Advanced Techniques for Mobile Robotics

Gaussian Mixture Models

Wolfram Burgard, Cyrill Stachniss,

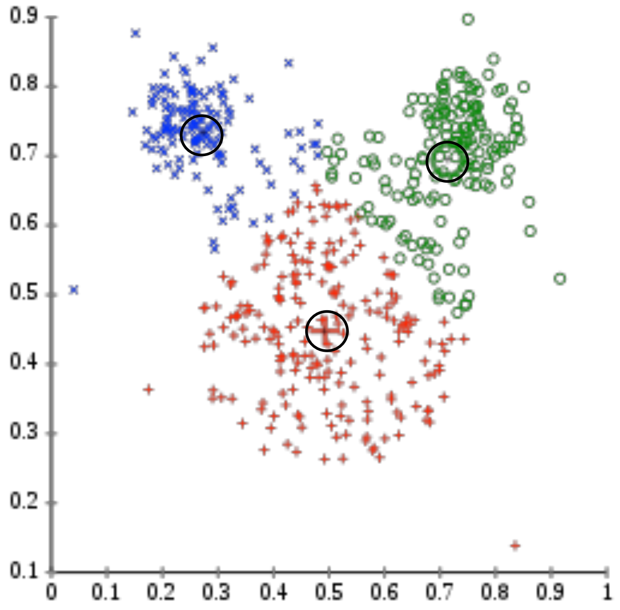
Kai Arras, Maren Bennewitz

UNI FREIBURG

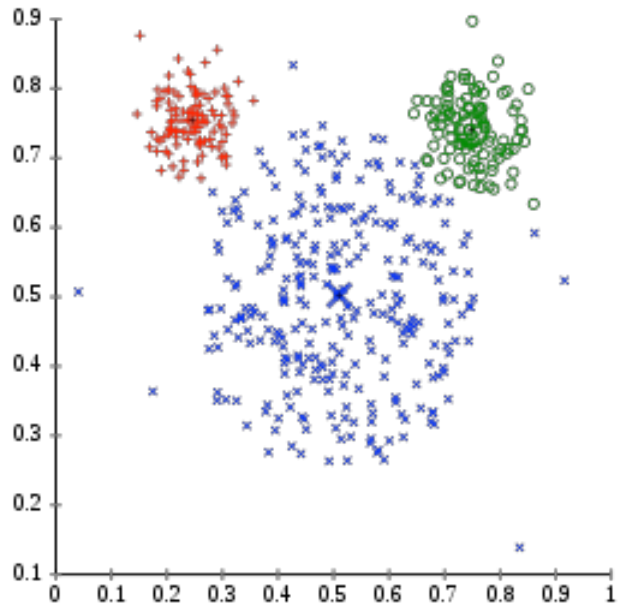
Recap K-Means

- Can be applied for clustering data
- Computes new centroids for the clusters in an iterative manner
- Converges to a local optimum
- Uses a fixed variance
- But the shapes of the clusters can be different in reality!

Motivation



k-means



Clustering based on a mixture of Gaussians

image source: wikipedia

Mixtures of Gaussians

- Assume that the data points are generated by sampling from a continuous function
- A mixture of Gaussians is such a generative model
- K Gaussians with means μ_k and covariance matrices Σ_k
- Each point is generated from one mixture component (but we don't know from which one)
- Use mixing coefficients π_k (probability that a data point is generated from component *k*)

EM for Gaussian Mixtures Models (GMMs)

- E-step: Softly assign data points to mixture components

$$c_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \mu_k, \mathbf{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n | \mu_j, \mathbf{\Sigma}_j)}$$

data point

mixture component

- Similar to k-means but considers mixing coefficients

EM for Gaussian Mixtures Models (GMMs)

- E-step: Softly assign data points to mixture components

$$c_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \mu_k, \mathbf{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n | \mu_j, \mathbf{\Sigma}_j)}$$

- M-step: Re-estimate the parameters for each mixture component based on the soft assignments

$$\mu_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} c_{nk} \mathbf{x}_n$$ (as in k-means)

$$\mathbf{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} c_{nk} (\mathbf{x}_n - \mu_k^{\text{new}})(\mathbf{x}_n - \mu_k^{\text{new}})^T$$

