Assignment 4: Convolutional Neural Networks

Homework assignments will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited. Electronic submission on Canvas is mandatory.

1. **Document Classification** (50 points) Implement a Convolutional Neural Network with word embeddings to classify paragraphs into three categories. Use the data in the first assignment.
   (a) Preprocess the train and validation data, build the vocabulary, tokenize, etc.
   (b) Initialize parameters for the model
   (c) Implement the forward pass for the model. Use an embeddings layer as the first layer of your network (i.e. `tf.nn.embedding_lookup`). Set zero paddings to the input matrix.
   (d) Calculate the loss of the model (cross-entropy loss is suggested).
   (e) Set up the training step: use a learning rate of $1e-3$ and an adam optimizer.
   (f) Train you model and report the classification error on validation/test data.
2. **Sentiment Analysis** (50 points)

   This is a standard Rotten Tomatoes dataset with sentiment annotations, deriving from the paper (which you'll need to cite, if you use the dataset): *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, Socher et al., Conference on Empirical Methods in Natural Language Processing (EMNLP, 2013).*

   Use the train, validation, test data split defined in the data. Report your root mean square error on the sentiment label prediction.