

AI611 μ Word Prediction with N -Grams Model using Python

Quiz 1: N -Grams model

This assessment evaluates the following competencies:

- AI101 – Understand the N -Grams model (+1)
- AI201 – Train an N -Grams model from a given text corpus (+1)
- AI102 – Formally describe N -Grams models thanks to probabilities (+1)

Three affirmations are given for the first assessed competency (AI101). For each of them, you have to decide whether it is true or false. To get a star for the competency, you must have the correct answer for the three affirmations.

| AI101 | True | False |
|---|--------------------------|--------------------------|
| Given the beginning of a sentence, a bigram model chooses the word with the lowest frequency in a given corpus, as the next word to complete the sentence. | <input type="checkbox"/> | <input type="checkbox"/> |
| The bigram model makes the one word length history assumption, that is, the probability of a word w_n given a history w_1^{n-1} is approximated by $P(w_n w_{n-1}^n)$. | <input type="checkbox"/> | <input type="checkbox"/> |
| A bigram model is always faster to train than a trigram model, but is not always as precise when used as a next word predictor. | <input type="checkbox"/> | <input type="checkbox"/> |

For the last assessed competency (AI201 and AI102), you have to train a small bigram model based on a very simple and small corpus consisting of the following two sentences:

- You love dance.
- You dance tonight.

Compute all the bigrams probabilities $P(w_n|w_{n-1})$ (16 probabilities):