Exercise 3- Exploring overfitting in NLP

When looking at a number of different types of layer for text classification this week you saw many examples of overfitting -- with one of the major reasons for the overfitting being that your training dataset was quite small, and with a small number of words. Embeddings derived from this may be over generalized also. So for this week's exercise you're going to train on a large dataset, as well as using transfer learning of an existing set of embeddings.

The dataset is from: https://www.kaggle.com/kazanova/sentiment140. I've cleaned it up a little, in particular to make the file encoding work with Python CSV reader.

The embeddings that you will transfer learn from are called the GloVe, also known as Global Vectors for Word Representation, available at: https://nlp.stanford.edu/projects/glove/

This course uses a third-party tool, Exercise 3- Exploring overfitting in NLP, to enhance your learning experience. No personal information will be shared with the tool.

☑ I agree to use this tool responsibly.

⧉ Open Tool