

<u>Unit 4 Unsupervised Learning (2</u>

Course > weeks)

Lecture 16. Mixture Models; EM

5. Mixture Model - Unobserved

> algorithm

> Case: EM Algorithm

5. Mixture Model - Unobserved Case: EM Algorithm The EM Algorithm

counts.

For each point, I can say how likely that this point actually

belongs to this cluster.

This will be a soft equivalent of this measure.

Here, I just say, is point i assigned to cluster j?

Now we say, how likely is it that the point

i comes from the cluster j?

So it's going to be 0 and 1.

It could be some other number.

But what it actually means is that instead of using delta,

I can use here the probability.





The point is assigned.

So as previously, when I did all this computation,

instead of taking this indicator function 0 and 1s.

Video

Download video file

Transcripts

<u>Download SubRip (.srt) file</u> <u>Download Text (.txt) file</u>

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

$$p(1 | 2) =$$

0.62245933

✓ Answer: 0.62246

$$p(1 | 3) =$$

0.65135486

✓ Answer: 0.65135

$$p(1 | 4) =$$

0.10669059

✓ Answer: 0.10669

$$p(1 | 5) =$$

0.05340333

✓ Answer: 0.053403

Solution:

Using the formula of the E-step

$$p\left(j\mid i
ight) = rac{p_{j}\,\mathcal{N}\left(x^{(i)};\mu^{(j)},\sigma_{j}^{2}
ight)}{p\left(x^{(i)}\mid heta
ight)},$$

we can obtain that

$$egin{array}{lll} p\left(1\mid 1
ight) &= 0.29421 \ p\left(1\mid 2
ight) &= 0.62246 \ p\left(1\mid 3
ight) &= 0.65135 \ p\left(1\mid 4
ight) &= 0.10669 \ \end{array}$$

$$p(1 \mid 5) = 0.053403.$$

Submit

You have used 1 of 3 attempts

1 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):

$$\hat{p}_1 =$$

0.34562462

✓ Answer: 0.34562

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

-0.53732895

✓ Answer: -0.53733

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

0.57578591

✓ Answer: 0.57579

Solution:

Using the formulae corresponding to the M-step,

$$\hat{n}_1 \ = \sum_{i=1}^5 p\left(1|i
ight) = 1.7281$$

$$\hat{p}_1 = rac{\hat{n}_1}{n} = rac{\hat{n}_1}{5} = 0.34562$$

$$\hat{\mu}_{1} \; = rac{1}{\hat{n}_{1}} \sum_{i=1}^{5} p\left(1|i
ight) x^{(i)} = -0.53733$$

$$\hat{\sigma}_{1}^{2} \; = rac{1}{\hat{n}_{1}} \sum_{i=1}^{5} p\left(1|i
ight) \left(x^{(i)} - \hat{\mu}^{(1)}
ight)^{2} = 0.57579.$$

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).

- 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.
- ☑ 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

1 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 =

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

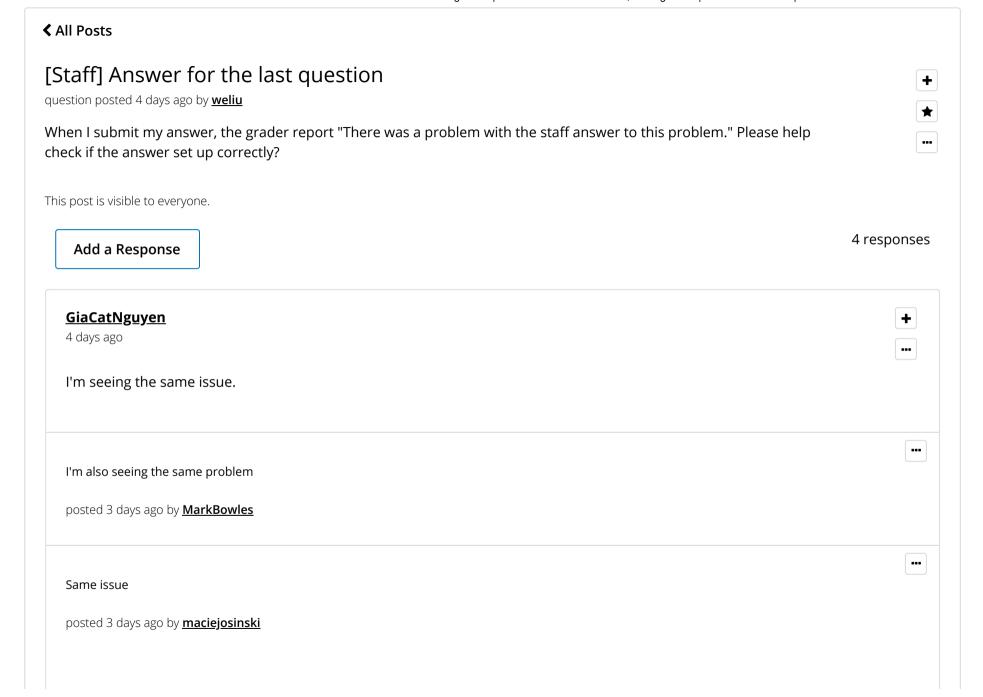
• Answers are displayed within the problem

