Mini-project n° 4 – RNN, LSTM and GRU vs MLP and CNN

The objective of this mini-project is to compare an RNN, a LSTM and a GRU, as well as the more traditional MLP and CNN models.

- Train a RNN on the IMDB dataset for the sentiment analysis task.
- Train a LSTM on the IMDB dataset for the sentiment analysis task.
- Train a GRU on the IMDB dataset for the sentiment analysis task.
- Train a MLP on the IMDB dataset for the sentiment analysis task. a CNN using 1-dimensional convolutions on the IMDB dataset for the sentiment analysis task.
- Describe and argue:
  - The choice of the architectures, the layer types, the layer sizes and the activations.
  - The choice of the different hyper-parameters: batch size, learning rate, number of epochs, regularization weight...
  - The choice of the loss function.
  - Compare the time taken by the different models, their memory footprint and their accuracy. What conclusions do you draw?
- Define a classifier by ensembling the 5 architectures above via the technique of the majority vote. Over which of the above models does the ensembling method enhance the results? How can you interpret this?
- BONUS: The IMDB dataset contains aditional unlabeled data. Can you device a way to pre-train the above models on the unlabled data and fine-tune them for the task of sentiment analysis? Are the results better?