DL4NLP - Assignment 2: RNNs and Variants

Total points: 50
Due date: October 18

1 Assignment Goal

In this assignment, you will train and evaluate RNN, LSTM and GRU models, on the task of predicting the sentiment of a short text.

2 Dataset

The dataset you will use to train and test your models is available at: https://nlp.stanford.edu/sentiment/ under 'Dataset Downloads.' Please download the zip file: "Main zip file with readme (6mb)" and use the train/dev/test split made available by the authors. Please read the README file carefully. The labels (positive/negative) are available in Blackboard in the file named "rt-polaritydata.tar.gz". The positive sentences are in the file 'rt-polarity.pos' and the negative sentences are in the file 'rt-polarity.neg.'

This original dataset ("rt-polaritydata.tar.gz") can be downloaded from http://www.cs.cornell.edu/people/pabo/movie-review-data/ under "Sentiment polarity datasets/" ('sentence polarity dataset v1.0').

3 Tasks

1. **Load and preprocess the data**

   We will use word-level representations for your models, specifically, use the glove word embeddings available from https://nlp.stanford.edu/projects/glove/. Please use the "glove.6B.zip" version trained on "Wikipedia 2014 + Gigaword 5" and 300d word vectors.

2. **Implement RNN, LSTM, and GRU models for the sentiment classification task.**

   - **Create the Model/Network:** Define models for vanilla RNN, LSTM and GRU networks, respectively.

- **Train the Model:** Train your models using the sentiment train subset. You are free to choose learning parameters (sequence length, learning rate, batch size, stopping criteria, etc.). Plot the loss and accuracy versus the number of iterations for both train and dev subsets and choose the parameter setting that works best on the dev subset.
- **Experiment with Network Structure:** We want to explore how a network performs when we change its structure. For the sentence encoding, use the element-wise max of all hidden states.
  - Number of hidden units. Try doubling and halving your number of hidden units. Like above, plot the loss and accuracy versus the number of iterations for both train and dev subsets. Discuss your findings.
  - Sequence length. Try doubling and halving your length of sequence that feeds into the network. Plot the loss and accuracy versus the number of iterations for both train and dev subsets. Discuss your findings.
  - Experiment with three word embedding dimensions (50d, 100d, and 300d) using pre-trained and fine-tuned word embeddings. Plot the loss and accuracy versus the number of iterations for both train and dev subsets. Discuss your findings.

3. **Evaluate the models:** Evaluate your best vanilla RNN, LSTM and GRU models on the test data and compare the results of the three models. Use precision, recall, and F1-measure to report the performance on the test set. Offer your intuition, high-level and briefly, behind any observed difference in performance between the models.