2. Recap of Maximum Likelihood Estimation for Multinomial and Gaussian Models

So far, in clustering we have assumed that the data has no probabilistic generative model attached to it, and we have used various iterative algorithms based on similarity measures to come up with a way to group similar data points into clusters. In this lecture, we will assume an underlying probabilistic generative model that will lead us to a natural clustering algorithm called the **EM algorithm** .

While a “hard" clustering algorithm like k-means or k-medoids can only provide a cluster ID for each data point, the EM algorithm, along with the generative model driving its equations, can provide the posterior probability (“soft" assignments) that every data point belongs to any cluster.

The EM algorithm will also form the basis for a portion of **Project 4** in which we explore collaborative filtering via Gaussian mixtures.

MLE for Multinomial and Gaussian Models

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

MLE under Gaussian Noise I

1/1 point (graded)

Let $Y_i = \theta + N_i, i = 1, \ldots, n$, where θ is an unknown parameter and N_i are iid Gaussian random variables with zero mean. Upon observing Y_i's, what is the maximum likelihood estimate of θ?

Choose the correct expression from options below.

Hint: For this problem, think of the distributional property of the random variables Y_i and how they are derived from N_i's. Also, remember that N_i are iid Gaussian random variables.

- $\hat{\theta} = \Pi_{i=1}^{n} Y_i$
- $\hat{\theta} = \frac{\Pi_{i=1}^{n} Y_i}{n}$
- $\hat{\theta} = \sum_{i=1}^{n} Y_i$
- (selected, correct) $\hat{\theta} = \frac{\sum_{i=1}^{n} Y_i}{n}$

Solution:

Y_i's are iid Gaussian with mean θ. As seen before, the ML estimator of θ is $\sum_{i=1}^{n} \frac{Y_i}{n}$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

MLE under Gaussian Noise II

0/1 point (graded)

Would the ML estimator of θ change if the N_i's are **independent** Gaussians with **possibly different variances** $\sigma_1^2, \ldots, \sigma_n^2$ but **same zero mean**? Assume that σ_i^2 are **known** constants.

- Yes ✔
- No (selected)

✖

Solution:

The log-likelihood (with possibly different variances) is

$$\log P\left(Y_1, \ldots, Y_n | \theta, \sigma_1^2, \ldots, \sigma_n^2\right) = -\frac{1}{2}\sum_{i=1}^{n} \log\left(2\pi\sigma_i^2\right) - \sum_{i=1}^{n} \frac{\left(Y_i - \theta\right)^2}{2\sigma_i^2}.$$

Note that here we cannot take σ_i^2 out of the summation. Thus, to solve this, we need to take the derivative of the log-likelihood with respect to θ set it equal to zero.

Doing so, we obtain

$$\sum_i \frac{Y_i}{\sigma_i^2} = \theta \sum_i \frac{1}{\sigma_i^2},$$

which yields

$$\hat{\theta} = \frac{\sum_i \frac{Y_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}},$$

which is not necessarily the same estimator as before. That is, the sample average of Y_i's is no longer the ML estimate of θ.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

MLE under Gaussian Noise III

1/1 point (graded)

Now, let $Y_i = \theta + N_i, i = 1, \ldots, n$, where θ is an unknown parameter. This time, let N_i for all $i = 1, \ldots, n$ each be a linear combination of ℓ **zero-mean, independent** Gaussian random variables with possibly different variances. That is, let $N_i = N_{i,1} + \cdots + N_{i,\ell}$, where $N_{i,j}$ for a given i are independent with possibly different variances. Let $\{N_{i,j} : i = 1, \ldots, n \text{ and } j = 1, \ldots, \ell\}$ be independent random variables, but with the condition that for a given j the variance of $N_{i,j}$ is the same for $i = 1, \ldots, n$.

Is the ML estimator of θ the same as the one we found in the problem "MLE under Gaussian Noise I"?

- (•) Yes
- ( ) No

✔

Solution:

Yes. The scenario explained here essentially is similar to the one in the problem "MLE under Gaussian Noise I".

First, note that N_i are all independent of each other for $i = 1, \ldots, n$ since $\{N_{i,j} : i = 1, \ldots, n \text{ and } j = 1, \ldots, \ell\}$ are independent random variables.

Second, $N_i, i = 1, \ldots, n$ are identically distributed Gaussian random variables. This is because of the two following reasons. First, a linear combination of Gaussian random variables is Gaussian, and in this case $N_i = N_{i,1} + \cdots + N_{i,\ell}$ for $i = 1, \ldots, n$ have the same variance because of the fact that for a given j the variance of $N_{i,j}$ is the same for $i = 1, \ldots, n$. Second, the fact that all $N_{i,j}$'s are zero-mean means that N_i are all zero-mean.

Since we have established that N_i are iid, zero-mean Gaussian random variables with the same variance, we conclude that the MLE for θ does not change.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 16. Mixture Models; EM algorithm / 2. Recap of Maximum Likelihood Estimation for Multinomial and Gaussian Models

Add a Post

Show all posts by recent activity

- [MLE under Gaussian Noise II] Any good reference? 3
  I understand the derivation explained in the answer, but it's a little too mathematical. I want t...
- [STAFF] Please link to explanation? 1
  I have been looking for hours to get a grasp on this but.... Any link to an explanation of the q...
- these were very hard. 3
  I wish this was discussed more in the lecture
- Confused 3
  Sometimes I feel we taught in the lecture video how to garden and then asked in the exercice...