Machine Learning
Exercises: language models (n-grams)

Laura Kallmeyer

Summer 2016, Heinrich-Heine-Universität Düsseldorf

Exercise 1 *Consider the following toy example (similar to the one from Jurafsky & Martin (2015)):*

Training data:

```
<s> I am Sam </s>
<s> Sam I am </s>
<s> Sam I like </s>
<s> Sam I do like </s>
<s> do I like Sam </s>
```

Assume that we use a bigram language model based on the above training data.

1. *What is the most probable next word predicted by the model for the following word sequences?*

   (1) `<s>` *Sam . . .*

   (2) `<s>` *Sam I do . . .*

   (3) `<s>` *Sam I am Sam . . .*

   (4) `<s>` *do I like . . .*

2. *Which of the following sentences is better, i.e., gets a higher probability with this model?*

   (5) `<s>` *Sam I do I like* `</s>`

   (6) `<s>` *Sam I am* `</s>`

   (7) `<s>` *I do like Sam I am* `</s>`

Solution:

Bigram probabilities:

$$\begin{array}{lll}
P(\texttt{Sam}|\texttt{<s>}) = \frac{3}{5} & P(\texttt{I}|\texttt{<s>}) = \frac{1}{5} & \\
P(\texttt{I}|\texttt{Sam}) = \frac{3}{5} & P(\texttt{</s>}|\texttt{Sam}) = \frac{2}{5} & \\
P(\texttt{Sam}|\texttt{am}) = \frac{1}{2} & P(\texttt{</s>}|\texttt{am}) = \frac{1}{2} & \\
P(\texttt{am}|\texttt{I}) = \frac{2}{5} & P(\texttt{like}|\texttt{I}) = \frac{2}{5} & P(\texttt{do}|\texttt{I}) = \frac{1}{5} \\
P(\texttt{Sam}|\texttt{like}) = \frac{1}{3} & P(\texttt{</s>}|\texttt{like}) = \frac{2}{3} & \\
P(\texttt{like}|\texttt{do}) = \frac{1}{2} & P(\texttt{I}|\texttt{do}) = \frac{1}{2} &
\end{array}$$

1. (1) and (3): "I".

   (2): "I" and "like" are equally probable.

   (4): `</s>`

2. Probabilities:

   (5): $\frac{3}{5} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{2}{3}$

   (6): $\frac{3}{5} \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{1}{2}$

   (7): $\frac{1}{5} \cdot \frac{1}{5} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{5} \cdot \frac{2}{5} \cdot \frac{1}{2}$

   (6) is the most probable sentence according to our language model.

Exercise 2 *Consider again the same training data and the same bigram model. Compute the perplexity of*

$$\texttt{<s>}\ \textit{I do like Sam}$$

Solution:

The probability of this sequence is $\frac{1}{5} \cdot \frac{1}{5} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{150}$.

The perplexity is then $\sqrt[4]{150} = 3.5$

Exercise 3 *Take again the same training data. This time, we use a bigram LM with Laplace smoothing.*

1. *Give the following bigram probabilities estimated by this model:*

   $P(\texttt{do}|\texttt{<s>})$ $P(\texttt{do}|\texttt{Sam})$ $P(\texttt{Sam}|\texttt{<s>})$ $P(\texttt{Sam}|\texttt{do})$
   $P(\texttt{I}|\texttt{Sam})$ $P(\texttt{I}|\texttt{do})$ $P(\texttt{like}|\texttt{I})$

   Note that for each word w_{n-1}, we count an additional bigram for each possible continuation w_n. Consequently, we have to take the words into consideration and also the symbol `</s>`.

2. *Calculate the probabilities of the following sequences according to this model:*

   (8) `<s>` *do Sam I like*

   (9) `<s>` *Sam do I like*

   Which of the two sequences is more probable according to our LM?

Solution:

1. If we include `</s>` (this can also appear as second element of a bigram), we get $|V| = 6$ for our vocabulary.

   $P(\texttt{do}|\texttt{<s>}) = \frac{2}{11}$ $P(\texttt{do}|\texttt{Sam}) = \frac{1}{11}$ $P(\texttt{Sam}|\texttt{<s>}) = \frac{4}{11}$ $P(\texttt{Sam}|\texttt{do}) = \frac{1}{8}$
   $P(\texttt{I}|\texttt{Sam}) = \frac{4}{11}$ $P(\texttt{I}|\texttt{do}) = \frac{2}{8}$ $P(\texttt{like}|\texttt{I}) = \frac{3}{11}$

2. (8): $\frac{2}{11} \cdot \frac{1}{8} \cdot \frac{4}{11} \cdot \frac{3}{11}$
   (9): $\frac{4}{11} \cdot \frac{1}{11} \cdot \frac{2}{8} \cdot \frac{3}{11}$
   The two sequences are equally probable.

References

Jurafsky, Daniel & James H. Martin. 2015. Speech and language processing. an introduction to natural language processing, computational linguistics, and speech recognition. Draft of the 3rd edition.