MiniProject 4: Reproducible Machine Learning

COMP 551, McGill University
Professor: William L. Hamilton

November 20th, 2019

Please read this entire document before beginning the assignment.

Preamble

- This mini-project is **due on December 3rd at 11:59pm. Late submissions will be accepted without penalty until December 14th at 11:59pm, but no submissions will be accepted after that date.**

- This mini-project is to completed in groups of three. All members of a group will receive the same grade. It is not expected that all team members will contribute equally to all components. However every team member should make integral contributions to the project.

- You will submit your assignment on MyCourses as a group. As with previous mini-projects, you must register your group on MyCourses and any group member can submit.

- You are free to use any programming language or toolbox you want with this project.

Background

Reproducibility (i.e., the ability for researchers to replicate others' work) is a critical issue in machine learning, and there is a growing movement towards ensuring reproducible machine learning. In this mini-project, you will aim to reproduce the results of a published machine learning paper. There are two different tracks: the first track aims to reproduce and extend on results from papers that did not publish code, and the second track aims to use published code to reproduce and extend on an existing work. You must either:

1. Choose a paper from the following curated list: `goo.gl/aVLcaY`.

2. Already be registered to participate in the NeurIPS Reproducibility Challenge—in which case you will reproduce the paper you registered for.

3. Receive special permission from the instructor to reproduce a different paper. If you choose a paper yourself, it should be published in well-known peer-reviewed machine learning venue (e.g., ICML, AAAI, NeurIPS, ACL, CVPR, or ICLR) and have been published no earlier than 2015.

You must form a team and choose a paper by November 27th at 11:59pm. One member of your team must fill in the following form by that date: https://forms.gle/vPiqmT27qM2U5SME7.

Disclaimer 1: Not all papers are easily reproducibile. Failing to reproduce the exact results of the original paper is not necessarily a bad thing, as long as your experiments are rigorous!

Disclaimer 2: Fully reproducing some papers or baselines will require more computational resources than you have access to. You should keep this in mind when you choose a paper, but it is also okay to only work on a subset of the data/tasks in a paper, in order to make things tractable.

Track 1: Implement and improve the baselines

When researchers propose new models, they often spend more time on fine-tuning their proposed model than fine-tuning the baselines. Simple baselines like Naive Bayes, Random Forests, k-NN if properly tuned, can sometimes provide performances that are hard to beat. In this track, your goal will be to select a paper (e.g., from the list of suggested papers) and improve/analyze the baselines for the tasks considered in the paper. First implement (a subset of) the baseline(s) mentioned in the paper (without using the author's code) and try to reproduce the baseline performance reported in the paper. Then try to fine-tune the baselines by doing extensive hyper-parameter tuning. Then, explore simple machine learning algorithms (e.g., discussed in the class) to improve the performance on the given task. Note that for some recent papers a simple/standard LSTM or CNN would count as a baseline. Your job is to act like an adversary to the paper and try to beat their performance by using simple algorithms that have less computational complexity. **At a minimum, in this track you should reproduce at least one baseline proposed in the paper and implement one modified baseline of your own. An outstanding project in this track would perform a rigorous and extensive evaluation many simple baselines and/or propose a clever new baseline that achieves strong performance.**

Track 2: Reproduce and modify the model

The goal of this track is to take a published machine learning model (e.g., from the given list of papers) **that comes with published code** and try to explore the proposed model in depth. Specifically, you will first reproduce the results reported in the paper by running the code provided by the authors. Then you will try to modify the model and perform ablation studies to understand the model's robustness and evaluate the importance of the various model components. (In this context, the term "ablation" is used to describe the process of removing different model components to see how it impacts performance.) You can also try to improve the model based on your experiments. You should do a thorough analysis of the model through an extensive set of experiments. Note that some experiments will be difficult to replicate due to computational resources. It is fine to reproduce only a subset of the original paper's results or to work on a smaller variant of the data—if necessary. **At a minimum, in this track you should use the authors code to reproduce a non-trivial subset of their results and explore how the model performs after you make minor modifications (e.g., changes to hyperparameters). An outstanding project in this track would perform very a detailed ablation study and/or implement significant/meaningful extensions of the model.**

Deliverables

You must submit two separate files to MyCourses (**using the exact filenames and file types outlined below**):

1. **code.zip**: A collection of supporting code files. Please submit a README detailing the packages you used and providing instructions to replicate your results.

2. **writeup.pdf**: Your project write-up as a pdf (details below).

In addition, you must form a team and choose a paper by November 27th at 11:59pm. One member of your team must fill in the following form by that date: `https://forms.gle/vPiqmT27qM2U5SME7`.

Project write-up

Your team must submit a project write-up that is a maximum of eight pages (single-spaced, 10pt font or larger; extra pages for references/bibliographical content and appendices can be used). We highly recommend that students use LaTeX to complete their write-ups and use the bibtex feature for citations. **You are free to structure the report how you see fit, but you should clearly summarize/describe the paper/task you are reproducing, reference relevant related works, describe the baselines/modifications you implemented, and clearly present your results.**

You must include a statement of contributions at the end of your report, which specifies what each group member contributed.

Clarifications for the NeurIPS Reproducibility Challenge

To be eligible for the bonus 10 points, you must:

- Already be registered for the NeurIPS Reproduciblity Challenge.
- Submit a project report to MyCourses by the class deadline on December 14th.
- Also submit a report to the challenge website by December 27th. This can be the same report you submitted to the class, or you can improve this report based on the feedback you receive.

Evaluation

The mini-project is out of 100 points. **Teams that participate in the NeurIPS Reproducibility Challenge will be graded out of 110 points, with the extra 10 points being a potential bonus.** Your grade will be based on your submitted write-up, as well as the supporting code. As with the previous mini-projects your write-up will be judged according its scientific quality (included but not limited to):

- Is your proposed methodology technically sound? For example, in Track 1, are your baselines appropriate for the task and properly implemented? For Track 2, are the modifications you made reasonable and properly implemented?
- How detailed/rigorous/extensive are your experiments?
- Does your report clearly describe the task you are working on (i.e., the paper you are reproducing), the experimental set-up, results, figures (e.g., don't forget axis labels and captions on figures, don't forget to explain figures in the text).
- Is your report well-organized and coherent?
- Is your report clear and free of grammatical errors and typos?
- Does your report include an adequate discussion of related work and citations?