

DS-GA 1008: Deep Learning, Spring 2020

Homework Assignment 3

Due: Tuesday, March 31, 2020 at 11:59pm

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

1. Fundamentals

1.1. Convolutions

For the following transformations from dimensionalities A to B , specify any combination of convolutions (with particular choices of kernels, stride), pooling, and other related modules that could get you from dimensionality A to B . [3 pts]

1. $A:(64, 111, 111) \rightarrow B:(4, 234, 234)$
2. $A:(256, 56, 56) \rightarrow B:(14, 126, 126)$
3. $A:(256, 56, 56) \rightarrow B:(42, 72, 72)$

1.2. Dropout

Dropout is a very popular regularization technique.

- (a) List the `torch.nn` module corresponding to 2D dropout. [2 pts]
- (b) Read on what dropout is and give a short explanation on what it does and why it is useful. You may find this chapter <https://www.deeplearningbook.org/contents/regularization.html> and the [original paper](#) helpful. [6 pts]

1.3. Batch Norm

- (a) What does mini-batch refer to in the context of deep learning? [3 pts]
- (b) Read on what batch norm is and give a short explanation on what it does and why it is useful. You might find [this blog post](#) and the [original paper](#) helpful. [6 pts]

DS-GA 1008: Deep Learning, Spring 2020

Homework Assignment 3

Due: Tuesday, March 31, 2020 at 11:59pm

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 on flowcharts in latex](#)) which highlights the main interacting components, illustrates their functionality, and provides an estimate of their computational time percentage (rough profiling). [5 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]
- (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 nonlinearities (pick just 3 of these parameters and experiment with 2-3 different values for each). [5 pts]
- (f) Why do we compute performance on a test set as well? What is this number good for? [5 pts]

DS-GA 1008: Deep Learning, Spring 2020

Homework Assignment 3

Due: Tuesday, March 31, 2020 at 11:59pm

3. PyTorch

If you haven't already, install most recent versions of [Python](#) (3.6 or higher), [PyTorch](#) (we recommend using [conda](#) for the installation), and [Jupyter](#).

Complete the programming exercises provided in the homework 3 directory of the course Google Drive folder ([link](#)).

4. Evaluation

Homework is worth a total of 100 points.

- Part 1 - 50 points
- Part 2 - 50 points

5. Submission

You are required to write up your solutions to Part 1 using markdown or \LaTeX .

Please submit the homework on the NYU classes assignment page. Please upload the following:

- `First-name_Last-name_netID_A3.tex` (or `.md`) file for Part 1
- `First-name_Last-name_netID_A3.pdf` file for Part 1
- `First-name_Last-name_netID_A3_code.ipynb` file for Part 2
- `First-name_Last-name_netID_A3_code.pdf` file for Part 2
(you can use Jupyter Notebook's "File → Download as → PDF" feature)

6. 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.