HU Extension Assignment 05 S89 DL for NLP

Handed out: 07/25/2019 Due by 11:59 PM EST on Wednesday, 07/31/2019

Problem 1. Please go https://github.com/joosthub/PyTorchNLPBook, the GitHub repository for the book we referenced in Lecture 10, Natural Language Processing with PyTorch by Delip Rao & Brian McMahan, O'Reilly 2019. Our recommendation is download the entire repository. You are asked to make Jupyter notebook `5_3_Document_Classification_with_CNN.ipynb` work. The script `get-all-data.py` will download sample data for all examples in the book. Your notebook requires only AG News examples. You will also need file `glove.6B.100d.txt`, which you already have on your system. Pease that file in the directory `chapter_5/data/glove`. The only slow part of your notebook is the training. Give it a try on your system, if you have GPU card. Please examine the code inside the notebook and first enable GPU usage. If your CUDA software does not crash, you are free to proceed. If you are getting funny errors, you will have to run on Google Colab. That means transferring your AG New and Glove files as well. Do it.
Anyway, train your network and then fetch 15 news titles from recent New York Times. Choose 3 moderately long titles from each of five categories: Sports, Science/Technology, World, Business and Cooking. Tell us how is your network classifying each one of those titles. Your network is trained on Sports, Science/Technology, World, and Business and not Cooking. Where does it places Cooking titles?
(100%)

Now that you have the network working properly, examine its code and keep on learning PyTorch.

Please submit working Jupyter notebook and its PDF image. If your can not generate PDF, please use Chrome> Print and then Save as PDF. Please comment every operation where appropriate. You can submit your report on the predictions made by the network as a Word document is that is more convenient.