6. MLE for Multinomial Distribution

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

Maximum Likelihood Estimate

0:00 / 0:00 1.0x

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 multino than 2 parameters. We will employ the method of lagrange multipliers for the optimizati

Let the document D be a sequence of words $w_1, \ldots, w_n$ from a collection W consisting of we assume that w_i's are independent, and that the probability of a word w is given by th denote by $\theta = \{\theta_w\}_{w \in W}$.

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

Find $P(D|\theta)$.

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

1/1 point (graded)

What are the constraints on the parameters θ_w in the model described in the previous p

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

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 fur

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

Maximizing $P(D|\theta)$ is equivalent to maximizing $\log P(D|\theta)$, so we take the natural logarith 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 mult

Method of Lagrange Multipliers

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

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

Now, apply the constraint that $\sum_w \theta_w = 1$ to the answer above to obtain λ.

Previous Next

$\lambda =$