Unsupervised Learning and Clustering

Chapter 10
Pattern Recognition

Prof. Vidya Manian
Dept. of Electrical and Comptuer Engineering

Introduction

- Unsupervised: use unlabeled samples
- Collecting and labeling large set of sample patterns is costly
- Train large amounts of unlabeled data, then use supervision to label the groupings found
- Useful for “data mining” applications
- Find features useful for categorization: “data dependent smart processing” or “smart feature extraction”
- Exploratory data analysis

Mixture densities and identifiability

- Functional form of the underlying probability densities are known, an unknown parameter vector is to be learnt
- Assumptions:
- The samples come from a known number of classes c
- Prior probabilities P(wj) are known
- Form of class-conditional densities are known
- The value for the c parameter vectors is unknown
- Category labels are unknown

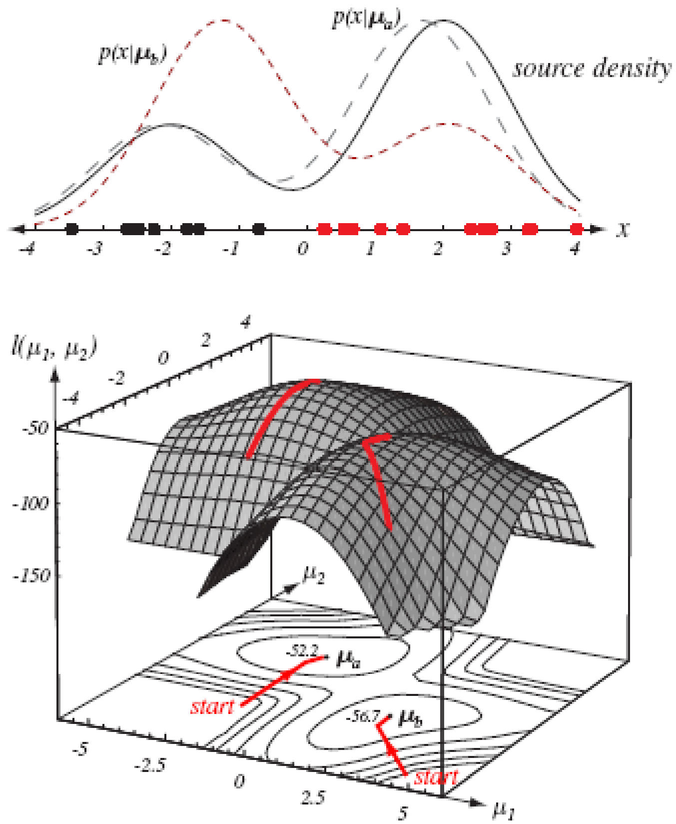

FIGURE 10.1. (Above) The source mixture density used to generate sample data, and two maximum-likelihood estimates based on the data in the table. (Bottom) Log-likelihood of a mixture model consisting of two univariate Gaussians as a function of their means, for the data in the table. Trajectories for the iterative maximum-likelihood estimation of the means of a two-Gaussian mixture model based on the data are shown as red lines. Two local optima (with log-likelihoods −52.2 and −56.7) correspond to the two density estimates shown above. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Clustering

- Classes themselves are initially undefined
- Process of grouping samples, so that the samples are similar within each group
- In Images - Group pixels ofsimilar gray level, color, local texture

Hierarchical clustering

- Organize data into large groups, which contain smaller groups (drawn as a tree)
- Agglomerative clustering algorithm
- 1. Begin with n clusters, each consisting of 1 sample
- 2. Repeat step 3 a total of n-1 times
- 3. Find the most similar clusters C_i and C_j and merge Ci and Cj into one cluster. If there is a tie, merge the first pair found

Single linkage and complete linkage algorithms

- Defining distance between 2 clusters to be the smallest distance between 2 points such that one point is in each cluster C_i and C_j
- $D_{SL}(C_i, C_j) = \min d(a,b)$ where $a \in C_i$ and $b \in C_j$
- D(a,b) – distance between samples a and b
- **Complete linkage algorithm** – maximum method or farthest neighbor method
- $D_{CL}(C_i, C_j) = \max d(a,b)$ where $a \in C_i$ and $b \in C_j$
- Largest distance between a sample in one cluster to a sample in another cluster

Ward's method

- Minimum variance method start with one cluster for each individual sample
- Merge pairs that produce the smallest squared error for the resulting set of clusters
- X_i is the feature vector, cluster has m samples $(x_1,\ldots,x_m)$

Partition Clustering

Forgy's algorithm

- 1. Initialize cluster centroids to the seed point
- 2. For each sample, find the cluster centroid nearest to it. Put the label of the nearest centroid to this sample
- 3. If no samples changed clusters in step 2 stop
- 4. Compute centroid of the resulting cluster and go to step 2

k-means clustering

Algorithm 1.

