Mixture Models and EM

Yoan Miche

CIS, HUT

March 17, 2007



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 1 / 2

Mise en Bouche



Mixture Models and EM

Mise en Bouche



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 2 / 2

- 1 Starters: K-Means
- 2 Main Course: EM Algorithm
- 3 Dessert: General Algorithm



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 3 / 23

Starters: K-Means On the idea of K-Means Going Further...

Setting

Data set of N observations of D-dimensional variable \mathbf{x}

 $\{\textbf{x}_1,...,\textbf{x}_N\}$

with each vector D-dimensional

$$\mathbf{x}_1 = (x_1^1, ..., x_1^D)$$



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 4 / 2

On the idea of *K*-Means Going Further...

K-Means Idea: Clustering

Idea

• Define the K prototype vectors (code-book vectors) D-dimensional

μ_k

- All prototypes of the corresponding cluster: Sum of squares of distances from each data point to its closest prototype μ_k is minimum.
- This is the condition to define the μ_k



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 5 / 23

On the idea of *K*-Means Going Further...

K-Means Idea: Clustering

Idea

• Define the K prototype vectors (code-book vectors) D-dimensional

μ_k

- All *prototypes* of the corresponding cluster: Sum of squares of distances from each data point to its closest prototype μ_k is minimum.
- This is the condition to define the μ_k



K-Means Idea: Clustering

Idea

• Define the K prototype vectors (code-book vectors) D-dimensional

μ_k

- All prototypes of the corresponding cluster: Sum of squares of distances from each data point to its closest prototype μ_k is minimum.
- This is the condition to define the μ_k



On the idea of *K*-Means Going Further...

Clustering idea (2)

Minimize the distortion measure J

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||\mathbf{x}_n - \boldsymbol{\mu}_k||^2$$
(1)

Sum of squares of distances of all data points to their prototype.

 $r_{nk} = \delta_{n,k}$ (Kronecker)



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 6 / 23

On the idea of *K*-Means Going Further...

The Old Faithful data set





Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 7 /

On the idea of K-Means Going Further...

The Old Faithful data set (2)





Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 8 / 2

On the idea of K-Means Going Further...

The Old Faithful data set (3)





Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 9 / 2

On the idea of *K*-Means Going Further...

How to determine r_{nk} and μ_k

Iterative procedure in two steps

- E Step: Fix initial values for the {μ_k} and minimize J with respect to the r_{nk}
- M Step: Fix the r_{nk} and minimize J with respect to the $\{\mu_k\}$

In practice..

$$\mathbf{r}_{nk} = \begin{cases} 1 & \text{if } k = \min_{j} ||\mathbf{x}_{n} - \boldsymbol{\mu}_{k}||^{2} \\ 0 & \text{else} \\ & \sum_{n} r_{nk} \mathbf{x}_{n} \end{cases}$$
(2)

Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 10 / 23

On the idea of *K*-Means Going Further...

How to determine r_{nk} and μ_k

Iterative procedure in two steps

- E Step: Fix initial values for the {μ_k} and minimize J with respect to the r_{nk}
- M Step: Fix the r_{nk} and minimize J with respect to the $\{\mu_k\}$

In practice...

$$r_{nk} = \begin{cases} 1 & \text{if } k = \min_{j} ||\mathbf{x}_{n} - \boldsymbol{\mu}_{k}||^{2} \\ 0 & \text{else} \end{cases}$$
(2)
$$\mu_{k} = \frac{\sum_{n} r_{nk} \mathbf{x}_{n}}{\sum_{n} r_{nk}}$$
(3)

March 17, 2007

10 / 23

Yoan Miche (CIS, HUT)

Mixture Models and EM

On the idea of K-Means Going Further...

Generalizing to other "distances"

Generalizing to the K-medoids algorithm.

$$\widetilde{J} = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \nu(\mathbf{x}_n - \boldsymbol{\mu}_k)$$



(4)

Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 11 / 23

On the idea of *K*-Means Going Further...

