Deep Learning Assignment

Eva Vanmassenhove

Deadline: October 5 2020

1 Practical Aspects

The sections below described important practical aspects for the group assignment for the TiU Deep Learning Course, BLOCK 1, 2020.

1.1 Group Assignment

The assignment is to be made in (small) groups consisting of 2–3 people. Groups will **not** be pre-assigned by the lecturer. As such, you will be required to organise yourselves into groups.

1.2 Grading

The hands-on assignment will count for 30% of your total course grade. The grades will be based on the quality of your work as judged by the instructed based on your report and code. Since all students are working from home, their might be hardware limitations in order to train very complex neural networks. You are encouraged to mention improvements to your classifier that you could implement if you had more resources available. The baseline model was trained on a CPU of a standard laptop, as such, the assignment itself does not necessarily require access to GPUs or a very powerful machine. If you do want to experiment with GPUs or TPUs, you could use the facilities provided by Google Colab.

Passing the assignment is not mandatory to pass the course but it is highly advisable. As it is not compulsory, there will be no resit. The exam may include questions that are easier to answer if you have worked actively on the assignment.

2 Task Description

The task consists of a classification task where you are given English sentence and you need to make a classifier that tries to predict whether this sentence is an original English text or a (human) translation from French into English. The data that you have been provided with has been balanced (i.e. there is an equal amount of each class). Similarly, the test data for which you will be asked to provide predictions will be balanced as well.

Warning: This is a difficult classification task – meaning: training a model on this data is not necessarily hard, but getting a model to perform well on it is. As such, the goal is not necessarily for you to get a high accuracy (as this might not even be possible). The baseline system only gets an accuracy of around 58%, as such don't be discouraged if your classifier doesn't reach a high(er) accuracy. This task is about you designing, developing, understanding the working of a deep neural classifier.

2.1 Data set

The data provided consists of transcribed English speech. Some of these sentences were originally uttered in English (label 0), others were originally uttered in French and then translated by a human translator into English (label 1).

More specifically, the *data.zip* file that has been shared with you on Canvas, contains a file called X_data containing the English sentences and y_data containing the labels for every sentence.

A test set (X_test) will also be provided on Canvas (without labels). Explanation on how the accuracy of the model will be tested can be found in Section 3.4.

3 Important Dates and Deliverables

3.1 Report

A one page report should be submitted by October 5, 2020. The report should include the following:

- Title
- Student names and numbers
- The Deep Learning Architecture used by making a diagram depicting your design
- A brief description of your experiments, including possible pre-processing steps, training, hyperparameters, activation function, optimization...
- Results presented in a table or a graph
- Discussion of the performance of your solution and how it relates to the literature used.
- Name of the account under which you submit your results to CodaLab

3.2 Code

Your code should be a plain Python script which can be run to generate your predictions. You do not need to include the training data or the trained model.

3.3 Submission format

Put the report and the code in a single zip file named with your group ID, e.g. group_1.zip and submit it to the Canvas assignment.

3.4 Performance

In addition, you will need to submit a file with the predictions to the competition server. The competition will be hosted on https:/competitions.codalab.org. You will need a Codalab account for the group: indicate the name of this account in your report as well. There will be a separate document on the submission to Codalab with additional details with this respect.