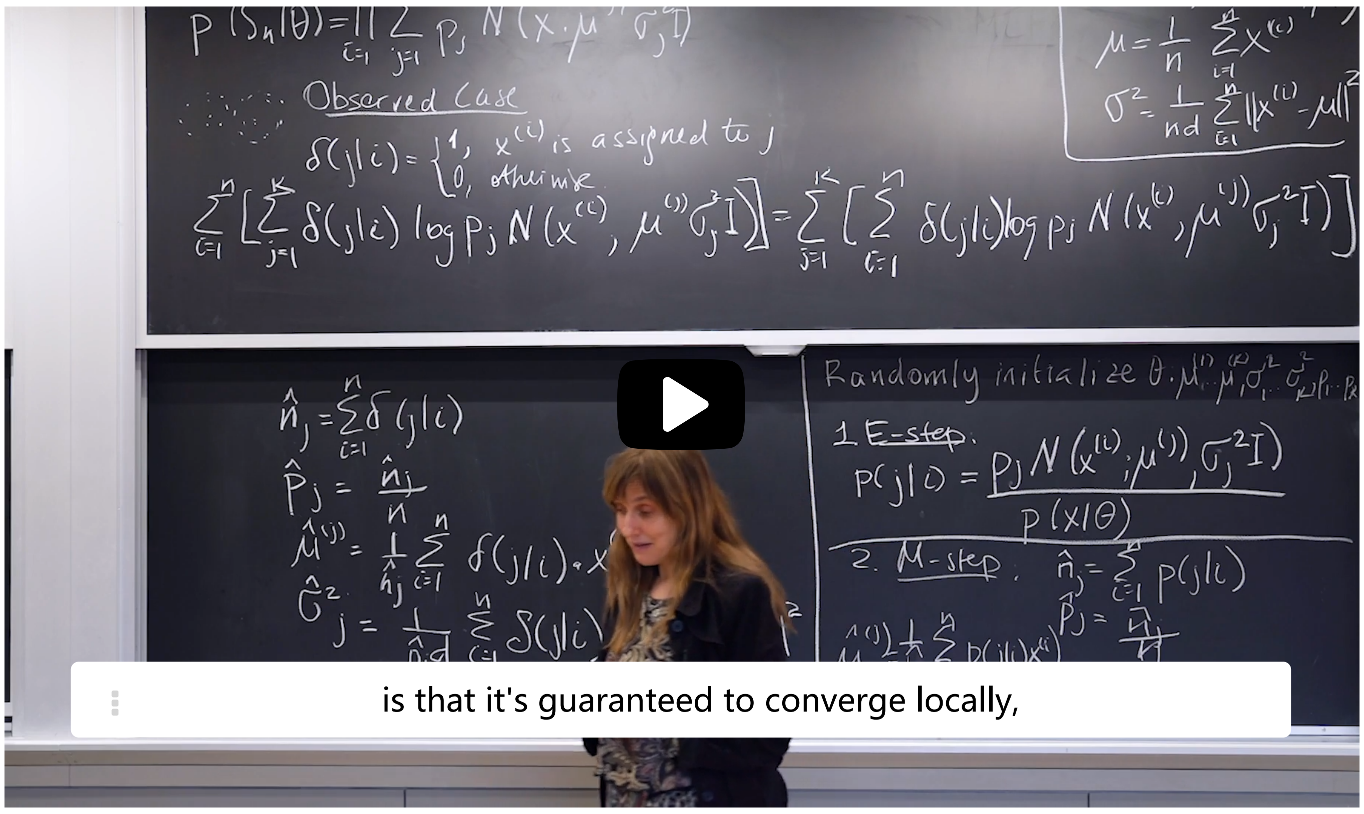
5. Mixture Model - Unobserved Case: EM Algorithm

The EM Algorithm



14:07 / 15:32 1.0x

And typically what people do in real problems--

and EM is really, really widely used--

sometimes people may look at a more simplified version

of the problem to utilize it as initialization

to more complex one.

And there are lots of interesting uses of EM,

and I hope that you will explore some of them in your exercises.

