Lecture 8

Chiranjib Bhattacharyya

chibha@chalmers.se

17th Feb, 2012

Agenda

- K-means algorithm
- Mixture models

The clustering problem

Given $D = \{x_1, \ldots, x_N\}$ find Clusters $C_i, i = 1, \ldots, K$ such that any x_j belongs to only one of the clusters.

Assume K, number of clusters, is given.

K-means algorithm

Assume that each observation is in $\mathbb{R}^d$. Let $D = \{x_1, \ldots, x_N\}$ find $\mu_1, \ldots, \mu_K$ such that

$$J(r, \mu) = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$$

is minimized. Each $r_{nk} = \{0, 1\}$ and $\sum_{k=1}^{K} r_{nk} = 1$

K-means algorithm

Suppose we know $\mu_1, \ldots, \mu_K$ then

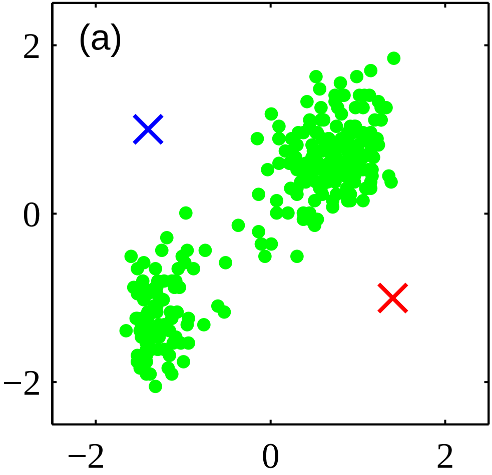
$$r_{nk} = \begin{cases} 1 & \text{if } k = \text{argmin}_j \|x_n - \mu_j\|^2 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we know r then

$$\mu_k = \frac{\sum_{n=1}^{N} r_{nk} x_n}{\sum_{n=1}^{N} r_{nk}}$$

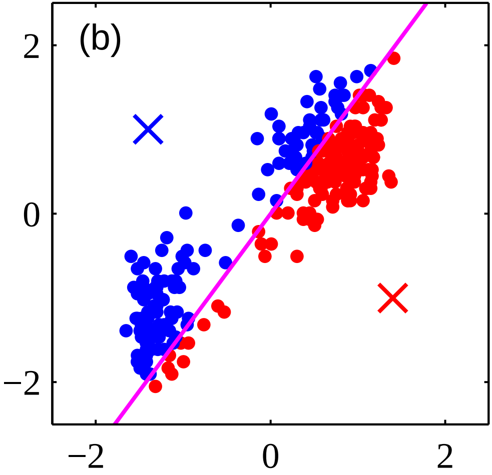
An example

Source: Pattern recognition and Machine Learning (Chris Bishop)



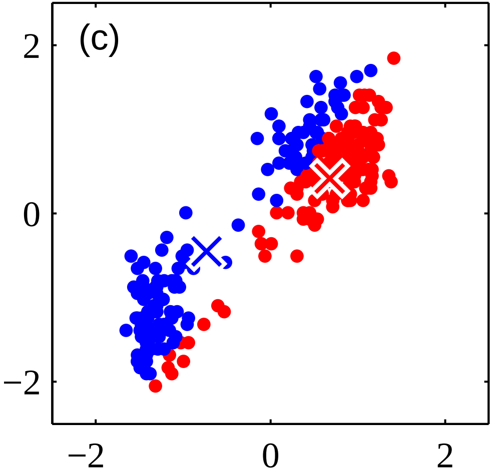
An example

Source: Pattern recognition and Machine Learning (Chris Bishop)



