





MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Course](#) / [Unit 4. Unsupervised Learning \(2 weeks\)](#) / [Lecture 15. Generative Models](#)[< Previous](#)

6. MLE for Multinomial Distribution

[Bookmark this page](#)

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

Maximum Likelihood Estimate**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.

Let the document D be a sequence of words w_1, \dots, w_n from a collection W consisting of $|W|$ words. We assume that w_i 's are independent, and that the probability of a word w is given by θ_w . We 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 $P(D|\theta)$.

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

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 fun

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

Maximizing $P(D|\theta)$ is equivalent to maximizing $\log P(D|\theta)$, so we take the natural logarithm 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 constraint $\sum_{w \in W} \theta_w = 1$, we use the Lagrange multi

Method of Lagrange Multipliers

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 $\sum_w \theta_w = 1$ to the answer above to obtain λ .

[< Previous](#)

[Next >](#)

$\lambda =$



edX

[About](#)

[Affiliates](#)

[edX for Business](#)

[Open edX](#)

[Careers](#)

[News](#)

Legal

[Terms of Service & Honor Code](#)

[Privacy Policy](#)

[Accessibility Policy](#)

[Trademark Policy](#)

[Sitemap](#)

[Cookie Policy](#)



© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)