldots, \pi_k$ and $P_1 = N(\mu_1, \Sigma_1), \ldots, P_k = N(\mu_k, \Sigma_k)$ in some manner.
2. Repeat until convergence:
   - Assign each point x_i fractionally between the k clusters:

$$w_{ij} = \Pr(\text{cluster } j \mid x_i) = \frac{\pi_j P_j(x_i)}{\sum_\ell \pi_\ell P_\ell(x_i)}$$

   - Now update the mixing weights, means, and covariances:

$$\pi_j = \frac{1}{n} \sum_{i=1}^{n} w_{ij}$$

$$\mu_j = \frac{1}{n\pi_j} \sum_{i=1}^{n} w_{ij} x_i$$

$$\Sigma_j = \frac{1}{n\pi_j} \sum_{i=1}^{n} w_{ij} (x_i - \mu_j)(x_i - \mu_j)^T$$