Resources

Next >

1 Audit Access Expires Dec 22, 2020 You lose all access to this course, including your progress, on Dec 22, 2020.

Progress

Dates

Course

Previous

Course / Unit 4 Unsupervised Learning (2 weeks) / Lecture 15. Generative Models

Discussion

6. MLE for Multinomial Distribution ☐ Bookmark this page Exercises due Nov 17, 2020 18:59 EST Completed **Maximum Likelihood Estimate** to find a martinormal winter fit tins data the best. And again, at this point, what we were discussing, I just told you all the time that we have a single document. In reality, you can assume that even if you have a collection of documents, the whole story is exactly the same because whenever we're making our assumption that we are generating the words, these words are generated independently. So pretty much, you can do exactly the same formula by concatenating many of your documents into a single document and then repeating the same story. So now we're done with the discussion of estimation for multinomial. 2:05 / 2:32 cc **66** ▶ 1.25x X And with that, we are ready to start

In the following problems, we will derive the maximum likelihood estimates for a multinomial model with more than 2 parameters. We will employ the method of lagrange multipliers for the optimization problem.

Deriving MLE for a General Multinomial Model: Likelihood

Let the document D be a sequence of words w_1, \ldots, w_n from a collection W consisting of N words. For simplicity, we assume that w_i 's are independent, and that the probability of a word w is given by the parameter θ_w , and denote

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Let $P(D|\theta)$ be the probability of D being generated by the simple model described above. Find $P(D|\theta)$.

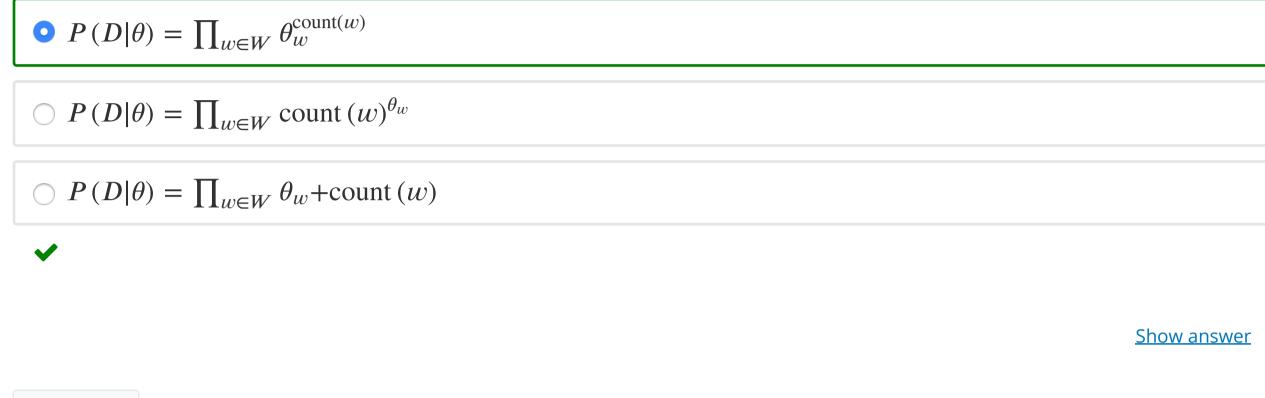
 $\bigcirc P(D|\theta) = \sum_{w \in W} \theta_w^{\text{count}(w)}$

Video

Download video file

1/1 point (graded)

by $\theta = \{\theta_w\}_{w \in W}$.



What are the constraints on the parameters θ_w in the model described in the previous problem?

Constraints on the Parameters

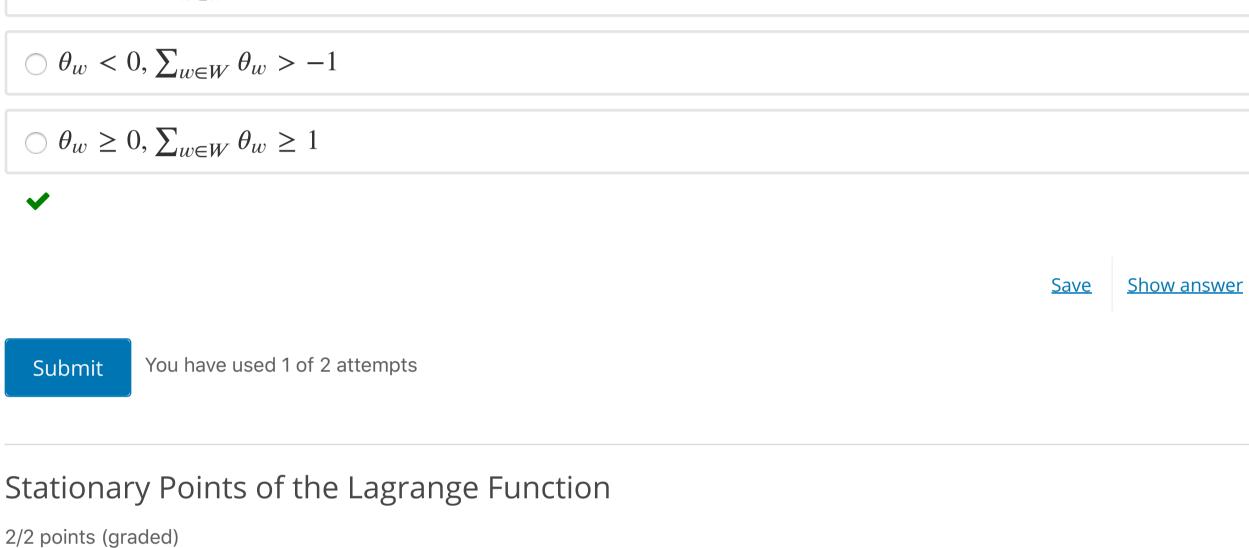
You have used 2 of 2 attempts

 $\bullet w \ge 0, \sum_{w \in W} \theta_w = 1$

1/1 point (graded)

