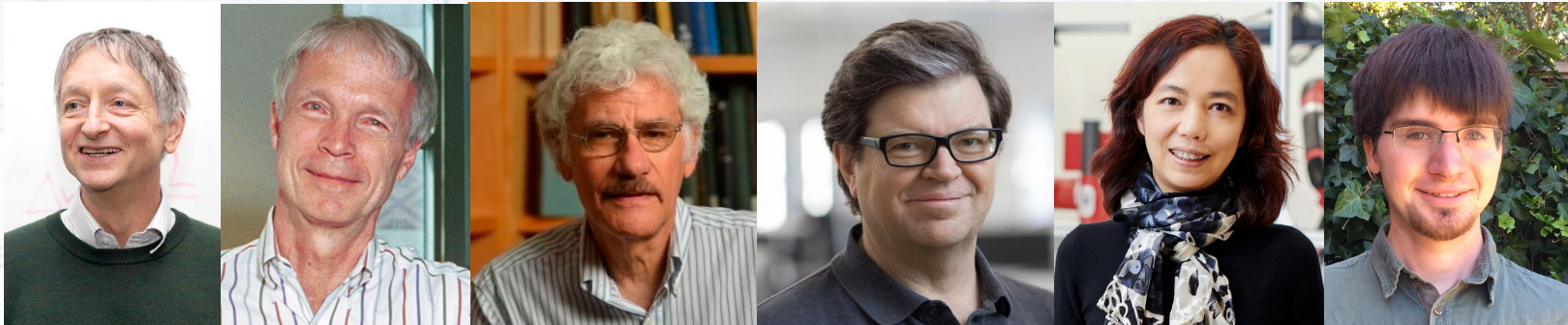
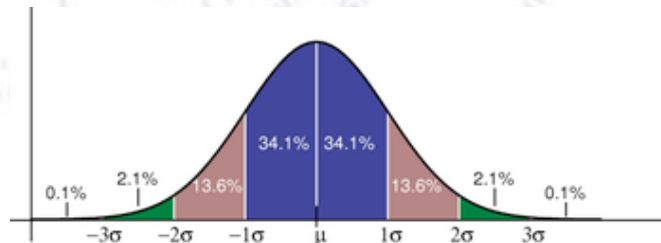


