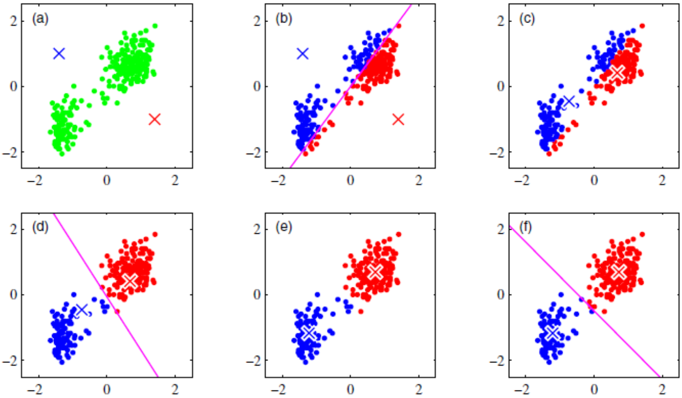
COMS30035, Machine learning: k-means and mixtures of Gaussians

James Cussens

School of Computer Science
University of Bristol

3rd September 2024

k-means for clustering



k-means optimisation

- $r_{nk} = 1$ if datapoint $\mathbf{x}_n$ is assigned to cluster k.
- $\boldsymbol{\mu}_k$ is the mean of cluster k.

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j ||\mathbf{x}_n - \boldsymbol{\mu}_j||^2 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$\boldsymbol{\mu}_k = \frac{\sum_n r_{nk}\mathbf{x}_n}{\sum_n r_{nk}} \tag{2}$$

Gaussian mixture distribution

Well, here it is [Bis06, §9.2]

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{3}$$

We can associate the *mixing coefficients* π_k with a K-dimensional random variable $\mathbf{x}$ where:

$$\begin{aligned} z_k &\in \{0, 1\} \\ \sum_k z_k &= 1 \\ p(z_k = 1) &= \pi_k \end{aligned}$$

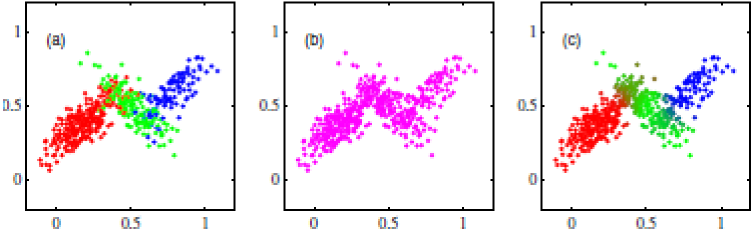
so we have

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z}) p(\mathbf{x}|\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{4}$$

Responsibility and sampling

- Now we have a full joint distribution $p(\mathbf{x}, \mathbf{z})$ we can define the *responsibility* that component k has for 'explaining' observation $\mathbf{x}$
$$\gamma(z_k) = p(z_k = 1|\mathbf{x}) \tag{5}$$
- To sample from a Gaussian mixture just use ancestral sampling: sample from $p(\mathbf{z})$, and then from $p(\mathbf{x}|\mathbf{z})$.

Soft clustering with Gaussian mixtures



More clustering with Gaussian mixtures

- If we want we can put restrictions on the covariance matrices of the Gaussians in the mixture.
- Let's have a look at [Mur22, p.729]

(Soft) clustering by MLE of a Gaussian mixture

- Given data $\mathbf{X}$ (and a fixed number K of component Gaussians) we can use MLE to get a particular Gaussian mixture distribution.
- This gives us a 'soft clustering'.
- Here's the log-likelihood [Bis06, 433]:

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{6}$$

- Number of problems:
  1. Possible singularities
  2. Symmetry/nonidentifiability
  3. No closed form for the MLE
- Queue the EM algorithm . . .

Christopher M. Bishop.
Pattern Recognition and Machine Learning.
Springer, 2006.

Kevin P. Murphy.
Probabilistic Machine Learning: An introduction.
MIT Press, 2022.