- Begin initialize n, c, $\mu_1, \mu_2, \ldots \mu_c$
  - **do** classify n samples according to nearest μ_i
    - Recompute μ_i
  - **until** no change in μ_i
  - **return** $\mu_1, \mu_2, \ldots, \mu_c$
- end

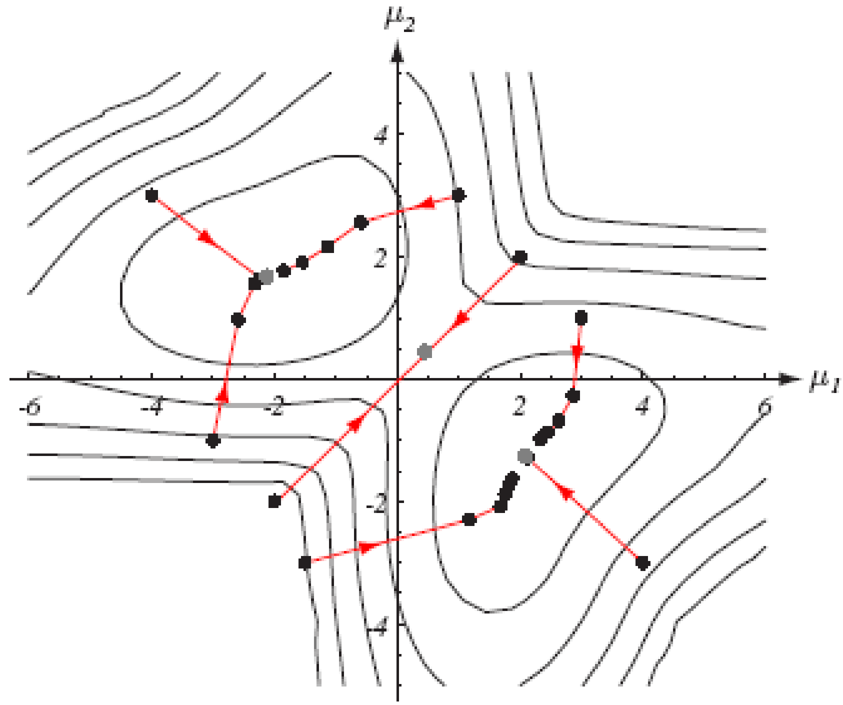

FIGURE 10.2. The k-means clustering procedure is a form of stochastic hill climbing in the log-likelihood function. The contours represent equal log-likelihood values for the one-dimensional data in Fig. 10.1. The dots indicate parameter values after different iterations of the k-means algorithm. Six of the starting points shown lead to local maxima, whereas two (i.e., $\mu_1(0) = \mu_2(0)$) lead to a saddle point near $\boldsymbol{\mu} = \mathbf{0}$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

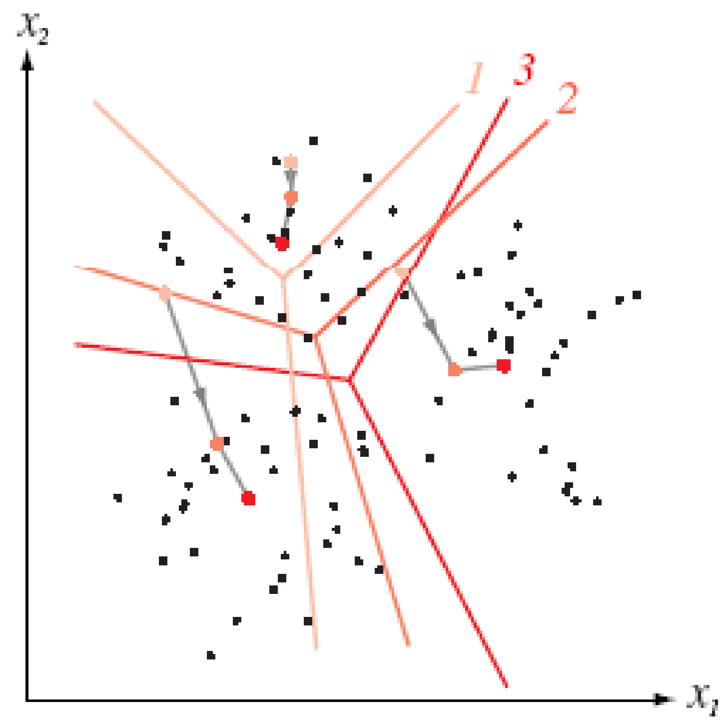

FIGURE 10.3. Trajectories for the means of the *k*-means clustering procedure applied to two-dimensional data. The final Voronoi tesselation (for classification) is also shown—the means correspond to the "centers" of the Voronoi cells. In this case, convergence is obtained in three iterations. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.