$$\pi_k^{\text{new}} = \frac{N_k}{N}$$

where

$$N_k = \sum_{n=1}^{N} c_{nk}$$ “soft” assignments to *k*

EM with GMMs

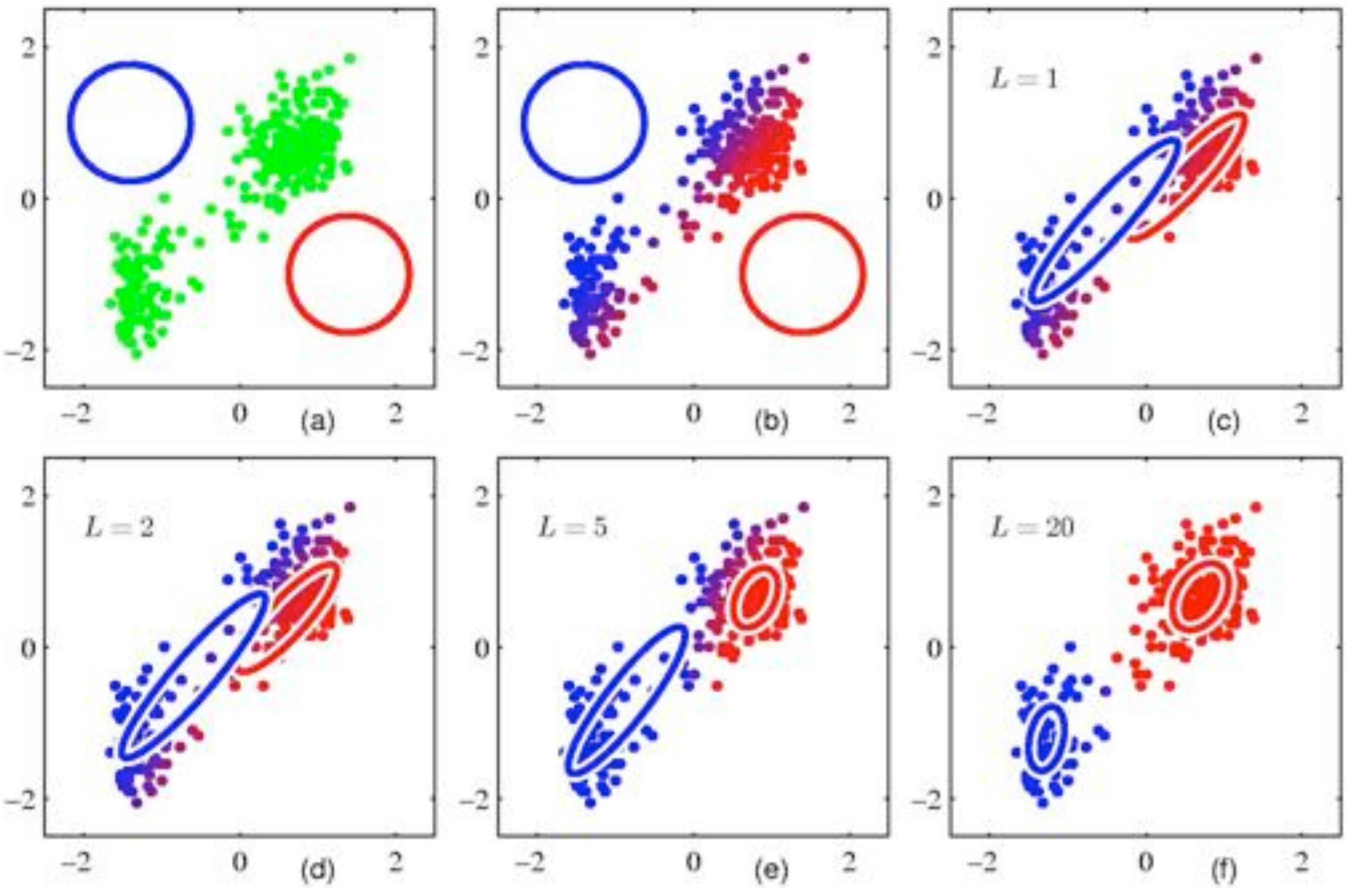


image source: C. M. Bishop

Properties of Gaussian Mixture Models

- Can represent any continuous distribution
- Number of mixture components must be estimated separately (as with k-means)
- EM for GMMs is computationally more expensive than for k-means
- EM converges slower than for k-means
- Results depend on the initialization
- K-means can be used for initialization (to speed up convergence and to find a “better” local optimum)

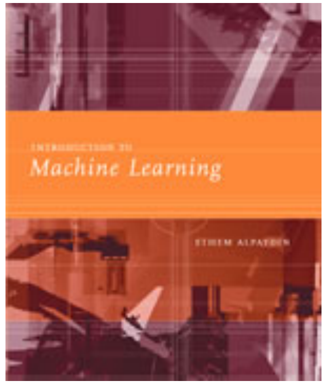
Initialization with K-Means

- Run k-means N times
- Take best result (highest likelihood)
- Use this result to initialize EM for the GMM
  - Set μ_j to the mean of cluster j from k-means
  - Set Σ_j to the covariance of the data points associated with cluster j

Further Reading

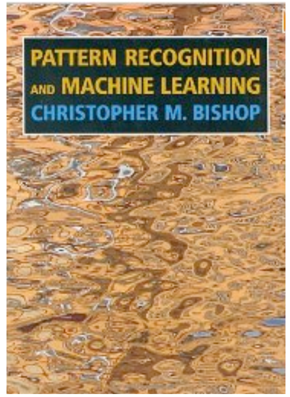
E. Alpaydin

Introduction to Machine Learning



C.M. Bishop

Pattern Recognition and Machine Learning



J. A. Bilmes

A Gentle Tutorial of the EM algorithm and its Applications to Parameter Estimation (Technical report)