

Unit 4 - Week 2

Course outline

How to access the portal

Week 0 : Assignment 0

Week 1

Week 2

- Lecture 6: Spelling Correction: Edit Distance

- Lecture 7: Weighted Edit Distance, Other Variations

- Lecture 8: Noisy Channel Model for Spelling Correction

- Lecture 9: N-Gram Language Models

- Lecture 10: Evaluation of Language Models, Basic Smoothing

- Lecture 11: Tutorial I

- Week 2: Lecture Materials

- Quiz : Assignment 2

- Feedback for Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Week 11

Week 12

DOWNLOAD VIDEOS

Solutions

Assignment 2

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2019-08-21, 23:59 IST.

1) Consider the following corpus C_1 of 4 sentences. What is the total count of unique bi-grams for which the likelihood will be estimated? Assume we do not perform any pre-processing. 1 point

today is Nayan's birthday
she loves ice cream
she is also fond of cream cake
we will celebrate her birthday with ice cream cake

- a. 24
- b. 28
- c. 27
- d. 23

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
a.

2) A 4-gram model is a _____ order Markov Model. 1 point

- a. Constant
- b. Five
- c. Four
- d. Three

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
d.

3) Arrange the words "blueberry, cranberry, raspberry, strawberry" in descending order, based on the frequency of their occurrence in the Google Books n-grams. The Google Books n-gram viewer is available at <https://books.google.com/ngrams>. 1 point

- a. raspberry, strawberry, blueberry, cranberry
- b. blueberry, cranberry, raspberry, strawberry
- c. strawberry, raspberry, cranberry, blueberry
- d. None of the above

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
c.

4) For the string 'mash', identify which of the following set of strings have a Levenshtein distance of 1. 1 point

- a. smash, mas, lash, mushy, hash
- b. bash, stash, lush, flash, dash
- c. smash, mas, lash, mush, ash
- d. None of the above

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
c.

5) Assume that we modify the costs incurred for operations in calculating Levenshtein distance, such that both the insertion and deletion operations incur a cost of 1 each, while substitution incurs a cost of 2. Now, for the string 'lash' which of the following set of strings will have an edit distance of 1? 1 point

- a. ash, slash, clash, flush
- b. flash, stash, lush, blush,
- c. slash, last, bash, ash
- d. None of the above

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
d.

6) Given a corpus C_2 , the Maximum Likelihood Estimation (MLE) for the bigram "dried berries" is 0.3 and the count of occurrence of the word "dried" is 580. for the same corpus C_2 , the likelihood of "dried berries" after applying add-one smoothing is 0.04. What is the vocabulary size of C_2 ? 1 point

- a. 3585
- b. 3795
- c. 4955
- d. 3995

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
b.

7) Calculate $P(\text{they play in a big garden})$ assuming a bi-gram language model. 1 point

- a. 1/8
- b. 1/12
- c. 1/24
- d. None of the above

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
b.

8) Considering the same model as in Question 7, calculate the perplexity of $\langle s \rangle \text{ they play in a big garden } \langle s \rangle$. 1 point

- a. 2.289
- b. 1.426
- c. 1.574
- d. 2.178

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
b.

9) Assume that you are using a bi-gram language model with add one smoothing. Calculate $P(\text{they play in a beautiful garden})$. 1 point

- a. 4.472×10^{-6}
- b. 2.236×10^{-6}
- c. 3.135×10^{-6}
- d. None of the above

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
b.

10) Which of the following sentences will be most likely to occur according to the language model of Question 9? 1 point

- a. $\langle s \rangle$ children play inside
- b. $\langle s \rangle$ in big garden
- c. $\langle s \rangle$ there is big
- d. $\langle s \rangle$ inside a play

- a.
- b.
- c.
- d.

No, the answer is incorrect.
Score: 0

Accepted Answers:
a.