Homework 9: Deep Learning

Out June 4; Due July 10, 12 a.m.

Kristian Kersting, Dominik Hintersdorf, Quentin Delfosse
{kersting, dominik.hintersdorf, quentin.delfosse}@cs.tu-darmstadt.de

1. Questions

In the exercise session on July 11, we want to discuss remaining questions. To give us some time for preparation, please submit your questions by the end of Sunday, **July 9**.

2. Word2Vec

Furthermore, we want to discuss a model which learns word representations called word2vec [1, 2]. In this homework, we want you to prepare the discussion by finding answers/explanations for questions such as:

- What are the disadvantages of one-hot encodings (for words)? What are good representations (of words)?
- What is a language model?
- How is word2vec related to language models and the distributional hypothesis?
- What are the inputs/outputs of word2vec? Which data is required to train word2vec?
- What does word2vec learn? In what are we interested in the end?
- What is the difference between the CBOW and the Skip-gram model?
- What does negative sampling mean? How is it different from the softmax? How is it related to logistic regression?
- Is there a connection between Autoencoders and word2vec?
- What does the famous example "king - man + woman = queen" mean (in vector space)?
- Is word2vec an unsupervised model?
- Is word2vec a generative model?
- How can we check empirically if word2vec learns something useful?
- How can we use word embeddings to represent sentences or documents? What are the advantages/disadvantages of this representation?

References

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *International Conference on Learning Representations*, 2013.

[2] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.