6. MLE for Multinomial Distribution

Bookmark this page

Exercises due Nov 17, 2020 18:59 EST Completed

Maximum Likelihood Estimate

2:05 / 2:32 1.25x

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.

And with that, we are ready to start

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Deriving MLE for a General Multinomial Model: Likelihood

1/1 point (graded)

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.

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

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

Find $P(D|\theta)$.

- ○ $P(D|\theta) = \sum_{w \in W} \theta_w^{\text{count}(w)}$
- ● $P(D|\theta) = \prod_{w \in W} \theta_w^{\text{count}(w)}$
- ○ $P(D|\theta) = \prod_{w \in W} \text{count}(w)^{\theta_w}$
- ○ $P(D|\theta) = \prod_{w \in W} \theta_w + \text{count}(w)$

✓

Show answer

Submit You have used 2 of 2 attempts

Constraints on the Parameters

1/1 point (graded)

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

- ● $\theta_w \geq 0, \sum_{w \in W} \theta_w = 1$
- ○ $\theta_w \geq 0, \sum_{w \in W} \theta_w < 1$
- ○ $\theta_w < 0, \sum_{w \in W} \theta_w > -1$
- ○ $\theta_w \geq 0, \sum_{w \in W} \theta_w \geq 1$

✓

Save | Show answer

Submit You have used 1 of 2 attempts

Stationary Points of the Lagrange Function

2/2 points (graded)

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

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

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:

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

Recall that θ is subject to the following constraint:

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

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

Method of Lagrange Multipliers
Show

Define the Lagrange function:

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

where λ is a constant scalar.

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

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

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

- ○ $\theta_w = \frac{-\lambda}{\text{count}(w)}$
- ○ $\theta_w = \lambda \text{count}(w)$
- ○ $\theta_w = -\lambda \text{count}(w)$
- ● $\theta_w = \frac{-\text{count}(w)}{\lambda}$

✓

Now, apply the constraint that $\sum_{w \in W} \theta_w = 1$ to the answer above to obtain λ.

$\lambda =$

- ● $\lambda = -\sum_{w \in W} \text{count}(w)$
- ○ $\lambda = \sum_{w \in W} \text{count}(w)$
- ○ $\lambda = -\sum_{w \in W} (\theta_w \text{count}(w))$
- ○ $\lambda = \sum_{w \in W} (\theta_w \text{count}(w))$

✓

Find θ_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.)

Save | Show answer

Submit You have used 1 of 3 attempts

Discussion

Hide Discussion

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

Add a Post

Show all posts | by recent activity

good video on langrange multipliers
https://www.youtube.com/watch?v=8mjcnxGMwFo&ab_channel=Dr.TreforBazett
1

Stationary Points of the Lagrange Function
How is the differentiation of the summation operator of theta_w, 1? Shouldn't we be left with summation(1) after differentiating it w.r.t theta_w ?
3