

Reproducibility

Simone Collier

University of Toronto

What is Reproducibility?

Reproducible work should be **repeatable by someone else using the same data and analytical tools** and the analysis should yield the same results.

- Scientific papers are meant to include any information that is necessary to repeat the analysis and arrive at the same conclusions.
- This format is not very suitable for complex methods that require a lot of detail to be reproducible.
- The code and data are required for reproducible work and these are rarely included in scientific papers.

Why is Reproducibility important?

For the scientific community:

- Reproducibility of results ensures the results are well founded.
- Falsifying a irreproducible study is very difficult.

For the non-scientific community:

- Non-academic machine learning practitioners will often use methods from existing studies and apply them to new problems.
- Reproducing the existing study is the first step to learning the method since it acts as sample code.
- As a result, ensuring reproducibility of work will increase the impact of the work.

Reproducibility challenges in ML

There are a number of reasons for why a scientific paper may be irreproducible.

- Training data is not available.
- Lack of details for the model or training procedure.
- The code to run the experiment is not available or contains errors.
- Improper use of statistics when analyzing the results.
- Overfitting.
- Selective reporting of the results.
- Drawing conclusions that are not supported by the results.

Many of these issues are often exacerbated by the push to publish positive results.

Levels of Reproducibility

There are many people who have made an effort to assign a reproducibility score to ML scientific papers based on their content. The guidelines presented in the paper by Tatman, et al. (2018) include three levels:

- Low reproducibility: finished paper only.
- Medium reproducibility: code and data.
- High reproducibility: code, data, and environment.

Low Reproducibility

A machine learning paper has low reproducibility if it is said to have "from scratch" reproducibility.

- It contains the methodology in just enough detail that another expert in the field should be able to reproduce it.
- In reality reproducing it is likely infeasible due to time or missing information.
- Releasing papers with this level of reproducibility is not reasonable given the ease of sharing code and data online.

Medium Reproducibility

A machine learning paper has medium reproducibility if it includes the data and code.

- Although there are many ethical considerations to sharing certain types of data, there are usually ways to anonymise it.
- Someone should be able to process the data, train the model, and evaluate the model results.
- Problems could still occur as a results of running the reproduction on a different machine.

High Reproducibility

A machine learning paper has medium reproducibility if it includes the data, code and environment used to run the code.

- The environment includes all the libraries and dependencies that are necessary to run the code on a new machine.
- This will allow someone to easily reproduce the results on a different machine than the original.

Reproducibility in Module 6

Although we have discussed reproducibility for machine learning in the context of scientific papers the same principles apply outside of the academic community.

Throughout this module we have practiced reproducibility in our work by:

- Ensuring a statistical method is appropriate for the problem at hand.
- Outlining the model fitting procedure including how the training and test sets were generated.
- Testing for overfitting.
- Computing test statistics where appropriate.
- Explaining results.
- Providing code and libraries used for the analysis.

Pineau, J. et al. 2021. Improving Reproducibility in Machine Learning Research. Journal of Machine Learning Research 22: 1–20.

Tatman, R., J. VanderPlas, and S. Dane. 2018. A Practical Taxonomy of Reproducibility for Machine Learning. 2nd Reproducibility in Machine Learning Workshop at ICML. Stockholm, Sweden.