Thank you.

End of transcript. Skip to the start.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

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: $\mu^{(1)} = -3$, $\mu^{(2)} = 2$, $\sigma_1^2 = \sigma_2^2 = 4$. 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 cluster.

In this problem and in the next, we compute the updated parameters corresponding to cluster 1. You may use any computational tool at your disposal.

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

$p(1 \mid 1) =$

0.29421497 ✔ **Answer:** 0.29421

$p(1 \mid 2) =$

0.62245933 ✔ **Answer:** 0.62246

$p(1 \mid 3) =$

0.65135486 ✔ **Answer:** 0.65135

$p(1 \mid 4) =$

0.10669059 ✔ **Answer:** 0.10669

$p(1 \mid 5) =$

0.05340333 ✔ **Answer:** 0.053403

Solution:

Using the formula of the E-step

$$p(j \mid i) = \frac{p_j \mathcal{N}(x^{(i)}; \mu^{(j)}, \sigma_j^2)}{p(x^{(i)} | \theta)},$$

we can obtain that

$$\begin{aligned} p(1 \mid 1) &= 0.29421 \\ p(1 \mid 2) &= 0.62246 \\ p(1 \mid 3) &= 0.65135 \\ p(1 \mid 4) &= 0.10669 \\ p(1 \mid 5) &= 0.053403. \end{aligned}$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Gaussian Mixture Model: An Example Update - M-Step

3/3 points (graded)

Compute the updated parameters corresponding to cluster 1 (provide at least five decimal digits):

$\hat{p}_1 =$

0.345624618355527464 ✔ **Answer:** 0.34562

$\hat{\mu}_1 =$

-0.5373289474340418 ✔ **Answer:** -0.53733

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

0.5757859076870628 ✔ **Answer:** 0.57579

Solution:

Using the formulae corresponding to the M-step,

$$\begin{aligned} \hat{n}_1 &= \sum_{i=1}^{5} p(1|i) = 1.7281 \\ \hat{p}_1 &= \frac{\hat{n}_1}{n} = \frac{\hat{n}_1}{5} = 0.34562 \\ \hat{\mu}_1 &= \frac{1}{\hat{n}_1} \sum_{i=1}^{5} p(1|i)\, x^{(i)} = -0.53733 \\ \hat{\sigma}_1^2 &= \frac{1}{\hat{n}_1} \sum_{i=1}^{5} p(1|i)\, (x^{(i)} - \hat{\mu}^{(1)})^2 = 0.57579. \end{aligned}$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

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 known (or estimated) parameters (means and variances of the Gaussians and the mixture weights).

- [x] A Gaussian mixture model can provide information about how likely it is that a given point belongs to each cluster. ✔
- [ ] The EM algorithm converges to the same estimate of the parameters irrespective of the initialized values.
- [x] An iteration of the EM algorithm is computationally more expensive when compared to an iteration of the K-means algorithm for the same number of clusters. ✔

✔

Solution:

The first and third statements are true. The first statement is true because the estimated posterior probabilities tell us how likely it is that a given point belongs to each cluster. The third statement is true because each iteration of the EM algorithm involves two steps that are each more computationally expensive than the updates involved in an iteration of the K-means algorithm.

The second statement is not true. As explained in the video, the EM algorithm is guaranteed (under some conditions) to only converge locally.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Mixture Models and Digit Classification

3/3 points (graded)
Assume that we have 100,000 black-and-white images of size 26×26 pixels that are the result of scans of hand-written digits between 0 and 9.

We can apply mixture models to effectively train a classifier based on clustering using the EM algorithm applied to the dataset.

Identify the following parameters (according to notation developed in the lecture, assuming that we use all the data for training):

$K =$

10 ✔ **Answer:** 10

$n =$

100000 ✔ **Answer:** 100000

$d =$

676 ✔ **Answer:** 676

Solution:

We are classifying $n = 100,000$ vectors each of length $d = 676$ into $K = 10$ clusters (one cluster for each digit).

Submit You have used 1 of 2 attempts