Assignment 3: Tutorial

GRAD-E1394

November 2, 2023

In groups, students will create a tutorial on a specific application of deep learning in the policy context. Tutorials are intended to provide the other students in class with an example of working code that illustrates in a pedagogical way how to go about implementing a deep learning approach. Groups can choose from a set of given topics. Rather than in a research project, the aim here is to lead classmates through a well-defined problem and provide them with a self-contained notebook that they can use in further projects that may involve a similar approach.

Any changes that will be made to these instructions will be announced through Moodle.

1 Instructions

1.1 Setting and task

Imagine you work in a data science unit in a ministry or public organization. There is a lot of workforce turnaround in your unit, and it helps to have working code on different areas of deep learning that can help new people get started. Your unit is therefore creating short tutorials on different topics relevant for the ministry to get new people up to speed. Those tutorials also come in handy when your unit is exploring new deep learning projects, and when you need to explain to the leadership why your work is relevant. Every tutorial therefore contains a short memo on the top explaining in layman's terms why this topic is relevant for public policy. Discuss relevant research works, real-world examples of successful applications, and/or organizations and governments that apply such approaches for policy making.

Your job is to create such a tutorial on a specific topic. The task is to create:

1. a notebook for colleagues with an example application, dataset, and all relevant instructions, including a short memo on why the topic is relevant for public policy or government, and a video explaining the tutorial

2. a presentation to the whole team about the topic and how to use the notebook.

1.2 Grading

Distribution of the overall grade is as follows,

1. Notebook: 55% (includes 10% for the memo)
2. Video: 15 %
3. Presentation: 30%.

To ensure that all members of the group participate actively in the development of the notebook, it is required for all members to make at least some meaningful commits on Github. If we see a group member making no or very minor commits, we reserve the right to apply a penalty to this group member's grade. Note that the share of commits does not have to be equally distributed among group members and that unless we see violations of the basic requirement described above, we aim to apply the same grade to all group members.

1.3 Tutorial structure

You find a template for the tutorial here: `https://colab.research.google.com/drive/17nmmxY0OciYuNNDZ640bfH62JnFuEMvn?usp=sharing`. An example of a tutorial from a previous iteration of the course will also be made available on Moodle.

While in principle you are free to create your tutorial as you see fit, we recommend to stay closely aligned with the sections in the template. At a minimum, every tutorial should contain the following components:

- Memo describing the relevance
- Introduction and task description
- Dataset download
- Data prepocessing
- Methodology
- Results and discussion
- References

The memo can be short, and should be informative but on point. The memo should be up to 300 words (approximately). In addition, we ask you to produce a video of maximum 10 minutes explaining the components of the tutorial to the user. Ideally, you will execute/show the tutorial and explain verbally what is being done in each part of the tutorial. There is no need to present the topic overall again as you have done in your presentation.

2 Developing the idea

2.1 Topics

Tutorial topics cover aspects that were missing in this course or were not addressed in the problem sets. Where applicable, we provide example readings but there is no need to involve that reading in the development of your tutorial. You may suggest your own topic but we recommend to follow the list below.

Core topics:

- Interpretability/explainability: For example, implementing a technique to interpret deep learning models. Starting points: Paper by Amarasinghe et al., paper by Rudin, book by Molnar.
- Sequence modeling for time series data: Implementing an approach introduced in Session 6.
- Bias and fairness: Implementing different metrics to measure fairness and approaches to reduce bias using an example.
- Generative adversarial networks (GANs): Implement an example of a GAN. Starting points: Panel discussion 'GANs for Good'
- Transfer learning and generalization. Starting point: Paper by Rolf et al.
- Energy-efficient deep learning: Show which carbon footprint measurement tools are available and how to use them, implement an energy-efficient model. Starting points: Paper by Kaack et al., Paper by Luccioni et al.
- Topics relating to optimization algorithms and compute
- Few-shot learning
- Semi- or self-supervised learning

More difficult topics that mostly only exist as cutting-edge research:

- Uncertainty quantification: Implementing a model and quantify its uncertainty as it can be useful for decision-makers. Starting point: Sun et al.
- Physics-informed models. Starting points: Paper by Beucler et al., Google flood forecasting
- Adversarial machine learning: What to do when things break because a malicious actor is fooling your model.
- Meta-learning
- Graph neural networks

2.2 Tips

- Set the tutorial up in pedagogical way. This is *not* a research project.
- You may follow a paper closely but you of course have to write your own code.
- Limit the computational time by working with smaller datasets and choosing your experiments wisely. The goal is not to achieve top performance, the goal is to provide your classmates with a starting point how to implement a particular model and learn about a new topic. Your notebook should run on Colab to be more useful for the future.
- **Follow plagiarism guidelines.** Do not copy or closely follow a tutorial from the internet. While we have been rather generous when we found out that content was adapted from online sources in the problem sets, we will not be generous here. Indicate all references used. If you leaned on other material in any way, make it abundantly clear which parts are adapted and how you modified them with your own work. Refer to the Hertie School Code of Conduct.

3 Submission

3.1 Proposal – Nov 16, 23:59

Form groups of 4 students (one group will have 5). On Moodle, submit a document that contains the following information:

1. Tentative tutorial title
2. Names of all group members. Also enter these on Moodle.
3. Description of **preferred topic**: Up to 300 words detailing what your tutorial will focus on, including resources you will use to develop it.
4. Description of **alternative topic**: Up to 200 words detailing what your tutorial will focus on in case you are not able to work on your preferred topic, including resources you will use to develop it.
5. Optional comments, such as why your group wants to focus on that topic.

This submission will not be graded. We use it to assign you the topic, in order to prevent having too many tutorials on the same topic.

3.2 Presentation – Session 12 on Dec 7, 16:00

Submit the slides on Moodle before the session (latest at 10:00). Present the project in class. The presentation should be 10 minutes, followed by 5 minutes of Q&A. In your presentation, you should pitch why the topic is relevant for

public policy and government, and then provide the background about the task and model that is covered in your tutorial. Finish by providing a very brief teaser of your tutorial (even if your tutorial is not finalized yet).

3.3 Tutorial submission – Dec 14, 23:59

Submit the tutorial on Github `https://classroom.github.com/a/wnYEo2Lv`. The final submission should contain 1) the notebook and readme files 2) a max 10-min video walking a user through the notebook (or a link to the video if the video is too large).

Unless otherwise agreed upon, please submit a notebook that runs on Colab with a sufficiently small dataset to keep the required computational time low.