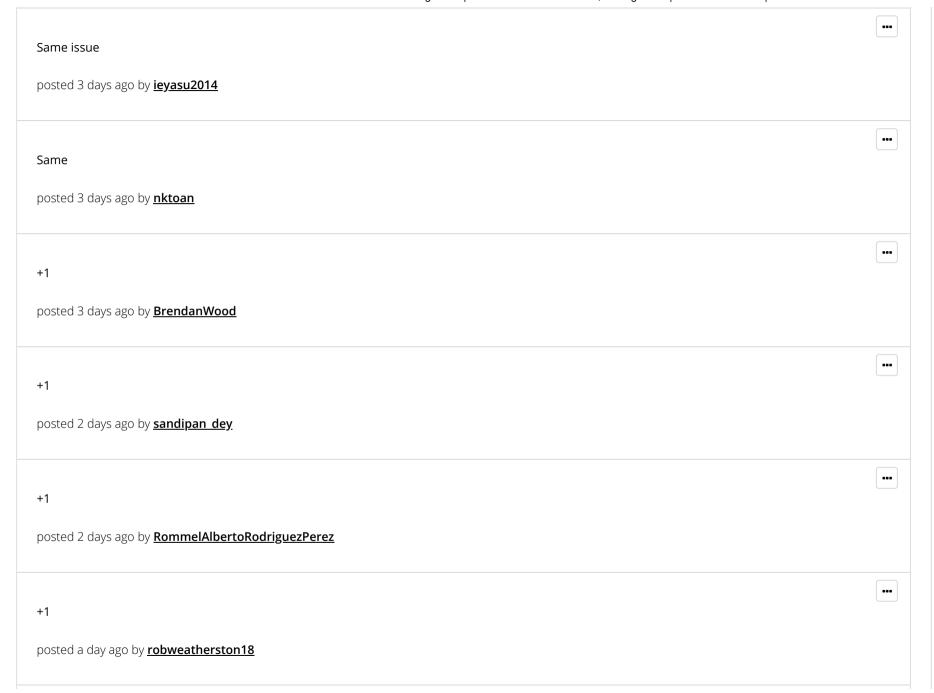
Discussion

Hide Discussion

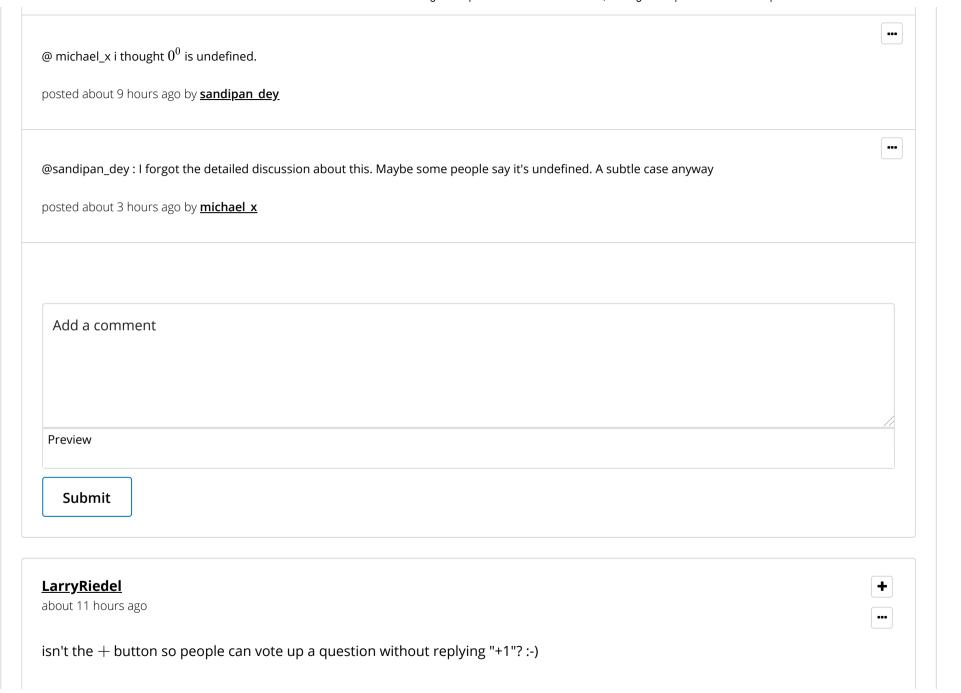
Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 16. Mixture Models; EM algorithm / 5. Mixture Model - Unobserved Case: EM Algorithm

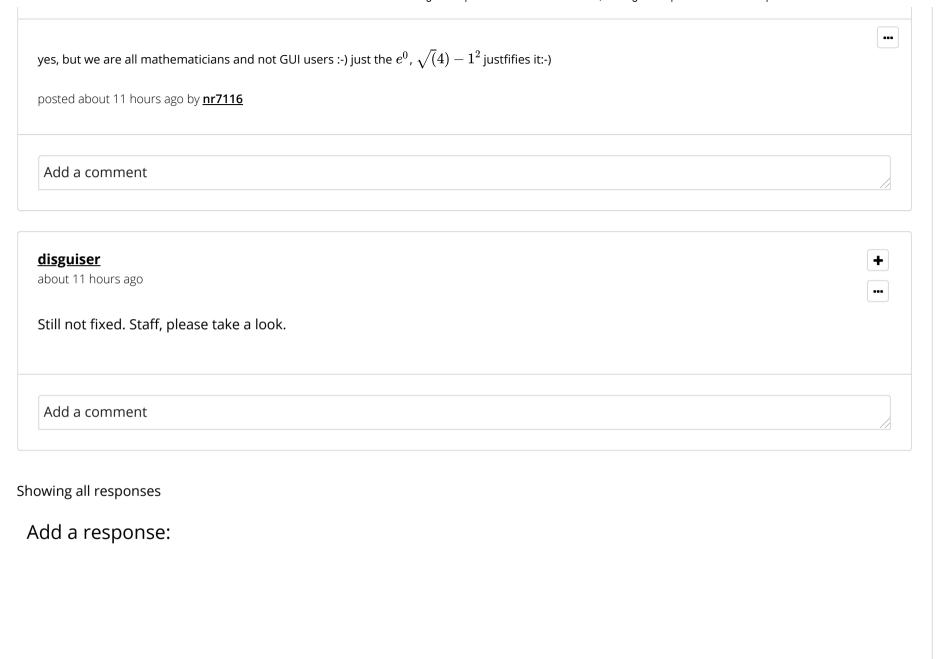
Add a Post





+1	
posted about 13 hours ago by <u>nbourbon</u>	
Add a comment	
HumphreyP 2 days ago	•••
$+\sqrt{4}-1^2$	•••
posted 2 days ago by <u>SniperWolf787</u>	
e^0 , lol posted a day ago by <u>Cool7</u> (Community TA)	•••
$+ 0^0$	•••
posted about 23 hours ago by <u>michael x</u>	





8/5/2019	5. Mixture Model - Unobserved Case: EM Algorithm Lecture 16. Mixture Models; EM algorithm 6.86x Courseware edX	
Preview		
Cubmit		
Submit		

© All Rights Reserved