Practical 1: Learning Sentence Representations

Steven van de Graaf
12234036
`steven.vandegraaf@student.uva.nl`

All code and output for this assignment are available in this GitHub repository.

While all coding is done and documented, the training is not. I have managed to train the `Baseline` model for the full 20 epochs on the full training set, but, as of yet, have been unable to train the LSTM, `BiLSTM` or the `MaxBiLSMT` models. While I did try to make use of SURFsara's LISA computing cluster to train my models, ever since the afternoon of Monday 22 April, LISA has not run anything for me.

Consequentially, I have few if any results or analysis to report here. As such, I plead you to please take a good and thorough look at my code and accompanying documentation. It took me a lot of painstaking hours to implement and document everything, and I don't want it to be a waste.

Baseline

In Fig. 1, you'll find a plot of the accuracy and loss curves obtained during training of the `Baseline` model.

As you can see, the losses and accuracies are constant over the epochs. This would mean that either my Baseline model did not train properly or I did not record the results correctly.

Furthermore, the accuracy obtained during training on the `DEV` set (as well as the `TRAIN` set) is 0.333. With $n = 3$ classes, this amounts to random guessing, which means that it performs rather poorly on the SNLI task (much more poorly than expected, anyway).

Lastly, the losses are remarkably low, which could explain the constant accuracies and losses, as the model won't learn anything with near-zero losses.

Figure 1: Accuracy and loss curves obtained for the `Baseline` model

Nonetheless, I evaluate this model on 3 transfer tasks from the `SentEval` framework, namely: MR, CR and MPQA, the results of which are (in terms of `DEV` accuracy): MR: 77.67%, CR: 79.84% and MPQA: 87.43%.

As you can see, these transfer task results are quite good. In fact, they're highly comparable to the `GloVe` BOW results listed in Table 4 of Conneau et al. [2017]. This makes sense, as the Baseline model is just that; a BOW model using `GloVe` features.
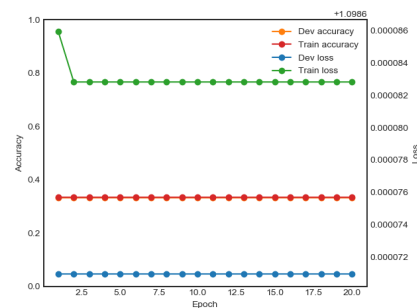
References

Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*, 2017.