

DS-GA 1008: Deep Learning, Spring 2019

Homework Assignment 3

Due: 6pm on Friday, Mar 29, 2019

If you get, give. If you learn, teach. Maya Angelou (1928 - 2014)

1. Fundamentals

1.1. Dropout

Dropout is a very popular regularization technique.

- (a) [2 pts] List the `torch.nn` module corresponding to 2D dropout.
- (b) [8 pts] Read on what dropout is and give a short explanation on what it does and why it is useful. You might find these references helpful:
 - 1. Section 7.12 of the Deep Learning Book by Ian Goodfellow et al:
<https://www.deeplearningbook.org/contents/regularization.html>.
 - 2. Original paper: [link](#)

1.2. Batch Norm

- (a) [2 pts] What does mini-batch refer to in the context of deep learning?
- (b) [8 pts] Read on what batch norm is and give a short explanation on what it does and why it is useful. You might find these references helpful:
 - 1. This blog post: <https://towardsdatascience.com/batch-normalization-in-neural-networks-1ac91516821c>
 - 2. Original paper: [link](#)

2. Language Modeling

This exercise explores the code from the `word_language_model` example in *PyTorch*.

- (a) Go through the code and draw a block diagram / flow chart (see [this tutorial](#)) which highlights the main interacting components, illustrates their functionality, and provides an estimate of their computational time percentage (rough profiling). [10 pts]
- (b) Find and explain where and how the back-propagation through time (BPTT) takes place (you may need to delve into *PyTorch* source code). [5 pts]
- (c) Describe why we need the `repackage_hidden(h)` function, and how it works. [5 pts]

DS-GA 1008: Deep Learning, Spring 2019

Homework Assignment 3

Due: 6pm on Friday, Mar 29, 2019

- (d) Why is there a `--tied` (tie the word embedding and softmax weights) option? [5 pts]
- (e) Compare LSTM and GRU performance (validation perplexity and training time) for different values of the following parameters: number of epochs, number of layers, hidden units size, input embedding dimensionality, BPTT temporal interval, and non-linearities (pick just 3 of these parameters and experiment with 2-3 different values for each). [10 pts]
- (f) Why do we compute performance on a test set as well? What is this number good for? [5 pts]

3. Programming

Complete exercises in `DS-GA-1008-HW_assignment_3.ipynb` [80 pts].

4. Submission

You are required to write up your solutions to Part 1 and Part 2 using \LaTeX .

Submit the following files to **NYU Classes** by the deadline (6pm on Friday, March 29, 2019):

- `First-name_Last-name_netID.A3.pdf` file for Part 1 and Part 2 (containing a read-only link to your Overleaf project for Part 1 and Part 2)
- `First-name_Last-name_netID.A3.tex` file for Part 1 and Part 2
- `First-name_Last-name_netID.A3.ipynb` file for Part 3

5. Disclaimers

You are allowed to discuss problems with other students in the class but have to write up your solutions on your own.

As feedback might be provided during the first days, the current homework assignment might be undergoing some minor changes. We'll notify you if this happens.