

Unit 4 Unsupervised Learning (2

Course > weeks)

5. Maximum Likelihood Estimate

> Lecture 15. Generative Models >

Audit Access Expires May 11, 2020

You lose all access to this course, including your progress, on May 11, 2020.

5. Maximum Likelihood Estimate Maximum Likelihood Estimate





Video

Download video file

Transcripts

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

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

Number of Parameters

1/1 point (graded)

For the following set of questions, let us consider generating documents that are English letter sequences (assume no spaces or punctuation), i.e. the vocabulary $W=\{a,b,c\ldots,z\}$ is made up of all the letters in the English alphabet.

We would like to generate documents using this vocabulary using a multinomial model M. As described in the lecture, what is the minimal number of parameters that the model M should have? Enter your answer below.

25 **✓ Answer:** 25

Solution:

Recall from the lecture that for multinomial generative models we have a parameter θ_w for each word $w \in W$. However, since the parameters should sum up to one, we can express one of the parameters as 1 minus the sum of all others. Since the vocabulary size for this example is 26, our model M can have 25 parameters to express the probability of each letter.

Submit

You have used 1 of 2 attempts

1 Answers are displayed within the problem

Note: For those of you who have completed **18.6501x** (**Fundamentals of Statistics**), it may be useful at this point to review or recall the lectures on Maximum Likelihood Estimation, which is perhaps the most powerful and general estimation method.

Maximum Likelihood Estimate

1/1 point (graded)

Let $\theta^* = \theta_a^*, \theta_b^*, \dots, \theta_z^*$ be the parameters of the multinomial model M^* that maximize the likelihood of generating a document D.

Further, it is known that the letter 'e' is twice as likely to occur as the letter 'z' in

document D.

Which of the following options is a correct expression relating θ_e^* and θ_z^* ?

- $\bigcirc heta_z^* = 2 heta_e^*$
- $lackbox{0}{ heta_e^*}=2 heta_z^*$
- $igcap heta_z^* = heta_e^*$
- $\bigcirc heta_z^* + heta_e^* = 2$



Solution:

Recall from the lecture that for any $w \in W$, the maximum

$$heta_{w}^{*} = rac{ ext{count}\left(w
ight)}{\sum_{w' \in W} count\left(w'
ight)}$$

Since $count\left(e\right)=2count\left(z\right)$, we can conclude that $heta_{e}^{*}=2 heta_{z}^{*}$

Submit

You have used 1 of 2 attempts

1 Answers are displayed within the problem

Maximum Likelihood Estimate for Poisson Distribution

2/2 points (graded)

Maximum Likelihood Estimate (MLE) is a very general method that can be applied to both continuous and discrete distributions. In this problem, we assume we have a training data $x_1, x_2 \ldots, x_n$ that are drawn from a Poisson distribution, with

probability mass function (pmf)

$$P(X=x) = rac{\lambda^x e^{-\lambda}}{x!}.$$

We want to use MLE to fit the parameter λ with the training data. To do so, we first compute the log likelihood of our training data, or in other words, log of the probability of obtaining the sample x_1, x_2, \ldots, x_n given the model and where x_i are independent. The log likelihood is...

- $\log \lambda \sum_i x_i + n\lambda + \sum_i \log \left(x_i!
 ight)$
- $ullet \log \lambda \sum_i x_i n\lambda \sum_i \log \left(x_i!\right)$
- $igcup_i \log \lambda \prod_i x_i n\lambda \prod_i \log \left(x_i!
 ight)$
- igcirc $\log \lambda \prod_i x_i + n\lambda + \prod_i \log \left(x_i!
 ight)$



In the next step, we maximize this log likelihood function by taking the derivative. What is the resulting estimate for λ ?

- $igotimes rac{1}{n} \sum_i x_i$
- $\bigcap rac{1}{n} \prod_i x_i$
- $igcap \sum_i x_i$
- $\bigcap \prod_i x_i$



Is it in accordance with the definition of λ in Poisson distribution? (There is no answer box for this question.)

Solution:

The loglikelihood of the data is:

$$egin{aligned} \log \prod_{i} P\left(X = x_i
ight) &= \log \prod_{i} rac{\lambda^{x_i} e^{-\lambda}}{x_i!} \ &= \sum_{i} \log\left(\lambda^{x_i}
ight) + \log\left(e^{-\lambda}
ight) - \log\left(x_i!
ight) \ &= \log \lambda \sum_{i} x_i - n\lambda - \sum_{i} \log\left(x_i!
ight) \end{aligned}$$

Take the derivative with respect to λ , we have

$$rac{1}{\lambda}\sum_i x_i - n \ \equiv 0$$
 $\lambda = rac{1}{n}\sum_i x_i.$

This is in accordance with the fact that λ is the expectation of a Poisson variable with parameter λ .

Submit

You have used 1 of 2 attempts

1 Answers are displayed within the problem

Discussion

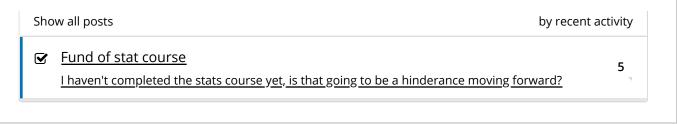
Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 15. Generative Models / 5. Maximum Likelihood Estimate

Add a Post

5. Maximum Likelihood Estimate | Lecture 15. Ge...

https://courses.edx.org/courses/course-v1:MITx+...



© All Rights Reserved