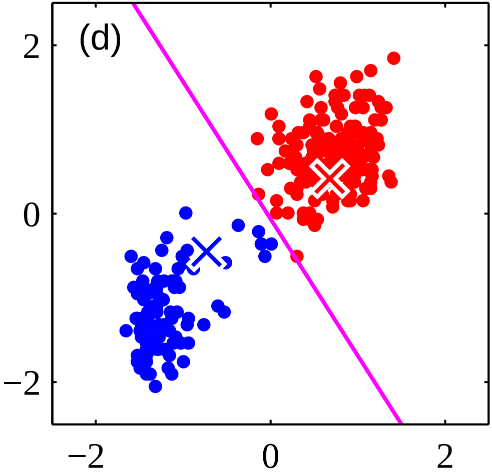
An example

Source: Pattern recognition and Machine Learning (Chris Bishop)



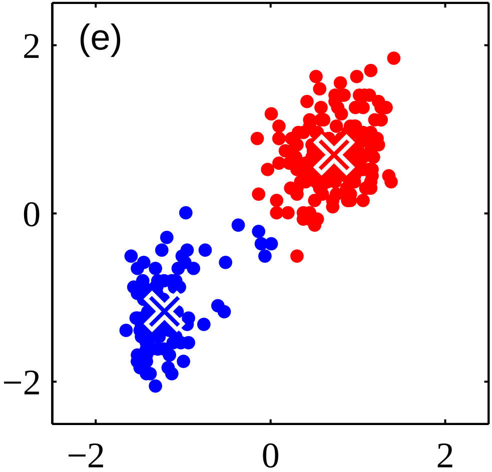
An example

Source: Pattern recognition and Machine Learning (Chris Bishop)



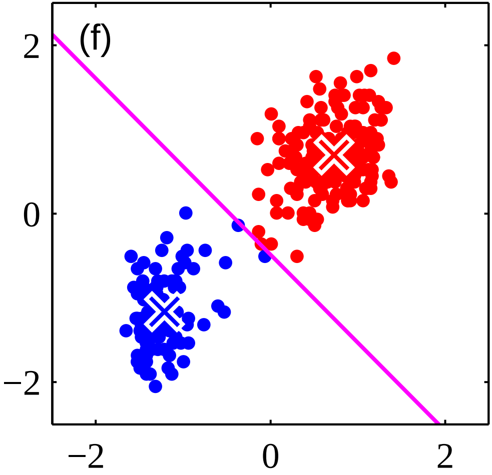
An example

Source: Pattern recognition and Machine Learning (Chris Bishop)



An example

Source: Pattern recognition and Machine Learning (Chris Bishop)



An example

Source: Pattern recognition and Machine Learning (Chris Bishop)

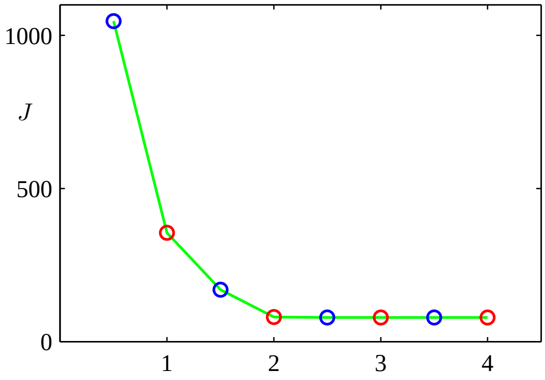


Image segmentation

Source: Pattern recognition and Machine Learning (Chris Bishop)

$K = 2$

Image segmentation

Source: Pattern recognition and Machine Learning (Chris Bishop)

$K = 3$

Image segmentation

Source: Pattern recognition and Machine Learning (Chris Bishop)

$K = 10$

Image segmentation

Source: Pattern recognition and Machine Learning (Chris Bishop)

Original image

Mixture of Gaussian distributions

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$$

Mixture of Gaussian distributions

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$$

$$P(X = x) = \sum_{k=1}^{K} P(X = x|z = k)P(z = k)$$

$$P(X = x|z = k) = N(x|\mu_k, \Sigma_k) \;\; P(z = k) = \pi_k$$

Mixture of Gaussian distribution

Maximum likelihood estimation:

MLE estimation

Compute $\pi_k, \mu_k, \Sigma_k, k = 1, \ldots, K$ to maximize

$$\sum_{i=1}^{N} logP(X = x_i | \{\pi_k, \mu_k, \Sigma_k \; k = 1, \ldots, K\})$$

Mixture of Gaussian distribution

An algorithm for finding MLE.

