Deriving MLE estimates for a general multinomial distribution

1/1 point (graded)
In this problem, we will derive the maximum likelihood estimates for a multinomial distribution with more than 2 parameters. For this we employ a powerful optimization strategy called method of lagrange multipliers.

First let $P\left(D|\theta\right)$ denote the probability of a multinomial model M with parameters $\theta = \{\theta_1, \theta_2 \ldots \theta_N\}$ generating a document D.

Which of the following option lists the correct expression for $P\left(D|\theta\right)$. Choose from options below.

- ○ $P\left(D|\theta\right) = \sum_{w\in W} \theta_w^{count(w)}$
- ● $P\left(D|\theta\right) = \Pi_{w\in W} \theta_w^{count(w)}$ ✓
- ○ $P\left(D|\theta\right) = \Pi_{w\in W} count\left(w\right)_w^{\theta}$
- ○ $P\left(D|\theta\right) = \Pi_{w\in W} \theta_w + count\left(w\right)$

Solution:

Recall from the lecture that

$$P\left(D|\theta\right) = \Pi_{i=1}^{n}\theta_{w_i} = \Pi_{w\in W}\theta_w^{count(w)}$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Constraints

1/1 point (graded)

Which of the following options lists the right set of constraints on the parameters θ_w of the multinomial model.

● $\theta_w \geq 0, \sum_{w\in W}\theta_w = 1$ ✓

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

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

○ $\theta_w \geq 0, \sum_{w\in W}\theta_w \geq 1$

Solution:

Note that θ_w denotes the probability of model M choosing the word w. Since it's a probability, its value must lie between 0 and 1. Therefore, $0 \leq \theta_w \leq 1$.

Further, all the above probability values must also sum up to 1. That is, $\sum_{w\in W}\theta_w = 1$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Stationary points for lagrange function

2/2 points (graded)

Let's recall the function that we're trying to optimize:

$$max_\theta P\left(D|\theta\right) = max_\theta \Pi_{w\in W}\theta_w^{count(w)}$$

Let us take natural logarithm on both sides of the equation in order to bring down the exponent

$$max_\theta \log P\left(D|\theta\right) = max_\theta \sum_{w\in W} count\left(w\right)\log\theta_w$$

subject to the following constraints,

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

Let's define a function L called the lagrange function for the sake of the above defined constrained optimization problem:

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

where λ is a constant.

Consider finding stationary points for L by solving for equation obtained by setting its derivative to zero. That is,

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

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

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

○ $\theta_w = \lambda \times count\left(w\right)$

○ $\theta_w = -\lambda \times count\left(w\right)$

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

Now, apply the constraint that $\sum_{w\in W}\theta_w = 1$, we get that λ is:

● $\lambda = -\sum_{w\in W} count\left(w\right)$ ✓

○ $\lambda = \sum_{w\in W} count\left(w\right)$

○ $\lambda = -\sum_{w\in W} count\left(w\right) \times \theta_w$

○ $\lambda = \sum_{w\in W} count\left(w\right) \times \theta_w$

Solution:

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

$$\frac{\partial \log P\left(D|\theta_w\right)}{\partial\theta_w} + \lambda = 0$$

$$\frac{\partial \log \Pi_{w\in W}\theta_w^{count(w)}}{\partial\theta_w} + \lambda = 0$$

$$\frac{\partial \sum_{w\in W}\log\theta_w \times count\left(w\right)}{\partial\theta_w} + \lambda = 0$$

$$\frac{count\left(w\right)}{\theta_w} + \lambda = 0$$

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

If we apply the constraint that $\sum_{w\in W}\theta_w = 1$ we get

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

$$\sum_{w\in W}-\frac{count\left(w\right)}{\lambda} = 1$$

$$\sum_{w\in W}count\left(w\right) = -\lambda$$

$$\lambda = -\sum_{w\in W}count\left(w\right)$$

Substituting this expression for λ back into our previous expression for θ_w we get

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

$$\theta_w = \frac{count\left(w\right)}{\sum_{w\in W}count\left(w\right)}$$

Note that $\theta_w > 0$ and $\sum_{w\in W}\theta_w = 1$ satisfying all the constraints. These set of θ_w parameters are the maximum likelihood estimates for this multinomial generative distribution.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 15. Generative Models / 6. MLEs for general multinomial distribution