Second Year Project:
Natural Language Processing and Deep Learning
Week 3

Barbara Plank, Rob van der Goot

February 11, 2020

This is your assignment for the third week of the course *Second Year Project*. You are expected to work on sections 1 and 2 in the Tuesday exercise class, and section 3 in the Friday class.

After completing the whole assignment, you should:

- be able to implement a Naive Bayes classifier from scratch
- be familiar with feedforward neural networks (FFNNs)
- be able to implement a feedforward neural network forward pass from scratch and compare it to an implementation with the Keras toolkit
- be familiar with the concept of word embeddings and have inspected different off-the-shelf word embeddings

Requested reading: Chapters 4 and 6 from *Speech and Language Processing* by Jurafsky and Martin [1].

1 Lyrics Era Classification with Naive Bayes

Complete the exercises provided in the notebook: `lyrics.ipynb`. They contain implementation exercises to create a Naive Bayes model from scratch for lyrics era detection.

2 A Feedforward Neural Network from Scratch

Complete the exercises provided in the notebook: `fnn.ipynb`. In this exercise you will work with a surname nationality predictor and implement a character-based FNN.

3 What are Word Embeddings?

Complete the exercises provided in the notebook: `semantics.ipynb`. In this exercise you will work with word embeddings.

References

[1] Jurfasky and Martin, In Preparation. *Speech and Language Processing (3rd ed. draft)*. Available at https://web.stanford.edu/~jurafsky/slp3/