VU VRIJE UNIVERSITEIT AMSTERDAM

Nik hef

Machine Learning:
a New Toolbox for Theoretical Physics
D-ITP Advanced Topics in Theoretical Physics

Study Guide

current version: October 16, 2019

Course coordinator: Dr Juan Rojo (j.rojo@vu.nl)

http://www.juanrojo.com

1   Course description

Under the generic umbrella denomination of Machine Learning (or Artificial Intelligence) one finds a number of statistical algorithms with the common feature that they manage to solve a specific task by learning by example from large datasets, as opposed to following a set of predefined rules. In recent years, Machine Learning tools have demonstrated their impressive potential to tackle a range of challenging problems related to classification, optimization, discrimination, inter- and extrapolation, efficient parameterizations, and pattern generalization, among several others.

In this course, we present an overview of the applications of Machine Learning to various areas of theoretical physics, from particle and astroparticle physics phenomenology to condensed matter and string theory to name a few. The course will be a mixture of theoretical foundations, discussion of practical applications, and hands-on tutorials.

2   Schedule

There will be four lectures, each taking place between 11am and 1pm at **room H331 of Nikhef**. In the corresponding afternoon the hands-on tutorial will take place, always in room H331. So the schedule of the course is:

- Monday 18th of November 2019: Lecture (11am to 1pm) and Tutorial (2pm to 5pm).

- Monday 25th of November 2019: Lecture (11am to 1pm) and Tutorial (2pm to 5pm).

- Monday 2nd of December 2019: Lecture (11am to 1pm) and Tutorial (2pm to 5pm).

- Monday 9th of December 2019: Lecture (11am to 1pm) and Tutorial (2pm to 5pm).

The exam will take place on Monday 16th of December. The requisites to obtain a pass in this module of the Advanced Topics in Theoretical Physics course are described below. Note that only registered students that aim to receive credits upon completion of the three course modules need to take part in the examination.

3   Examination

The examination of the course will be composed by two parts:

- A short report (4 pages long) about a specific application of Machine Learning algorithms to a physics problem that you find interesting or relevant for your research. This can be one of the examples discussed in the course of an altogether new application.

  This report can be either written individually or in groups of at most three students. You should describe the physics problem that one is trying to use, the details of the

machine learning algorithm adopted, and how this approach improves over more traditional strategies.

Enthusiastic students might want to include with their submission a piece of code that implements and executes this ML algorithm for some toy application, though this is not a formal requirement to achieve a pass in the course.

This report should be submitted to the course coordinator by email **by Friday 13th of December** at the latest.

- On Monday 16th of December starting at 11am each student/group should present their report to the course coordinator and the rest of participants. Students should aim for a short (10 min max) presentation followed by a discussion. Providing full background of the relevant physics is not necessary (further references if needed can be added in the written report).

Note that both handing in the reports and presenting their content on the 16th of December are formal requisites (necessary but not sufficient) to pass the course. If the quality of the report and the corresponding presentation is deemed sufficient (taking into account the topics covered during the course), the student/group will be granted a pass for this module.

As mentioned above, **only registered students that take the course for credits** are required to participate in the examination.

4   Course contents

Some of the contents that will be covered during the course include:

- Artificial neural networks

- Supervised versus unsupervised learning

- Hyper-parameter optimisation

- Training strategies: from genetic algorithms to gradient descent methods

- Generative models and generative adversarial networks

- Deep learning

- Under-learning and over-fitting

- Machine learning application for parametrization

- Machine learning applications for classification and discrimination

- Machine learning applications for inter- and extrapolation

- Machine learning applications for complex optimisation and search problems

- Convolutional neural networks

- Variational autoencoders.

Further topics can be added to the above list if there is sufficient interest from the participating students.

5   Hands-on tutorials

As mentioned above, this course will include four hands-on tutorial sessions following each of the lectures. The idea of these tutorial sessions is for the participants to learn how to use, modify, and apply some of the machine learning algorithms presented in the course. These will be implemented in `Python` programs and/or notebooks. While the emphasis of the course is not on programming itself but rather in the algorithms and their applications, fully benefiting from the tutorials requires some knowledge of `Python`.

Participants should **bring their own laptops** to the tutorial sessions, and make sure that they have an up-to-date functional installation of `Python` in their system. The code that will be used for the tutorials, together with all other course materials such as the lecture slides, can be obtained from a public GitHub repository:

<div align="center">

https://github.com/juanrojochacon/ml-ditp-attp

</div>

This repository will be updated during the course and students should remind themselves to pull the latest version before the lectures.

Many of the examples that we will use require specific `Python` libraries such as `SciPy`, `NumPy`, and `MatplotLib`. Participants in the course should make sure that their local `Python` installation is up to date and that all relevant packages are installed. This can be verified by executing some of the example programs that will be found in the `Code` folder in the repository.

There exist a large number of possible ways to install Python. A convenient one is via the `Conda` package manager

<div align="center">

https://docs.conda.io/projects/conda/en/latest/user-guide/install/

</div>

which is available for Windows, macOS and Linux, but other package managers such as `Homebrew` or `Pip` can also be used. Regardless of the choice, an important recommendation is to use a single package manager, since combining two or more usually leads to issues and conflicts with the `Python` installation.

6   Useful reading

Here we provide an incomplete list of potentially useful references about the topic of machine learning in physics in general, in some cases with emphasis in theoretical and high energy physics. The interested student will find in these references sufficient material if she is interested in deepening some of the concepts that will be covered in the lectures.

References

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborová, "Machine learning and the physical sciences," arXiv:1903.10563 [physics.comp-ph].

[2] P. Mehta, M. Bukov, C. H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher and D. J. Schwab, "A high-bias, low-variance introduction to Machine Learning for physicists," Phys. Rept. **810**, 1 (2019) doi:10.1016/j.physrep.2019.03.001 [arXiv:1803.08823 [physics.comp-ph]].

[3] M. Abdughani, J. Ren, L. Wu, J. M. Yang and J. Zhao, "Supervised deep learning in high energy phenomenology: a mini review," Commun. Theor. Phys. **71**, no. 8, 955 (2019) doi:10.1088/0253-6102/71/8/955 [arXiv:1905.06047 [hep-ph]].

[4] M. Paganini, "Machine Learning Solutions for High Energy Physics: Applications to Electromagnetic Shower Generation, Flavor Tagging, and the Search for di-Higgs Production," arXiv:1903.05082 [hep-ex].

[5] Giles Louppe, lectures on Advanced Machine Learning: https://github.com/glouppe/info8004-advanced-machine-learning

[6] Machine Learning in Condensed Matter Physics course. Roger Melko, Lei Wang, Eun-Ah Kim, Zohar Ringel. Course repository: https://github.com/iamc/ML-CM-2019