3. EM Algorithm

Bookmark this page

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

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 $\theta=\left\{\pi_1,\pi_2,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\right\}$. They correspond to the mi 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: $x^{(0)}=-1$, $x^{(1)}=0$, $x^{(2)}=4$,

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

$$\underset{\theta}{\operatorname{argmax}}\sum_{i=0}^{4}\log p\left(x^{(i)};\theta\right).$$

Recall that we can do this with the EM algorithm. The algorithm optimizes a lower boun thus iteratively pushing the data likelihood upwards. The iterative algorithm is specifie 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

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

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:

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

Limit Theorems and Classical Statistics, Additional Theoretical Material, 2. Jensen's Ine

Likelihood Function

0/1 point (graded)

What is the log-likelihood of the data given the initial setting of ? Please roun

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

2.13 ✗

Submit You have used 3 of 3 attempts

E-Step

1/1 point (graded)

What is the formula for ? Write in terms of , , , , , and).

pi_k * N_k / (pi_1 * N_1 + pi_2 * N_ ✓

? STANDARD NOTATION

Submit You have used 1 of 3 attempts

E-Step Weights

5/5 points (graded)

For each of the given data points say which Gaussian (1 or 2) they are given more weig E-step using the given setting of . This is, answer 2 if otherwise.

2 ✓

2 ✓

Fixing , we want to update such that our lower bound is maximized.

What is the optimal ? For simplicity, assume we only have two data points and question. Answer in terms of , , and , , which are defined to be

(For ease of input, use subscripts instead superscripts, i.e. type x_i for . Type gamma_

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

✓

What is the optimal ? Answer in terms of , , and , which are defined a
, and .

(Type hatmu_k for . As above, for ease of input, use subscripts instead superscripts, i.
gamma_ki for .)

(gamma_k1 * (x_1 - hatmu_k)^2 + gamma_k2 * (x_2 - hatmu_k)^2)/ (gamma_k1 + gamr

✓

What is the optimal ? Answer in terms of and , which are defined as above to
,

(As above, type gamma_ki for .)

Note: that you must account for the constraint that where .

Note: If you know that some aspect of your formula equals an exact constant, simplify a
.

(gamma_k1 + gamma_k2) / 2

✓

? STANDARD NOTATION

Submit You have used 1 of 1 attempt

Training 2

0/1 point (graded)

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

● Gaussian 1

○ Gaussian 2

✖

Submit You have used 1 of 1 attempt

Training 3

0/1 point (graded)

After convergence, which variance will be larger?

○

●

✖

< Previous Next >

Submit You have used 1 of 1 attempt

edX

About

Affiliates

edX for Business

Open edX

Careers

- ☑ Likelihood Function: which one should we calculate?
- ? In the first part, where should I use np.float64?
  where should I use np.float64? I started my script with import numpy as np from scipy.stats import norm dty
- ? Training 3 - after convergence
  I don't really see how to qualitatively answer this question. Do we need to run EM to see how the variance ch
- ? M - step (why not drop the denominator?)
  In the M - step described above, why don't we drop the denominator inside the log function, which is assum
- "hatmu_k" is not accepted as a variable in the formula

Connect

Blog
Contact Us
Help Center
Security
Media Kit

GET IT ON Google play

Download on the App Store