AIPHY PhD School

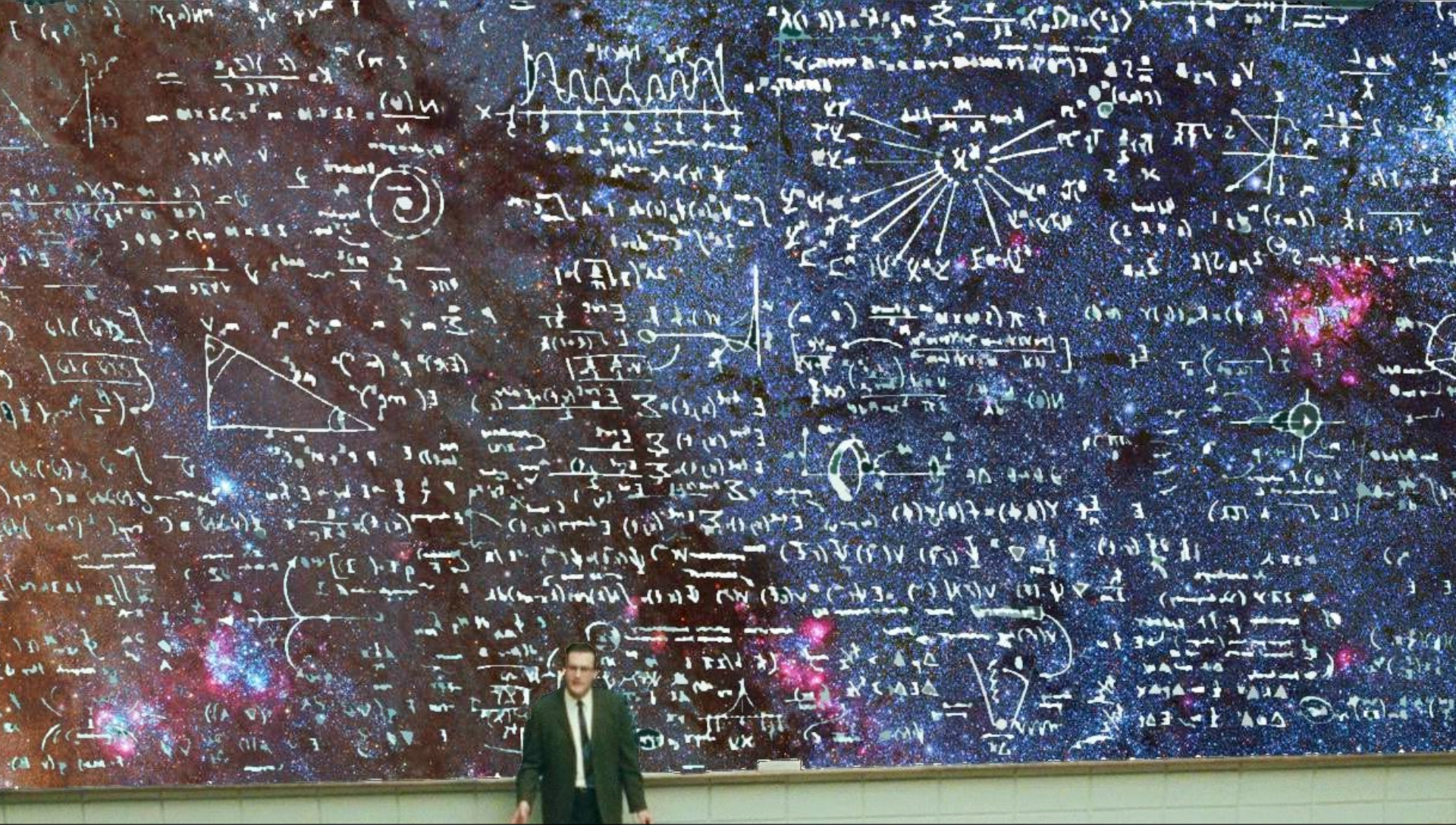
Exercise information



Troels C. Petersen (NBI)



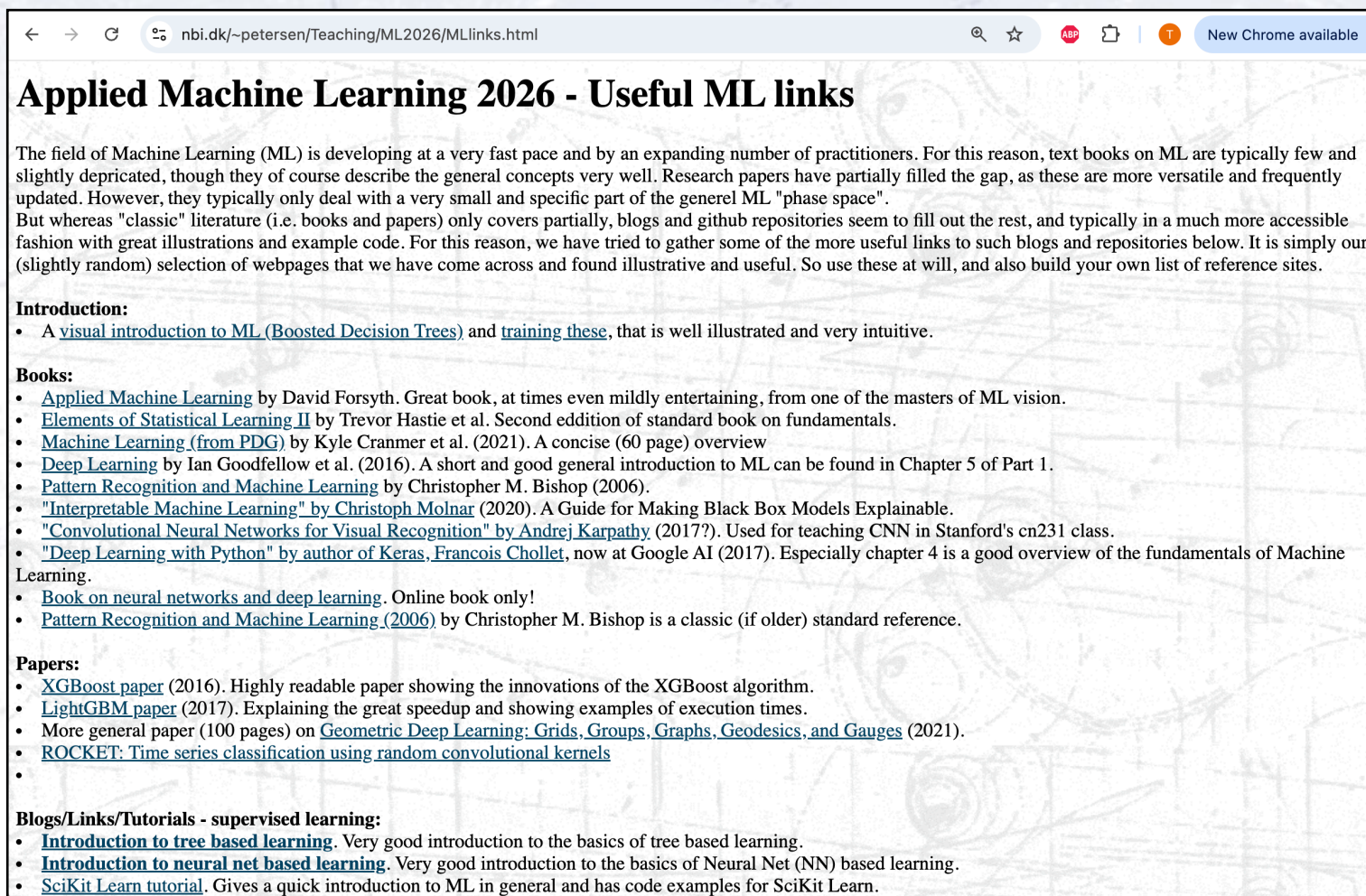
"Statistics is merely a quantisation of common sense - Machine Learning is a sharpening of it!"



