



MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Course](#) / [Unit 4. Unsupervised Learning \(2 w...](#) / [Lecture 16. Mixture Models; E](#)[← Previous](#)

5. Mixture Model - Unobserved Case: EM Algorithm

[Bookmark this page](#)

Exercises due Apr 19, 2023 08:59 -03 Completed

The EM Algorithm



Video

[!\[\]\(de95854c7ee024cfadc48187bbb781b2_img.jpg\) Download video file](#)

Transcripts

[!\[\]\(6059a5aa8b4ca7bb793408023d6c6e42_img.jpg\) Download SubRip \(.srt\) file](#)[!\[\]\(c50c8b7b2cc2cf9ff925edec0ee94c0d_img.jpg\) Download Text \(.txt\) file](#)

Estimates of Parameters of GMM: The Expectation Maximization (EM) Algorithm

We observe n data points $\mathbf{x}_1, \dots, \mathbf{x}_n$ in \mathbb{R}^d . We wish to maximize the GMM likelihood with parameter set $\theta = \{p_1, \dots, p_K, \mu^{(1)}, \dots, \mu^{(K)}, \sigma_1^2, \dots, \sigma_K^2\}$.

Maximizing the log-likelihood $\log(\prod_{i=1}^n p(\mathbf{x}^{(i)} | \theta))$ is not tractable in the setting of GMMs. A common solution to finding the parameter set θ that maximizes the likelihood. The **EM algorithm** is an iterative algorithm that finds a locally optimal solution $\hat{\theta}$ to the GMM likelihood maximization problem.

E Step

The **E Step** of the algorithm involves finding the posterior probability that point $\mathbf{x}^{(i)}$ was generated by component j for every $i = 1, \dots, n$ and $j = 1, \dots, K$. This step assumes the knowledge of the parameter

$$\ell(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \mid \theta) = \sum_{i=1}^n \log \left[\sum_{j=1}^K p(\mathbf{x}^{(i)} \text{ generated by cluster } j \mid \theta) \right].$$

Maximizing the proxy function over the parameter set θ , one can verify by taking derivative equal to zero that

$$\hat{\mu}^{(j)} = \frac{\sum_{i=1}^n p(j \mid i) \mathbf{x}^{(i)}}{\sum_{i=1}^n p(j \mid i)}$$

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^n p(j \mid i),$$

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^n p(j \mid i) \|\mathbf{x}^{(i)} - \hat{\mu}^{(j)}\|^2}{d \sum_{i=1}^n p(j \mid i)}.$$

The E and M steps are repeated iteratively until there is no noticeable change in the accuracy after M step using the newly estimated parameters or if the parameters do not vary by

Initialization

As for the initialization before the first time E step is carried out, we can either do a random parameter set θ or we can employ k-means to find the initial cluster centers of the K clusters. The global variance of the dataset as the initial variance of all the K clusters. In the latter case, $\hat{\mu}^{(j)}$ can be initialized to the proportion of data points in the clusters as found by the k-means.

Gaussian Mixture Model: An Example Update - E-Step

5/5 points (graded)

Assume that the initial means and variances of two clusters in a GMM are as follows: $\mu^{(1)} = 0.2$, $\sigma_1^2 = \sigma_2^2 = 4$. Let $p_1 = p_2 = 0.5$.

Let $x^{(1)} = 0.2$, $x^{(2)} = -0.9$, $x^{(3)} = -1$, $x^{(4)} = 1.2$, $x^{(5)} = 1.8$ be five points that we wish to

In this problem and in the next, we compute the updated parameters corresponding to any computational tool at your disposal.

Compute the following posterior probabilities (provide at least five decimal digits):

$$p(1 \mid 1) =$$

$$p(1 \mid 5) =$$



You have used 3 of 3 attempts

Gaussian Mixture Model: An Example Update - M-Step

2/3 points (graded)

Compute the updated parameters corresponding to cluster 1 (provide at least five decim

$$\hat{p}_1 =$$



$$\hat{\mu}_1 =$$



$$\hat{\sigma}_1^2 =$$



You have used 3 of 3 attempts

Gaussian Mixture Model and the EM Algorithm

1/1 point (graded)

Which of the following statements are true? Assume that we have a Gaussian mixture model with estimated parameters (means and variances of the Gaussians and the mixture weights).



A Gaussian mixture model can provide information about how likely it is that a given data point belongs to each cluster.



The EM algorithm converges to the same estimate of the parameters irrespective of the initial values.

to the dataset.

Identify the following parameters (according to notation developed in the lecture, assume data for training):

$K =$



$n =$



$d =$



Submit

You have used 1 of 2 attempts

Note: The Gaussian mixture model can be extended to the case where each mixture component has its own covariance matrix Σ_j . The case that we have studied so far is a special case where $\Sigma_j = \sigma^2 I$, the identity matrix of size $d \times d$. The EM algorithm can also be extended to work in this general case.

Discussion

< Previous

Next >

Topic: Unit 4. Unsupervised Learning (2 weeks) / Lecture 16. Mixture Models; EM algorithm / 5. Mixture Model - Unobserved Case: EM Algorithm



edX

[About](#)

[Affiliates](#)

[edX for Business](#)

[Open edX](#)

[Careers](#)

[News](#)



Useful resource WITH example calculation

https://www.cs.toronto.edu/~rgrosse/csc321/mixture_models.pdf Page 5ff. for posterior calculation... etc.



Should $d = 5$ in the E step?

Should $d = 5$ in the E step?

Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Security](#)

[Media Kit](#)



© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)