EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the Privacy Policy.





Course > Final exam (1 week) > Final Exam > Problem 4

Problem 4

4. (1)

3/3 points (graded)

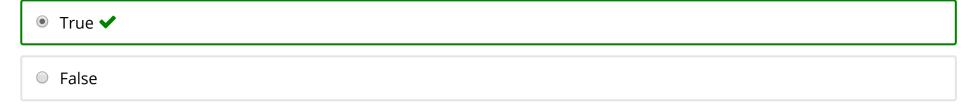
Mark the following statements as true or false.

The EM algorithm monotonically increases the likelihood of the data with each iteration. In other words, the likelihood after iteration i+1 is greater than or equal to the likelihood after iteration i, for all i.



O False

Depending on the initialization, the likelihood of the data the algorithm converges to may be different.



We are estimating a mixture model with K components. During random initialization, p_1 was assigned to be zero. p_1 could become non-zero as the algorithm iterates.

O True	
● False ✓	

Solution:

The M-step chooses the parameter values that maximize the likelihood. Therefore, the previous iteration cannot have parameters of a greater likelihood than the current iteration.

The initial values provide a "starting point" for the algorithm, different initializations lead to different results.

The expected value after the E-step will be zero, therefore p_1 is recomputed in the M-step to be a sum of zeros.

Submit You have used 3 of 3 attempts

1 Answers are displayed within the problem

4. (2)

1/1 point (graded)

Consider a 1-dim Gaussian mixture model with two components. We set the mixture with $\mu_1=1$, $\mu_2=1$, $\sigma_1=0.5$, $\sigma_2=0.5$. The mixing proportions are set differently for the two components: $p_1=0.01$ and $p_2=0.99$. Is it the case that $\mu_1=\mu_2$ and $\sigma_1=\sigma_2$ after running the EM algorithm regardless of the data?

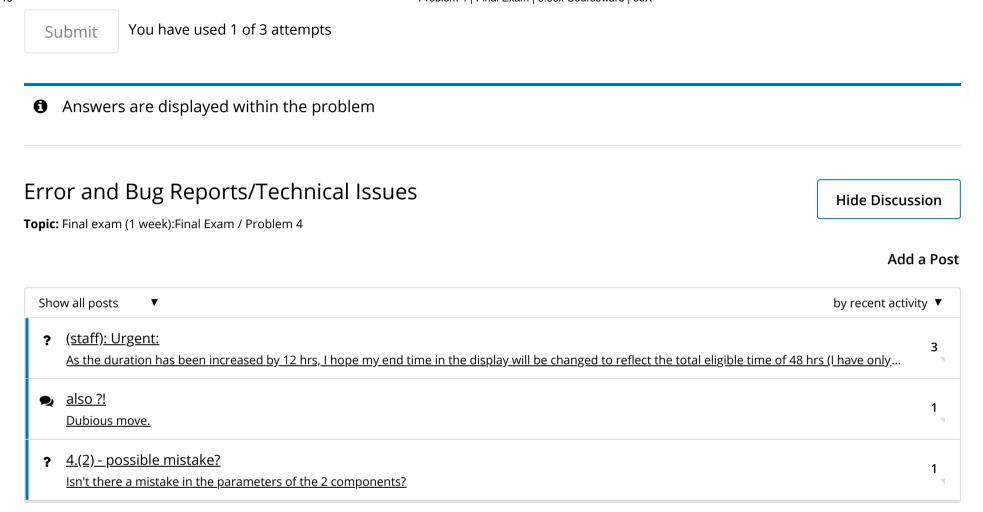
- Yes
- No

Will $p_1 = 0.01$ and $p_2 = 0.99$ also hold at convergence? (There is no answer box for this question.)

Solution:

Initially, it will be 99 times more likely that the points come from cluster 2 as opposed to cluster 1. However, the μ and σ that maximize the probability of the clusters generating the given points are the same.

 $p\left(cluster1|examplei
ight) = rac{0.01x}{0.01x + 0.99x} \ p\left(cluster2|examplei
ight) = rac{0.99x}{0.01x + 0.99x}$ Since all $p\left(cluster1|examplei
ight) * 0.99 = p\left(clusterz|examplei
ight)$ for all $i, p_1 * 99 = p_2$. This holds by induction.



© All Rights Reserved