Prof. Dr. Andreas Hotho, Albin Zehe, Jonas Kaiser
Data Science Chair

02.05.2024

1. Assignment in "Machine Learning for Natural Language Processing"

Summer Term 2024

1 General Questions

1. Identify three concrete cases where Deep Learning is used for language processing in real-world applications!
2. Name two reasons why word embeddings are better suited for most NLP tasks than 1-hot encodings!
3. Without covering the details of how the training is actually performed, explain the intuition behind Word2Vec!

2 Bag of Words and Linear Classifiers

Suppose we have the input sentences "I really like cats" and "I mostly like dogs" and want to classify them into the classes "is about cats" and "is about dogs". Let's create a linear classifier for this!

1. Pretend that only the words from the sentences above are in the vocabulary. Create a bag of words representation for each sentence. For a better overview, sort the vocabulary alphabetically.
2. Create a linear classifier that correctly classifies the sentences into the classes "is about cats" and "is about dogs". How many inputs does the classifier need? How many outputs? Assign weights by hand, since optimization was not covered in the lecture yet.
3. Given only the sentences above for selecting the weights of the linear classifier; are there multiple solutions to correctly classify both sentences? Can you construct a valid input sentence that leads to an incorrect classification for an alternative solution? What could we do to prevent such wrong solutions?

3 Python

3.1 Installing and Testing Python and Numpy

- Install Python on your computer. The most recent version can be found at `https://www.python.org/downloads/`.
- Install Numpy on your computer. You can install it using your command line: `pip3 install numpy`. The documentation can be found at `https://numpy.org`.
- Create a Python file, insert the matrices and vectors from the previous task in Python and compute the outputs of the linear classifier to get a feeling for the syntax of Python and Numpy. If you want, you can also implement a function that creates the bag of words representation for a sentence. Feel free to experiment with the language and tools, since you will need them for the upcoming programming tasks.