You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

Course › Unit 4 Unsupervised Learning (2 weeks) › Homework 5 › 3. EM Algorithm

3. EM Algorithm

Extension Note: Homework 4 due date has been extended by 1 day to **August 17 23:59UTC** .

Consider the following mixture of two Gaussians:

$$p\left(x;\theta\right) = \pi_1\mathcal{N}\left(x;\mu_1,\sigma_1^2\right) + \pi_2\mathcal{N}\left(x;\mu_2,\sigma_2^2\right)$$

This mixture has parameters $\theta = \{\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$. They correspond to the mixing proportions, means, and variances of each Gaussian. We initialize θ as $\theta_0 = \{0.5, 0.5, 6, 7, 1, 4\}$.

We have a dataset $\mathcal{D}$ with the following samples of x: $x^{(0)} = -1$, $x^{(1)} = 0$, $x^{(2)} = 4$, $x^{(3)} = 5$, $x^{(4)} = 6$.

We want to set our parameters θ such that the data log-likelihood $l\left(\mathcal{D};\theta\right)$ is maximized:

Recall that we can do this with the EM algorithm. The algorithm optimizes a lower bound on the log-likelihood, thus iteratively pushing the data likelihood upwards. The iterative algorithm is specified by two steps applied successively:

1. E-step: infer component assignments from current $\theta_0 = \theta$ (complete the data)

$$p\left(y=k \mid x^{(i)}\right) := p\left(y=k \mid x^{(i)}; \theta_0\right), \text{ for } k=1,2, \text{ and } i=0,\ldots,4.$$

2. M-step: maximize the expected log-likelihood

$$\tilde{l}\,(D;\theta) := \sum_i \sum_k p\left(y=k \mid x^{(i)}\right) \log \frac{p\left(x^{(i)}, y=k;\theta\right)}{p\left(y=k \mid x^{(i)}\right)}$$

with respect to θ while keeping $p\left(y=k \mid x^{(i)}\right)$ fixed.

To see why this optimizes a lower bound, consider the following inequality:

$$\log p\left(x;\theta\right) \quad = \log \sum_y p\left(x,y;\theta\right)$$

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

$$= \log \mathbb{E}_{y \sim q(y|x)} \left[\frac{}{q(y|x)} \right]$$

$$\geq \mathbb{E}_{y \sim q(y|x)} \left[\log \frac{p(x, y; \theta)}{q(y|x)} \right]$$

$$= \sum_{y} q(y|x) \log \frac{p(x, y; \theta)}{q(y|x)}$$

where the inequality comes from **Jensen's inequality** . EM makes this bound tight for the current setting of θ by setting $q(y|x)$ to be $p(y \mid x; \theta_0)$.

Note: If you have taken *6.431x Probability–The Science of Uncertainty*, you could review the video in *Unit 8: Limit Theorems and Classical Statistics, Additional Theoretical Material, 2. Jensen's Inequality*.

Likelihood Function

1/1 point (graded)

What is the log-likelihood of the data $l(\mathcal{D}; \theta)$ given the initial setting of θ? Please round to the nearest tenth.

Note: You will want to write a script to calculate this, using the natural log (np.log) and np.float64 data types.

-24.51253233

✓ **Answer:** -24.5

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

$$\begin{aligned} P(\mathcal{D};\theta) &= \prod_{i=0}^{4} p(x;\theta) \\ &= \prod_{i=0}^{4} \pi_1 \mathcal{N}\left(x^{(i)};\mu_1,\sigma_1^2\right) + \pi_2 \mathcal{N}\left(x^{(i)};\mu_2,\sigma_2^2\right) \end{aligned}$$

Taking the log gives:

$$l(\mathcal{D};\theta) = \sum_{i=0}^{4} \log\left(\pi_1 \mathcal{N}\left(x^{(i)};\mu_1,\sigma_1^2\right) + \pi_2 \mathcal{N}\left(x^{(i)};\mu_2,\sigma_2^2\right)\right)$$

We then evaluate each Gaussian using the standard formulation:

$$\mathcal{N}\left(x;\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Submit

You have used 2 of 3 attempts

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

E-Step

1/1 point (graded)

What is the formula for $p\left(y=k \mid x,\theta\right)$? Write in terms of π_k, π_1, π_2, N_k, N_1, and N_2 (where $N_k = \mathcal{N}\left(x \mid \mu_k, \sigma_k^2\right)$).

pi_k*N_k/(pi_1*N_1+pi_2

✓ **Answer:** (pi_k * N_k) / (pi_1 * N_1 + pi_2 * N_2)

$$\frac{\pi_k \cdot N_k}{\pi_1 \cdot N_1 + \pi_2 \cdot N_2}$$

STANDARD NOTATION

Solution:

