Exercise Sheet 9

Neural Computation
University of Birmingham

Week 9 2025

Exercise 1.

The following table shows word counts from a small training corpus: The <EOS> token is used to mark the end of the sequence and should be the last token in any sequence.

Word	**Count**
the	15
cat	4
sat	3
on	6
mat	3
dog	4
chased	2
and	4
ran	2
quickly	2
`<EOS>`	15

1. How can you compute the probability table for a unigram model from the word counts?
2. Compute the probability of the sentence:

   "the cat sat on the mat"

   including the end-of-sequence token.
3. What is the most probable sequence according to the model? What is it's proability?
4. What is the most probable sequence of length 4 (including the <EOS> token)? What is it's probability?

Exercise 2. Now consider the following bigram counts. Rows represent the **previous word** w_{i-1}, columns represent the **next word** w_i. Here the <SOS> token comes at the beginning of all sequences to indicate the start of the sequence.

	the	cat	sat	on	mat	dog	chased	and	ran	quickly	`<EOS>`
`<SOS>`	5	0	0	0	0	0	0	0	0	0	0
the	0	4	0	3	2	3	0	0	0	0	0
cat	0	0	3	0	0	0	0	2	0	0	1
sat	0	0	0	2	0	0	0	0	0	0	1
on	4	0	0	0	1	0	0	0	0	0	1
mat	0	0	0	0	0	0	0	0	0	0	1
dog	0	0	1	0	0	0	2	3	1	0	1
chased	2	0	0	0	0	0	0	0	0	0	1
and	2	0	0	0	0	1	0	0	0	1	0
ran	0	0	0	0	0	0	0	0	0	2	1
quickly	0	0	0	0	0	0	2	1	2	0	1

- Compute the probability of the sentence:

  "the cat sat on the mat"

  including the <SOS> and <EOS> tokens.

- Use the greedy search algorithm for finding high probability sequences. Start with the <SOS> token and continue until you reached a sequence length of 8.