You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

2. Maximum Likelihood Estimation

Extension Note: Homework 5 due date has been extended by 1 day to **August 17 23:59UTC** .

Consider a general multinomial distribution with parameters θ. Recall that the likelihood of a dataset $\mathcal{D}$ is given by:

$$P(\mathcal{D};\theta) = \prod_{i=1}^{|\theta|} \theta_i^{c_i}$$

where c_i is the occurrence count of the i-th event.

The MLE of θ is the setting of θ that maximizes $P(\mathcal{D};\theta)$. In lecture we derived this to be

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

Unigram Model

4/4 points (graded)

Consider the sequence:

A B A B B C A B A A B C A C

A unigram model considers just one character at a time and calculates $p(w)$ for $w \in \{A, B, C\}$.

What is the MLE estimate of θ? Give your result to three decimal places.

θ^*_A 0.42857142857142855 ✔ **Answer:** 0.4285714286

θ^*_B 0.35714285714285715 ✔ **Answer:** 0.3571428571

θ^*_C 0.21428571428571427 ✔ **Answer:** 0.2142857143

Using the MLE estimate of θ on $\mathcal{D}$, which of the following sequences is most likely?

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

- (•) ABB ✔
- ( ) AAC

Solution:

We calculate the MLE as $\frac{\text{count}(w)}{N}$ where $N = 14$ and the counts are 6, 5, and 3.

For comparing probabilities in part two, we simply multiply. We only need to compare the numerators: $6 \times 5 \times 3$, 5^3, 6×5^2, and $6^2 \times 3$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Bigram Model 1

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

$$p(\mathcal{D};\theta) = \prod_{w_1,w_2 \in \mathcal{D}} p(w_2|w_1)$$

where w_2 is a word that follows w_1 in the corpus.

This is also a multinomial model. Assume the vocab size is N. How many parameters are there?

Grading note: The formula above contains an error: the probability $p(\mathcal{D};\theta)$ in a bigram model is generally:

$$p(\mathcal{D};\theta) = p(w_0) \prod_{w_1,w_2 \in \mathcal{D}} p(w_2|w_1)$$

where w_0 is the first word, and (w_1, w_2) is a pair of consecutive words in the document. In this case, the number of parameters is $(N-1) + (N^2 - N) = N^2 - 1$. However, with the model as written above, there are only parameters $N^2 - N$.

The grader is now fixed to accept both as correct and regrading is happening.

N^2-1 ✔ **Answer:** N^2 - 1

STANDARD NOTATION

$$p\left(\mathcal{D};\theta\right) = p\left(w_0\right) \prod_{w_1,w_2 \in \mathcal{D}} p\left(w_2|w_1\right)$$

where w_0 is the first word, and $\left(w_1, w_2\right)$ is a pair of consecutive words in the document.
Denote the set of all N words by V. The set of parameters is

$$\left\{p\left(w_0\right) : w_0 \in V\right\} \cup \left\{p\left(w_1|w_2\right) : w_1 \in V, w_2 \in V\right\}$$

and the only constraints on these parameters are

$$\begin{aligned} \sum_{w_0 \in V} p\left(w_0\right) &= 1 \\ \sum_{w_1 \in V} p\left(w_1|w_2\right) &= 1 \quad \text{for all } w_2 \in V. \end{aligned}$$

Hence, the number of parameters is $(N-1) + (N^2 - N) = N^2 - 1$. (Note that this is also the number of parameters $p\left(w_1, w_2\right)$ where $w_1 \in V, w_2 \in V$, which determine the joint distribution.

Solution to the problem as written:
The likelihood of D in bigram model was given as

without taking into account the likelihood $p(w_0)$ of the first word. In this case, the parameters are

$$\{p(w_1|w_2) : w_1 \in V, w_2 \in V\}$$

where $\sum_{w_1 \in V} p(w_1|w_2) = 1$ for all $w_2 \in V$. Hence, the number of parameters is $N^2 - N$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Bigram Model 2

1/1 point (graded)

Which of the following represents the MLE for the **conditional probability** $p(w_2 \mid w_1)$?

- $\frac{\text{count}(w_1, w_2)}{\sum_{w_1', w_2' \in \mathcal{D}} \text{count}(w_1', w_2')}$
- $\frac{\text{count}(w_1, w_2)}{\sum_{w_1', w_2 \in \mathcal{D}} \text{count}(w_1', w_2)}$

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

$$\frac{\text{count}(w_1, w_2)}{\sum_{w_1, w_2' \in \mathcal{D}} \text{count}(w_1, w_2')}$$

Solution:

This is a simple application of Bayes Rule:

$$p(w_2|w_1) = \frac{p(w_1, w_2)}{p(w_1)}$$

To compute $p(w_1)$, we marginalize out w_2.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Bigram Model 3

1/1 point (graded)

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

If you estimate θ on this, what probability will be assigned to the following test sequence? Assume the starting probabilities of all characters $p\left(w|\text{null}\right)$ is uniform. Give your answer to three decimal places.

A A B C B A B

0 ✔

Submit You have used 1 of 3 attempts

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Homework 5 / 2. Maximum Likelihood Estimation

Add a Post

Show all posts ▼ by recent activity ▼

? Bigram Model 1 27

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**

☑	Bigram Model 2. Bayes rule or conditional probability? Hi. Is it a Bayes Rule or simple application of cond. probability? Thanks in advance.	2
💬	Bigram Model I - Clarification for the students	13
☑	[Staff] Bigram Model 1	3
💬	[STAFF] I disagree with Q1 answer	4
?	[Staff] Bigram Model 1: there maybe a Solution smaller than standard p(w_i|w_j) = p(w_i, w_j) / p(w_j) p(D;theta) = prod(p(w_2 | w_1)) = prod(p(w_1, w_2) / p(w_1)) p(w_i) = sum_j(p(w_i, w_j)) So we o...	1
?	[Staff] Please check my answer for Bigram Please check my answer to bigram 3. It says to give answer to 3 decimal places but the grader mark me wrong. thanks	1
💬	Bigram Model 1 - Hints for those trying I will try to give some hints on the Bigram Model 1 question since many seem to have been confused with this problem. In case I am saying too ...	6
☑	[STAFF] Bigram model 1 - not clear at all	20
?	Bigram - what is the assumption about p(null|w) Dear Staff, I see we are considering p(w|null). How about p(null|w)? I do not see it mentioned any where in the bigram questions. Should we add...	5
💬	Yes it is a Markov chain...also, that's not the issue for the number of parameters Community TA	1

You are taking "Final Exam " as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

End My Exam **46:28:56**