You can not teach a person anything!
You can only help them discovering it in themselves...
[Galileo Galilei]

WebBlogs as literature

In ML, blogs/articles/tutorials are a very common (and great) source of literature on ML. For this reason, we've made a list of links that we find good:



Applied Machine Learning 2026 - Useful ML links

The field of Machine Learning (ML) is developing at a very fast pace and by an expanding number of practitioners. For this reason, text books on ML are typically few and slightly deprecated, though they of course describe the general concepts very well. Research papers have partially filled the gap, as these are more versatile and frequently updated. However, they typically only deal with a very small and specific part of the general ML "phase space". But whereas "classic" literature (i.e. books and papers) only covers partially, blogs and github repositories seem to fill out the rest, and typically in a much more accessible fashion with great illustrations and example code. For this reason, we have tried to gather some of the more useful links to such blogs and repositories below. It is simply our (slightly random) selection of webpages that we have come across and found illustrative and useful. So use these at will, and also build your own list of reference sites.

Introduction:

- A [visual introduction to ML \(Boosted Decision Trees\)](#) and [training these](#), that is well illustrated and very intuitive.

Books:

- [Applied Machine Learning](#) by David Forsyth. Great book, at times even mildly entertaining, from one of the masters of ML vision.
- [Elements of Statistical Learning II](#) by Trevor Hastie et al. Second edition of standard book on fundamentals.
- [Machine Learning \(from PDG\)](#) by Kyle Cranmer et al. (2021). A concise (60 page) overview
- [Deep Learning](#) by Ian Goodfellow et al. (2016). A short and good general introduction to ML can be found in Chapter 5 of Part 1.
- [Pattern Recognition and Machine Learning](#) by Christopher M. Bishop (2006).
- ["Interpretable Machine Learning" by Christoph Molnar](#) (2020). A Guide for Making Black Box Models Explainable.
- ["Convolutional Neural Networks for Visual Recognition" by Andrej Karpathy](#) (2017?). Used for teaching CNN in Stanford's cn231 class.
- ["Deep Learning with Python" by author of Keras, Francois Chollet](#), now at Google AI (2017). Especially chapter 4 is a good overview of the fundamentals of Machine Learning.
- [Book on neural networks and deep learning](#). Online book only!
- [Pattern Recognition and Machine Learning \(2006\)](#) by Christopher M. Bishop is a classic (if older) standard reference.

Papers:

- [XGBoost paper](#) (2016). Highly readable paper showing the innovations of the XGBoost algorithm.
- [LightGBM paper](#) (2017). Explaining the great speedup and showing examples of execution times.
- More general paper (100 pages) on [Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges](#) (2021).
- [ROCKET: Time series classification using random convolutional kernels](#)
-

Blogs/Links/Tutorials - supervised learning:

- [Introduction to tree based learning](#). Very good introduction to the basics of tree based learning.
- [Introduction to neural net based learning](#). Very good introduction to the basics of Neural Net (NN) based learning.
- [SciKit Learn tutorial](#). Gives a quick introduction to ML in general and has code examples for SciKit Learn.

Code of Conduct

This Ph.D. school is a community event intended for presentations, networking and collaboration. We value a civil and respectful environment which encourages the free expression and exchange of scientific ideas.

All attendees are expected to adhere to the [CERN Code of Conduct](#).

Should a lapse of professional decorum occur, attendees are encouraged to bring issues, in a confidential setting, to the advisors appointed by the conference organisers.

The advisors will suggest ways of redressing the matter and counsel the parties involved. The conference organisers may, after due consideration, take action as they deem appropriate, including, in severe cases, expulsion from the school.

If you believe someone is violating the code of conduct, we ask that you report it by emailing Anja Butter (anja.butter@lpnhe.in2p3.fr) or Carsten Littek (littek@thphys.uni-heidelberg.de).

Exercises

For this Ph.D. school we have prepared some exercises, which are pieces of code that describe an illustrate a particular ML method. There are five such exercises:

1. Classification, Regression, and Clustering Kaggle:

https://www.nbi.dk/~petersen/PhDsSchool2026/PhDsSchool2026_InitialProject.html

Solutions can be submitted here (we will attempt a daily leaderboard!):

<https://docs.google.com/forms/d/1vvhfiVb8f96uUkVqrtO6Mg5paUwzpRhF7Dv-hJm-IDI/edit>

2. VAE on MNIST data
3. CNN on MNIST data
4. GNN on HIV molecule data
5. Transformer on Shakespearean data

The last four exercises can be found on GitHub:

https://github.com/troelspetersen/PhDsSchool2026_AppML

Choose exercises at will, play around, and most importantly... **discuss these with your peers.**

Essential Links

ML links:

<https://www.nbi.dk/~petersen/Teaching/ML2026/MLlinks.html>

GitHub for school exercises:

https://github.com/troelspetersen/PhDschoo2026_AppML

Link to “Kaggle Challenge” setup/rules:

https://www.nbi.dk/~petersen/PhDschoo2026/PhDschoo2026_InitialProject.html



ML is a big field - so don't worry if you don't understand everything!