Assignment 5

The deadline for submitting the model is 6th of January 2021

The homework will be presented as following:

・Group A 7 January 2021
・Group B 7 January 2021

Problem:

Implement a classifier, based on a neural network, for the reviews from the IMDB dataset. (https://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz)

You can also load it directly into keras: `from tensorflow.keras.datasets import imdb`

Use the keras framework for training and testing together with theano or tensorflow. You can use any configuration for the network, as long as you respect the following requirements:

- Use any architecture for the network, any optimizer you wish (cost functions, adjustments to solve overfitting, etc.) as long as you are able to explain why did you choose it and how it works・The number of trainable parameters must be less than 250.000
- Do not use any other training data than the current data set (i.e. you can extend your current dataset, but do not use other datasets available on the internet). DO NOT USE the test dataset for training.
- The model has to contain at least one recurrent layer (i.eLSTM, GRU, etc.)

Scoring:

The score for the current homework is 25 points and will be given as follows:

Everyone that has solved the task and is able to explain it will receive 15 points. The rest of the 10 points will be given according to the Gauss curve. (For setting the thresholds only the accuracy for the tasks that were submitted at term will be used)

Every homework where you've used at least one technique that you don't fully understand will be scored with maximum 15 points. (I'm not interested in the mathematical equation, but the reason you've chosen it. I would like an explanation of how the technique works, why the network is working better with the technique you've used and the advantages the technique brings).

The penalty for late presentation is 5 points. (You will first get the points according to the previously computed Gauss curve and then 5 points will be subtracted)

If you do not respect the requirements stated above (i.e. more parameters being used or using another training set) will lead to invalidation of the homework. (the homework will be scored with 0 points)

Evaluation of the model:

For evaluating the neural network, you will submit the trained model, in .h5 format as well as the .py code.

The evaluation will be performed with the following script. Test your model with the script below so you make sure you achieve the same accuracy. Models that can not be loaded and evaluated with the script below will not be taken into account.

I recommend you to use Google Colab. It is free, you can use a GPU and you'll avoid possible serialization conflicts. Use google drive to save your model.

.

```python
from tensorflow.keras.datasets import imdb
from tensorflow.keras.models import load_model
import sys

#numarul maxim de cuvinte pe care vreti sa le considerati
nr_cuv_diferite = sys.argv[1] #ex: 5000
#dimensiunea maxima a unui review
dim_max = sys.argv[2] #ex: 500
if __name__ == '__main__':
# Load the dataset
    (X_train, Y_train), (X_test, Y_test) = imdb.load_data(num_words =
    nr_cuv_diferite)
    model = load_model("model.h5")
    X_test = sequence.pad_sequences(X_test, maxlen=dim_max)
    scores = model.evaluate(X_test, Y_test)
    model.summary()
    print('Loss: %.3f' % scores[0])
    print('Accuracy: %.3f' % scores[1])
```
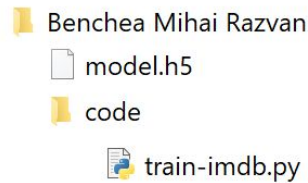
How to submit:

In order to submit the model::

・In one folder that has you name, insert your model named: **model.h5**
・In one sub-folder named **code**, insert the python code used. The python code is necessary for testing.
  Example:

・Make a **zip** archive and submit it to the following google form.
・ **You are allowed only one submission.** In case you submit multiple times, only the first submission will be considered. Test with the above script that the model can be loaded. ・ **The deadline for submitting your homework is 6th of January 2021. After 6th of January, the submission will no longer be possible.**

Use the following url for submitting:
https://forms.gle/RRtMsssnkmxPxuhz7