Submit

 $\theta_w \ge 0, \sum_{w \in W} \theta_w < 1$



Maximizing $P(D|\theta)$ is equivalent to maximizing $\log P(D|\theta)$, so we take the natural logarithm on both sides of the

equation above to bring down the exponents:

Recall that θ is subject to the following constraint:

Method of Lagrange Multipliers

Define the Lagrange function:

where λ is a constant scalar.

 $\theta_w = \frac{-\lambda}{\operatorname{count}(w)}$

 $\theta_w = -\lambda \operatorname{count}(w)$

 $\bullet_{w} = \frac{-\operatorname{count}(w)}{\lambda}$

 $\bigcirc \lambda = \sum (\theta_w \text{count}(w))$

Submit

Discussion

You have used 1 of 3 attempts

Topic: Unit 4 Unsupervised Learning (2 weeks): Lecture 15. Generative Models / 6. MLE for

Previous

 $\log P(D|\theta) = \sum_{w \in W} \operatorname{count}(w) \log \theta_w.$

 $\sum_{w \in W} \theta_w = 1.$

The maximum likelihood estimate of θ is the value of θ that maximizes the likelihood function:

 $P(D|\theta) = \prod (\theta_w)^{\operatorname{count}(w)}.$

To maximize
$$\log P(D|\theta)$$
 subject to the contraint $\sum_{w \in W} \theta_w = 1$, we use the Lagrange multiplier method.

 $L = \log P(D|\theta) + \lambda \left(\sum_{w \in W} \theta_w - 1\right)$

Then, find the stationary points of L by solving the equation $\nabla_{\theta}L=0$. The components of this equation are

$$\frac{\partial}{\partial \theta_w} \left(\log P(D|\theta) + \lambda \left(\sum_{w \in W} \theta_w - 1 \right) \right) = 0 \quad \text{for all } w \in W.$$

Solve for θ_w from the above equation. Choose the right answer for θ_w from options below.

 $\theta_w = \lambda \text{count}(w)$

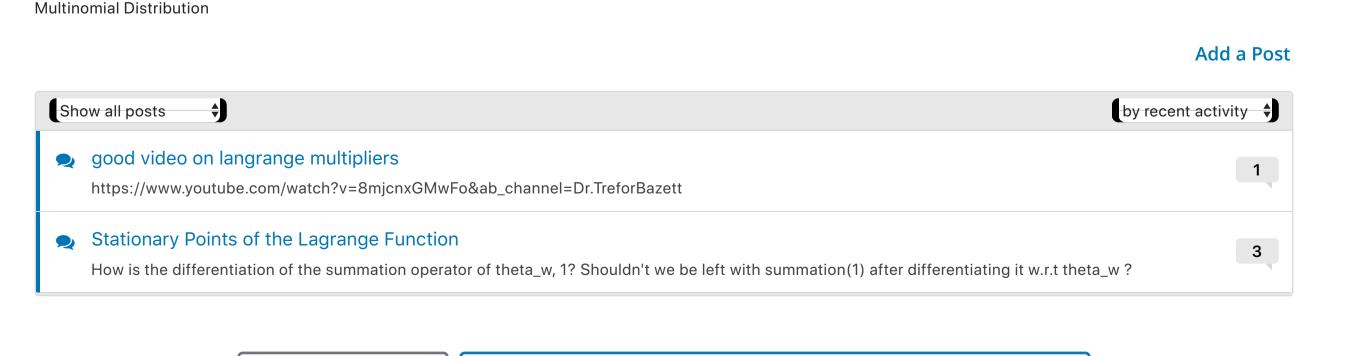
Now, apply the constraint that
$$\sum_{w \in W} \theta_w = 1$$
 to the answer above to obtain λ .
$$\lambda = \frac{1}{1 + 1} \sum_{w \in W} \operatorname{count}(w)$$

$$0 \quad \lambda = \sum_{w \in W} \operatorname{count}(w)$$

$$0 \quad \lambda = \sum_{w \in W} \operatorname{count}(w)$$

$$0 \quad \lambda = \sum_{w \in W} \operatorname{count}(w)$$

Find
$$\theta_w$$
 that maximizes $\log P(D|\theta)$ subject to $\sum_{w\in W}\theta_w=1$. (There is no answer box for this final question.)



© All Rights Reserved

Show





Terms of Service & Honor Code Privacy Policy **Accessibility Policy** Trademark Policy Sitemap

Legal

Blog Contact Us Help Center Media Kit Donate

Connect

Next >



深圳市恒宇博科技有限公司 <u>粤ICP备17044299号-2</u>

© 2020 edX Inc. All rights reserved.

Hide Discussion