Initialize: Initialize π_k, μ_k, Σ_k

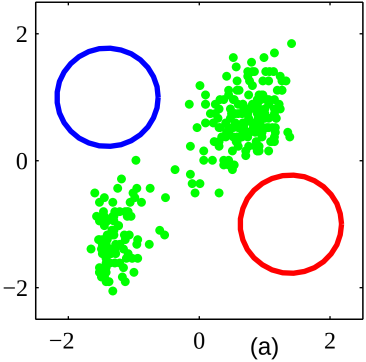
Step 1: Compute $\gamma(z_k) = P(z_k = 1|x)$

Step 2: Recompute μ_k, Σ_k, π_k

Step 3: Check for convergence

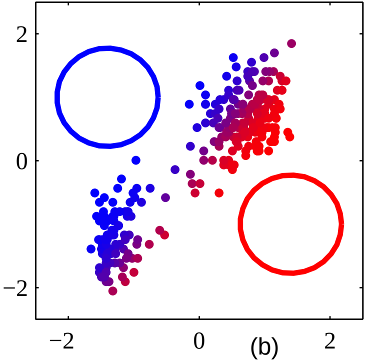
EM algorithm

Source: Pattern recognition and Machine Learning (Chris Bishop)



EM algorithm

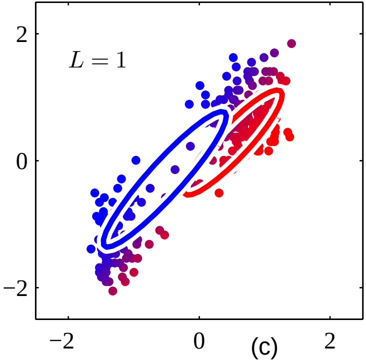
Source: Pattern recognition and Machine Learning (Chris Bishop)



(b)

EM algorithm

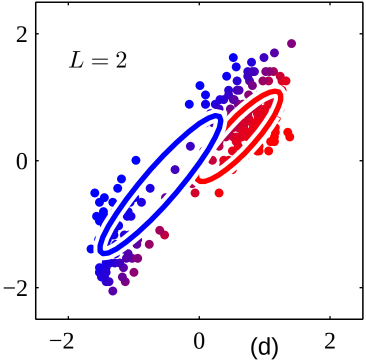
Source: Pattern recognition and Machine Learning (Chris Bishop)



(c)

EM algorithm

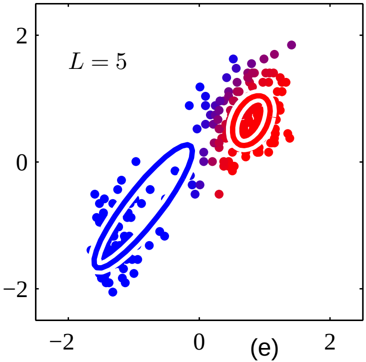
Source: Pattern recognition and Machine Learning (Chris Bishop)



(d)

EM algorithm

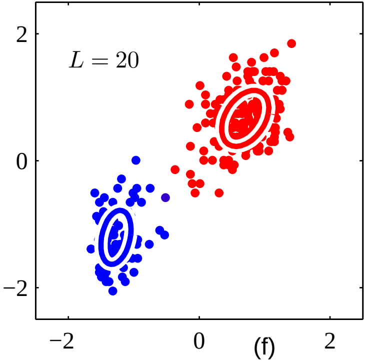
Source: Pattern recognition and Machine Learning (Chris Bishop)



(e)

EM algorithm

Source: Pattern recognition and Machine Learning (Chris Bishop)



(f)