CIS 620, Advanced Topics in Deep Learning, Spring 2024

Homework 0
Due: Monday, January 22, 11:59pm
Submit to Gradescope

Learning Objectives

After completing this assignment, you will:

- Know what the course's homework policies are
- How to submit written assignments
- How to run a Jupyter Notebook and submit coding assignments
- How to import a model from Huggingface and train it
- Learn about Colin's method on Role-Playing Paper-Reading Seminars

Deliverables

This is an **individual** assignment for both the written and coding portions.

1. **A PDF with your name in the agreement**

    Copy and edit the google doc to enter your name in the Student Agreement and answer the questions mentioned below.

2. **hw0.ipynb file with the functions implemented**

    Complete the coding assignment in the Jupyter Notebook and upload the file to Gradescope.

Note that there is a separate assignment for the papers, which will be handed in separately.

Homework Submission Instructions

Written Homeworks

All written homework must be submitted as a PDF to Gradescope. **Handwritten assignments (scanned or otherwise) will not be accepted.**

Coding Homeworks

All coding assignments will be done in Jupyter Notebooks. We will provide a .ipynb template for each assignment as well as function stubs for you to implement. You are free to use your own installation of Jupyter for the assignments, or Google Colab, which provides a Jupyter Environment connected to Google Drive along with a hosted runtime containing a CPU and GPU.

Assignment Policies

Collaboration Policy

You are allowed and encouraged to work together. Homework will specify the allowed team size. You may discuss the homework with other people to understand the problem and reach a solution. For non-coding homework, it is recommended that each student also write down the solution independently, without referring to written notes from others. I.e., you must understand the solution well enough in order to reconstruct it by yourself. Coding will generally be done in teams of at most two.

On each assignment, you should try to work with a different set of people, to maximize learning, but this is optional, not mandatory.

Homework Related Questions

If you have questions about how to approach a homework problem, you should go to one of the many scheduled instructor/TA office hours, or post your question to Ed. **All questions MUST be directed to Ed. If you send us an email, the only response you will receive is "Please Post to Ed."** If you choose, Ed allows for you to post either anonymously or to the instructors privately. Please refrain from posting solutions or code snippets publicly in Ed. If you want us to review your code, please make the question private.

Homework Due Dates

Any homework turned in late will be penalized 25% per late day – thus a homework four or more days late will receive no credit. Note that there will be instances when we will not accept late homeworks in order to release solutions in a timely fashion. **Your lowest homework score will be dropped.** Thus, if you miss (or are late) a homework due to illness, travel, or any other reason, you can effectively avoid a penalty by dropping that homework.

Plagiarism Policy

When taking phrases or facts from any source (books, articles, or the web), be sure to acknowledge where the information was taken from. Anytext quoted directly must be enclosed in quotation marks, again with the source noted. The format of the citation does not matter, as long as it is sufficiently detailed for us to find the original source.

We will run homeworks through plagiarism detection software, and the graders will be alert for cases of copied homework. **Do not share your code or written solutions with friends to "let them look at it." If they include a piece of your homework in their submission and it is caught, you are** *not* **doing them a favor.**

Suspected violations will be reported to the Office of Student Conduct for investigation and possible disciplinary action. We will not contact you before reporting, but we will let you know that we have done so.

Student Agreement

I have read the above policies and agree to abide by
them. Student Name: **_Vyaas Valsaraj_**

Questions

1. ## Hugging Face

    a. Complete the *hw0.ipynb* file to import the *bert-base-uncased* model and fine
        tune it on the *glue-mrpc* dataset to perform text classification task. Use the
        hyperparameters given in the notebook for this part. You can use the
        Huggingface's course on NLP for reference.

    b. We want to see the effects of batch size on the model training. For this rerun the
        model with 5 widely different mini-batch sizes (e.g. 1, 10,100). How does
        compute time (for a fixed training set size) change with minibatch size? How (if
        at all) does test accuracy change with mini-batch size?

    c. We used AdamW optimizer here, is it the same as Adam?

2. ## Colin's Roles

    a. Read the blog on Role-Playing Paper-Reading Seminars by Alec Jacobson and
        Colin Raffel. This will be important for paper discussions in the coming weeks.

    b. What role do you think you would most enjoy? why?

ANSWERS:-

1.b) Initially, I ran the same steps for the mini batch sizes- 1,10,50,75 and 100 because I was
getting out of memory errors. I noticed that the compute time for training took the longest for a
mini batch of size 1 and kept decreasing as the batch size increased. The test accuracy was
lowest for a batch of size 1 but it kept increasing with increasing batch size. However, for the
final batch size of 100, it decreased a little and was actually much lower than the test accuracy
for batch size of 75. As I reran the same code block, I noticed that the batch size of 50 had the
highest test accuracy while the sizes 75 and 100 had a slightly lower and similar accuracy,
which may mean the batch size after a certain point may not really matter that much with
respect to accuracy.

I managed to run the above process for batch sizes-[1,10,50,100,150] and noticed a similar
trend. The maximum I could go was 150 because I am getting out of memory errors for 200 and
above inspite of clearing cache and deleting the existing model. I ran the training process with a

batch size of 200 alone and still faced memory issues. Transformers are really memory intensive as they run parallel computations and we do need superior computing power to train such large batch sizes.

1.c) No, AdamW is different from Adam wherein AdamW makes a tweak in the way the original Adam optimizer handles weight decay. Instead of applying the weight decay directly within the optimizer's update rule, it separates the weight decay term, basically decoupling the weight decay.

2. As a person with a stutter, I would enjoy playing any role that involves the least amount of speaking. I would guess the investigator would be able to make it through without talking so much but a key factor of stuttering which every single stuttering person experiences irrespective of the spectrum of stuttering they fall under is that they always use synonyms of words to avoid stuttering on a word or a phrase. The alternative word or phrase may be so much more complicated than the original word they intended to say but for some reason, it would come out of our mouths way easier without any obstruction. Having this in mind, I would avoid the investigator's role as it involves hard facts about a person where I would not be able to rely on synonyms. So, I would really consider the researcher's role because I feel that has the most scope of being as general as possible where I will be able to mix in a lot of synonyms in case I stutter and there would be a good flow in my speech but ultimately, I feel I would still be better off as an investigator because I would not be nervous about technical information and would not be thinking too much. I would also not have to face any questions asked by my classmates and can avoid being put on the spot.