Image Clustering

Pixels considered as 3-dimensional vectors



(j) K = 2 (k) K = 3 (l) K = 5 (m) K = 10



Reminders and Notations

Gaussian mixture distribution:

Recall that a Gaussian mixture can be written as a sum of Gaussians:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$
(5)

And let us introduce

whose distribution, considering $p(z_k = 1) = \pi_k$ can be written in the form

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 13 / 23

Reminders and Notations (2)

We also have

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{zk}$$
(8)

giving finally, by summing the joint distribution

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z}) p(\mathbf{x}|\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (9)$$

Only interest in these development in the introduction of the latent variable \mathbf{z}_n



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 14 / 23

 \mathbf{Z}

 \mathbf{X}

Reminders and Notations (3)

Lets us note, further, $\gamma(z_k) = p(z_k = 1 | \mathbf{x})$

$$\gamma(z_k) = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$
(10)

the conditional probability of z given x.

 $\gamma(z_k)$ is in fact the "responsibility" that k has in the observation ${f x}$



Mixture Models and EM

Example for mixture of Gaussians





Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 16 / 23

This time, instead of codebook vectors, we use Gaussians:

Now...

Maximizing the log-likelihood instead of the distortion measure
 μ_k = 1/N_k Σ^N_{n=1} γ(z_{nk})x_n with N_k = Σ^N_{n=1} γ(z_{nk})
 Σ_k = 1/N_k Σ^N_{n=1} γ(z_{nk})(x_n - μ_k)(x_n - μ_k)^T



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 17 / 2

This time, instead of codebook vectors, we use Gaussians:

Now...

Maximizing the log-likelihood instead of the distortion measure
μ_k = 1/N_k Σ^N_{n=1} γ(z_{nk})x_n with N_k = Σ^N_{n=1} γ(z_{nk})
Σ_k = 1/N_k Σ^N_{n=1} γ(z_{nk})(x_n - μ_k)(x_n - μ_k)^T



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 17 / 23

This time, instead of codebook vectors, we use Gaussians:

Now...

• Maximizing the log-likelihood instead of the distortion measure

•
$$\boldsymbol{\mu}_{k} = \frac{1}{N_{k}} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_{n}$$
 with $N_{k} = \sum_{n=1}^{N} \gamma(z_{nk})$
• $\boldsymbol{\Sigma}_{k} = \frac{1}{N_{k}} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_{n} - \boldsymbol{\mu}_{k}) (\mathbf{x}_{n} - \boldsymbol{\mu}_{k})^{T}$



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 17 / 23

This time, instead of codebook vectors, we use Gaussians:

Now...

• Maximizing the log-likelihood instead of the distortion measure

•
$$\boldsymbol{\mu}_{k} = \frac{1}{N_{k}} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_{n}$$
 with $N_{k} = \sum_{n=1}^{N} \gamma(z_{nk})$
• $\boldsymbol{\Sigma}_{k} = \frac{1}{N_{k}} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_{n} - \boldsymbol{\mu}_{k}) (\mathbf{x}_{n} - \boldsymbol{\mu}_{k})^{T}$



This time, instead of codebook vectors, we use Gaussians:

Now...

• Maximizing the log-likelihood instead of the distortion measure

•
$$\boldsymbol{\mu}_{k} = \frac{1}{N_{k}} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_{n}$$
 with $N_{k} = \sum_{n=1}^{N} \gamma(z_{nk})$
• $\boldsymbol{\Sigma}_{k} = \frac{1}{N_{k}} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_{n} - \boldsymbol{\mu}_{k}) (\mathbf{x}_{n} - \boldsymbol{\mu}_{k})^{T}$



In images...





Main Course: EM Algorithm Mixtures of Gaussians

EM for Gaussian Mixtures (1)

EM Algorithm (1)

- Initialize μ_k , $\mathbf{\Sigma}_k$ and π_k and evaluate log-likelihood with these
- E Step: $\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 19 / 23

EM for Gaussian Mixtures (1)

EM Algorithm (1)

• Initialize μ_k , Σ_k and π_k and evaluate log-likelihood with these

• E Step:
$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 19 / 23

EM for Gaussian Mixtures (2)

EM Algorithm (2)

• M Step:
$$\mu_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n,$$

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



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 20 / 23

Main Course: EM Algorithm Mixtures of Gaussians

EM for Gaussian Mixtures (3)

EM Algorithm (3)

• Evaluate log-likehood

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



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 21 / 23

Dessert: General Algorithm

Kidding...

The proposed version of the general algorithm in the book is way too heavy to be presented on slides...

Meanwhile, I invite you to read section 9.4 about this general algorithm since it gives a general framework that is given in Chapter 10

Or not. . .



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 22 / 23

Had fun ?

Hope you enjoyed...



Yoan Miche (CIS, HUT)

Mixture Models and EM

March 17, 2007 23 / 23