# The EM algorithm for Gaussian mixture models

Sanjoy Dasgupta

University of California, San Diego

## Topics we'll cover

1. Gaussian mixture models
2. The optimization problem
3. The EM algorithm
4. Examples

## $K$-means: the good and the bad

The good:

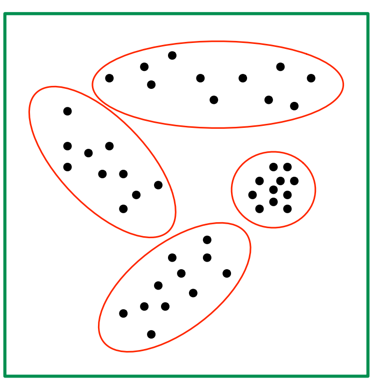
- Fast and easy.
- Effective in quantization.

The bad:

- Geared towards data in which the clusters are spherical, and of roughly the same radius.

Is there is a similarly-simple algorithm in which clusters of more general shape are accommodated?

## Mixtures of Gaussians



Each of the $k$ clusters is specified by:

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

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

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

## The clustering task

We are given data $x_1, \ldots, x_n \in \mathbb{R}^d$.

For any mixture model $\pi_1, \ldots, \pi_k,\ P_1 = N(\mu_1, \Sigma_1), \ldots, P_k = N(\mu_k, \Sigma_k)$,

$$\begin{aligned}
&\Pr\left(\text{data} \mid \pi_1 P_1 + \cdots + \pi_k P_k\right) \\
&= \prod_{i=1}^{n} \left(\pi_1 P_1(x_i) + \cdots + \pi_k P_k(x_i)\right) \\
&= \prod_{i=1}^{n} \left( \sum_{j=1}^{k} \frac{\pi_j}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \exp\left( -\frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right) \right)
\end{aligned}$$

Find the **maximum-likelihood mixture of Gaussians**:
the parameters $\{\pi_j, \mu_j, \Sigma_j : j = 1 \ldots k\}$ that maximize this function.

## Optimization surface

Minimize the negative log-likelihood,

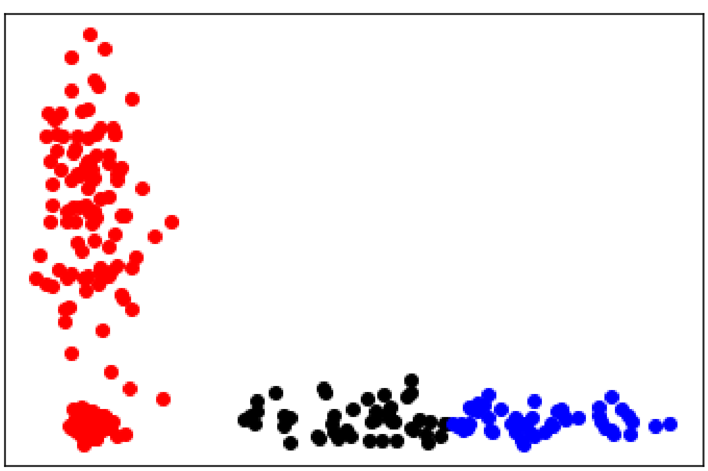
$$L(\{\pi_j, \mu_j, \Sigma_j\}) \ = \ \sum_{i=1}^{n} \ln \left( \sum_{j=1}^{k} \frac{\pi_j}{(2\pi)^{d/2} |\Sigma_j|^{1/2}} \exp\left( -\frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right) \right)$$

## The EM algorithm

1. Initialize $\pi_1, \ldots, \pi_k$ and $P_1 = N(\mu_1, \Sigma_1), \ldots, P_k = N(\mu_k, \Sigma_k)$.
2. Repeat until convergence:
   - Assign each point $x_i$ fractionally between the $k$ clusters:
     $$w_{ij} = \Pr(\text{cluster } j \mid x_i) = \frac{\pi_j P_j(x_i)}{\sum_\ell \pi_\ell P_\ell(x_i)}$$
   - Update mixing weights, means, and covariances:
     $$\pi_j = \frac{1}{n}\sum_{i=1}^n w_{ij}$$
     $$\mu_j = \frac{1}{n\pi_j}\sum_{i=1}^n w_{ij} x_i$$
     $$\Sigma_j = \frac{1}{n\pi_j}\sum_{i=1}^n w_{ij}(x_i - \mu_j)(x_i - \mu_j)^T$$

## Example

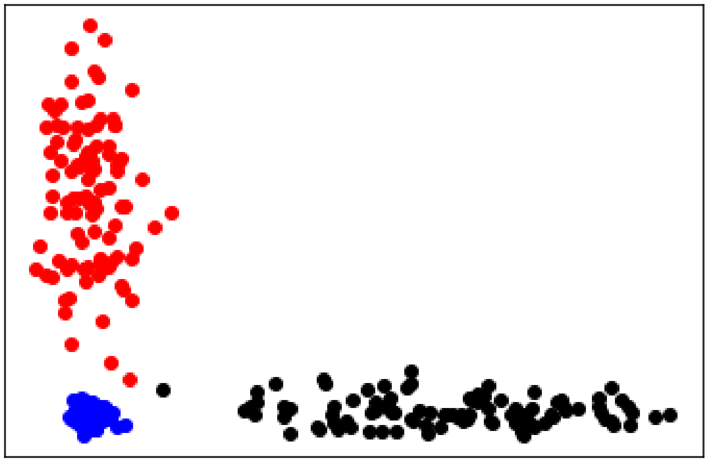
Data with 3 clusters, each with 100 points.



$k$-means solution 1

# Example

Data with 3 clusters, each with 100 points.



EM for mixture of Gaussians