Project 1 CNNs, RNNs, Autoencoders and more

FYS5429/9429, Advanced machine learning and data analysis for the physical sciences, University of Oslo, Norway

Spring semester 2025, deadline March 21

Discriminative methods

This project explores different discriminative methods, starting with standard neural networks, then moving to Comvolutional NNs (CNNs) or Recurrent NNs (RNNs), Autoencoders and/or Graphical NNs (GNNs). It allows also for a smooth transition to generative methods like Boltzmann machines, VAEs, GANs and Diffusion models. These methods are the topic of the second project.

Images and more. A data set which many of you are familiar with is the MNIST set of handwritten numbers, see https://en.wikipedia.org/wiki/MNIST_database. Here you can study this data set with NNs, CNNs and eventually autoencoders or GNNs. For project 2 you could continue with the same data set and explore generative methods. You can obviously replace this data set with other images, such as images from modeling of phase transitions in physical systems, see https://www.arxiv.org/abs/2501.05547 or other data sets of your choice.

Time series. Another frequently occuring type of data are so-called time series. Here again you can start with NNs, move over to RNNs and autoencoders. Similarly, for project 2, you can include VAEs, Diffusion models or other generative methods. At the end you will have studied a broad set of deep learning methods, allowing you thereby to assess their pros and cons.

Feel free to develop own codes or use libraries like TensorFlow and/or PyTorch.

Introduction to numerical projects

Here follows a brief recipe and recommendation on how to write a report for each project.

- Give a short description of the nature of the problem and the eventual numerical methods you have used.
- Describe the algorithm you have used and/or developed. Here you may find it convenient to use pseudocoding. In many cases you can describe the algorithm in the program itself.
- Include the source code of your program. Comment your program properly.
- If possible, try to find analytic solutions, or known limits in order to test your program when developing the code.
- Include your results either in figure form or in a table. Remember to label your results. All tables and figures should have relevant captions and labels on the axes.
- Try to evaluate the reliabilty and numerical stability/precision of your results. If possible, include a qualitative and/or quantitative discussion of the numerical stability, eventual loss of precision etc.
- Try to give an interpretation of you results in your answers to the problems.
- Critique: if possible include your comments and reflections about the exercise, whether you felt you learnt something, ideas for improvements and other thoughts you've made when solving the exercise. We wish to keep this course at the interactive level and your comments can help us improve it.
- Try to establish a practice where you log your work at the computerlab. You may find such a logbook very handy at later stages in your work, especially when you don't properly remember what a previous test version of your program did. Here you could also record the time spent on solving the exercise, various algorithms you may have tested or other topics which you feel worthy of mentioning.

Format for electronic delivery of report and programs

The preferred format for the report is a PDF file. You can also use DOC or postscript formats or as an ipython notebook file. As programming language we prefer that you choose between C/C++, Fortran2008 or Python. The following prescription should be followed when preparing the report:

- Send us an email in order to hand in your projects with a link to your GitHub/Gitlab repository.
- In your GitHub/GitLab or similar repository, please include a folder which contains selected results. These can be in the form of output from your code for a selected set of runs and input parameters.

Finally, we encourage you to collaborate. Optimal working groups consist of 2-3 students. You can then hand in a common report.