Following Bayes Rule we have:

$$p\left(y \mid x\right) = \frac{p\left(y\right)p\left(x \mid y\right)}{\sum_{y'} p\left(y'\right)p\left(x|y'\right)}$$

For this problem, this equates to:

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

E-Step Weights

5/5 points (graded)

For each of the given data points say which Gaussian (1 or 2) they are given more weight towards in the first E-step using the given setting of θ_0. This is, answer 2 if $p\left(y=2 \mid x, \theta_0\right) > p\left(y=1 \mid x, \theta_0\right)$ and 1 otherwise.

$x^{(0)}$: 2 ✓

$x^{(1)}$: 2 ✓

$x^{(2)}$: 2 ✓

$x^{(3)}$: 1 ✓

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

M-Step

3/3 points (graded)

Fixing $p\left(y=k \mid x, \theta_0\right)$, we want to update θ such that our lower bound is maximized.

What is the optimal $\hat{\mu}_k$? Answer in terms of $x^{(1)}$, $x^{(2)}$, and γ_{k1}, γ_{k2}, which are defined to be $\gamma_{ki} = p\left(y=k \mid x^{(i)}; \theta_0\right)$

(For ease of input, use subscripts instead superscripts, i.e. type `x_i` for $x^{(i)}$. Type `gamma_ki` for γ_{ki}.)

(gamma_k1*x_1+gamma_k2*x_2)/(gamma_k1+gamma_k2) ✓

Answer: (gamma_k1 * x_1 + gamma_k2 * x_2) / (gamma_k1 + gamma_k2)

$$\frac{\gamma_{k1} \cdot x_1 + \gamma_{k2} \cdot x_2}{\gamma_{k1} + \gamma_{k2}}$$

What is the optimal $\hat{\sigma}_k^2$? Answer in terms of $x^{(1)}$, $x^{(2)}$, γ_{k1} and γ_{k2}, which are defined as above to be $\gamma_{ki} = p\left(y=k \mid x^{(i)}; \theta_0\right)$, and $\hat{\mu}_k$.

(Type `hatmu_k` for $\hat{\mu}_k$. As above, for ease of input, use subscripts instead superscripts, i.e. type `x_i` for $x^{(i)}$. Type `gamma_ki` for γ_{ki}.)

$\gamma_{k1}+\gamma_{k2}$

What is the optimal $\hat{\pi}_k$? Answer in terms of γ_{k1} and γ_{k2}, which are defined as above to be $\gamma_{ki} = p\left(y=k \mid x^{(i)};\theta_0\right)$,

(As above, type `gamma_ki` for γ_{ki}.)

Note: that you must account for the constraint that $\pi_1+\pi_2=1$ where $\pi_1, \pi_2 \geq 0$.

Note: If you know that some aspect of your formula equals an exact constant, simplify and use this number, i.e. $\gamma_{11}+\gamma_{21}=1$.

(gamma_k1+gamma_k2)/2

✓ **Answer:** (gamma_k1 + gamma_k2) / 2

$\frac{\gamma_{k1}+\gamma_{k2}}{2}$

STANDARD NOTATION

Solution:

The function we are optimizing is now:

Taking $\frac{\partial}{\partial \mu_k}$ and setting to 0 gives:

$$\begin{aligned}\frac{\partial}{\partial \mu_k}\sum_i\sum_k \gamma_{ki}\log\left(\pi_k \mathcal{N}\left(x^{(i)};\mu_k,\sigma_k^2\right)\right) &= \sum_i \gamma_{ki}\frac{\partial}{\partial \mu_k}\log\left(\pi_k \mathcal{N}\left(x^{(i)};\mu_k,\sigma_k^2\right)\right)\\ &= \sum_i \gamma_{ki}\frac{\partial}{\partial \mu_k}\left(\log\left(\frac{1}{\sqrt{2\pi\sigma_k^2}}\right) - \frac{\left(x^{(i)}-\mu_k\right)^2}{2\sigma_k^2}\right)\\ &= \sum_i \gamma_{ki}\frac{x^{(i)}-\mu_k}{\sigma_k^2} = 0\end{aligned}$$

Separating out μ_k gives:

$$\mu_k = \frac{\sum_i \gamma_{ki} x^{(i)}}{\sum_i \gamma_{ki}}$$

We can interpret this as a weighted average of the data points, normalized by the "total mass" assigned to Gaussian k. The weight is the probability that point $x^{(i)}$ "belongs" to Gaussian k.

Solving for σ_k^2 is similar:

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

$$\sum_i \quad \partial\sigma_k^2 \quad \sqrt{2\pi\sigma_k^2} \quad 2\sigma_k^2$$

$$= \sum_i \gamma_{ki} \left(-\frac{1}{2\sigma_k^2} + \frac{\left(x^{(i)} - \mu_k\right)^2}{2\sigma_k^4} \right) = 0$$

Separating out σ_k^2 gives:

$$\sigma_k^2 = \frac{\sum_i \gamma_{ki} \left(x^{(i)} - \mu_k\right)^2}{\sum_i \gamma_{ki}}$$

Finally we solve for π_k while including a lagrange multiplier for the constraint that $\sum_k \pi_k = 1$.

$$\begin{aligned} \frac{\partial}{\partial \pi_k} \sum_i \sum_k \gamma_{ki} \log\left(\pi_k \mathcal{N}\left(x^{(i)}; \mu_k, \sigma_k^2\right)\right) + \lambda\left(\sum_k \pi_k - 1\right) &= \sum_i \gamma_{ki} \frac{\partial}{\partial \pi_k} \log\left(\pi_k\right) + \frac{\partial}{\partial \pi_k} \lambda\left(\sum_k \pi_k - 1\right) \\ &= \frac{\sum_i \gamma_{ki}}{\pi_k} + \lambda = 0 \end{aligned}$$

Giving $\pi_k = -\frac{\sum_i \gamma_{ki}}{\lambda}$.

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

Combining the two gives:

$$\lambda = -\sum_i \sum_k \gamma_{ki}$$

which we recognize as N, the total number of points. Thus $\hat{\pi}_k$ is $\frac{\sum_i \gamma_{ki}}{N}$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Training 1

1/1 point (graded)

In the first M-step, which Gaussian will shift to the left more (relatively)?

○ Gaussian 1

● Gaussian 2 ✓

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

Intuitively, Gaussian 2 is influenced most by the points $x^{(0)}$,$x^{(1)}$, and so it will move to the left. Gaussian 1 will be more influenced by the points at $x^{(2)}$, $x^{(3)}$ and $x^{(4)}$ and so it will not move very much to the left. If we computed the actual values, we would see that the updated means for the two Gaussians are approximately $\mu_1 = 5.1317$ and $\mu_2 = 1.4710$.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Training 2

1/1 point (graded)

In the first M-step, which Gaussian's variance will increase more (relatively)?

○ Gaussian 1

● Gaussian 2 ✓

Solution:

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Training 3

1/1 point (graded)

After convergence, which variance will be larger?

● σ_1^2 ✓

○ σ_2^2

Solution:

Gaussian 1 will be centered around the cluster of 3 points on the right, while Gaussian 2 will be centered around the 2 points on the left. Gaussian 1 will have larger variance because of the larger spread of the right cluster.

Submit You have used 1 of 1 attempt

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Homework 5 / 3. EM Algorithm

Add a Post

Show all posts ▼ by recent activity ▼

Likelihood Function
I have a doubt regarding the way we're supposed to take the likelihood in the EM algorithm. I initially thought we were supposed to use the likeli...
3

Is there something wrong with my code about caculating the log-likelihood？
e.g. x0 1.caculating the N(x;μ1,σ21) and N(x;μ2,σ22) p1_x0=multivariate_normal.pdf(-1,6,1) p2_x0=multivariate_normal.pdf(-1,7,2) 2.caculating the...
6

(Staff) E-Step
Staff, my E-Step is correct, since there are only two K.
1

? M-Step without Norm
I think time has finally run out on this homework. I did well on most but I was unable to figure out how to answer "What is the optimal σ^2k" with...
2

? M-Step
Hi there! The M-step problem looks like rather adopting the definitions given on the lecture for the case of just 2 points, but for some reason all ...
3

? [STAFF]Could you check my answer?
Could you check my answer of M-step? I'm pretty sure I'm right.
3

? M-step write in terms of ...
Hi, so I was doing the work fine yesterday, but today I looked at it again, and I can't seem to understand, why we are using only two points, x1 an...
5

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:17**

☑ E-Step, how to enter a sum from 1 to k? 5
E-Step, how to enter a sum from 1 to k?

☑ Likelihood Function 6

💬 E-Step Weights: results surprised me! 2
By running my script for the calculation I got the correct answers. But they were very different from my initial naive guess by inspection! Another ...

☑ E-step - symbols and notation 5
I think I know how to calculate $p(y=k|x,\theta)$ but I don't see how to do this with the symbols given. Any pointers?

? [Training 3] How to debug this part? 4