Machine Learning/artificial intelligence and quantum technologies at the university of Oslo; research, education and exciting possibilities in the future job market

Morten Hjorth-Jensen[1]

Department of Physics and Center for Computing in Science Education,
University of Oslo, Norway[1]

Aug 15, 2024

What is this talk about?

The main emphasis is to give you a short and pedestrian introduction to the whys and hows we can use (with several examples) machine learning methods and quantum technologies in science and perhaps in general. And why this could (or should) be of interest. I will also try to link to potential job possibilities and educational activities.

AI/ML and some statements you may have heard (and what do they mean?)

1. Fei-Fei Li on ImageNet: **map out the entire world of objects** (The data that transformed AI research)
2. Russell and Norvig in their popular textbook: **relevant to any intellectual task; it is truly a universal field** (Artificial Intelligence, A modern approach)
3. Woody Bledsoe puts it more bluntly: **in the long run, AI is the only science** (quoted in Pamilla McCorduck, Machines who think)

If you wish to have a critical read on AI/ML from a societal point of view, see Kate Crawford's recent text Atlas of AI. And Inga Strumke, Machines which think

Here: with AI/ML we intend a collection of machine learning methods with an emphasis on statistical learning and data analysis

Types of machine learning

The approaches to machine learning are many, but are often split into two main categories. In *supervised learning* we know the answer to a problem, and let the computer deduce the logic behind it. On the other hand, *unsupervised learning* is a method for finding patterns and relationship in data sets without any prior knowledge of the system.

An important third category is *reinforcement learning*. This is a paradigm of learning inspired by behavioural psychology, where learning is achieved by trial-and-error, solely from rewards and punishment.

Main categories

Another way to categorize machine learning tasks is to consider the desired output of a system. Some of the most common tasks are:

- ▶ Classification: Outputs are divided into two or more classes. The goal is to produce a model that assigns inputs into one of these classes. An example is to identify digits based on pictures of hand-written ones. Classification is typically supervised learning.

- ▶ Regression: Finding a functional relationship between an input data set and a reference data set. The goal is to construct a function that maps input data to continuous output values.

- ▶ Clustering: Data are divided into groups with certain common traits, without knowing the different groups beforehand. It is thus a form of unsupervised learning.

The plethora of machine learning algorithms/methods

1. Deep learning: Neural Networks (NN), Convolutional NN, Recurrent NN, Boltzmann machines, autoencoders and variational autoencoders and generative adversarial networks, stable diffusion and many more generative models

2. Bayesian statistics and Bayesian Machine Learning, Bayesian experimental design, Bayesian Regression models, Bayesian neural networks, Gaussian processes and much more

3. Dimensionality reduction (Principal component analysis), Clustering Methods and more

4. Ensemble Methods, Random forests, bagging and voting methods, gradient boosting approaches

5. Linear and logistic regression, Kernel methods, support vector machines and more

6. Reinforcement Learning; Transfer Learning and more

Example of discriminative modeling, taken from Generative Deep Learning by David Foster
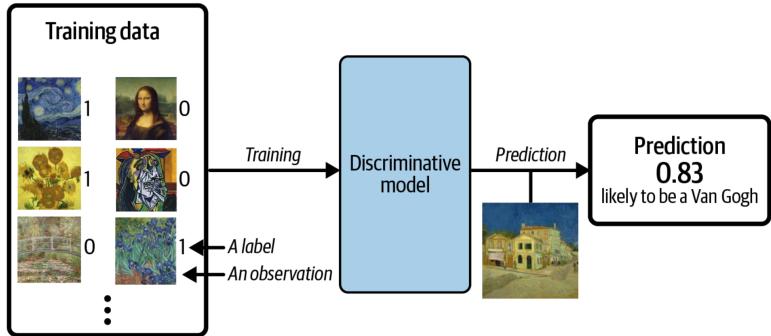
Discriminative Modeling

When performing discriminative modeling, each observation in the training data has a label. For a binary classification problem such as our data could be labeled as ones and zeros. Our model then learns how to discriminate between these two groups and outputs the probability that a new observation has label 1 or 0

In contrast, generative modeling doesn't require the dataset to be labeled because it concerns itself with generating entirely new data (for example an image), rather than trying to predict a label for say a given image.

What Is Generative Modeling?

Generative modeling can be broadly defined as follows:
Generative modeling is a branch of machine learning that involves
training a model to produce new data that is similar to a given
dataset.

What does this mean in practice? Suppose we have a dataset
containing photos of horses. We can train a generative model on
this dataset to capture the rules that govern the complex
relationships between pixels in images of horses. Then we can
sample from this model to create novel, realistic images of horses
that did not exist in the original dataset.

Example of generative modeling, taken from Generative Deep Learning by David Foster
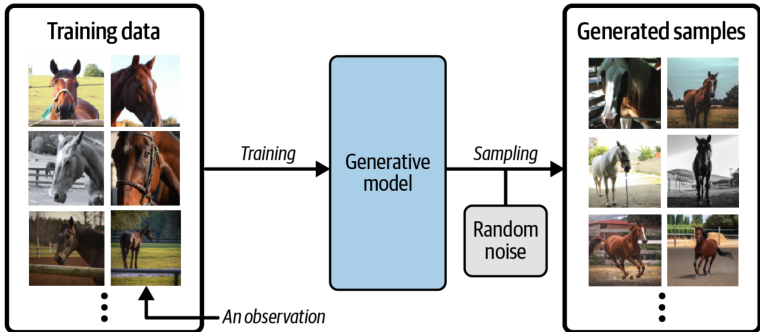
Good books with hands-on material and codes

- Sebastian Rashcka et al, Machine learning with Sickit-Learn and PyTorch
- David Foster, Generative Deep Learning with TensorFlow
- Bali and Gavras, Generative AI with Python and TensorFlow 2

All three books have GitHub addresses from where one can download all codes. We will borrow most of the material from these three texts as well as from Goodfellow, Bengio and Courville's text Deep Learning

What are the basic Machine Learning ingredients?

Almost every problem in ML and data science starts with the same ingredients:

- ▶ The dataset $\boldsymbol{x}$ (could be some observable quantity of the system we are studying)
- ▶ A model which is a function of a set of parameters $\boldsymbol{\alpha}$ that relates to the dataset, say a likelihood function $p(\boldsymbol{x}|\boldsymbol{\alpha})$ or just a simple model $f(\boldsymbol{\alpha})$
- ▶ A so-called **loss/cost/risk** function $\mathcal{C}(\boldsymbol{x}, f(\boldsymbol{\alpha}))$ which allows us to decide how well our model represents the dataset.

We seek to minimize the function $\mathcal{C}(\boldsymbol{x}, f(\boldsymbol{\alpha}))$ by finding the parameter values which minimize $\mathcal{C}$. This leads to various minimization algorithms. It may surprise many, but at the heart of all machine learning algortihms there is an optimization problem.

Machine learning. A simple perspective on the interface between ML and Physics
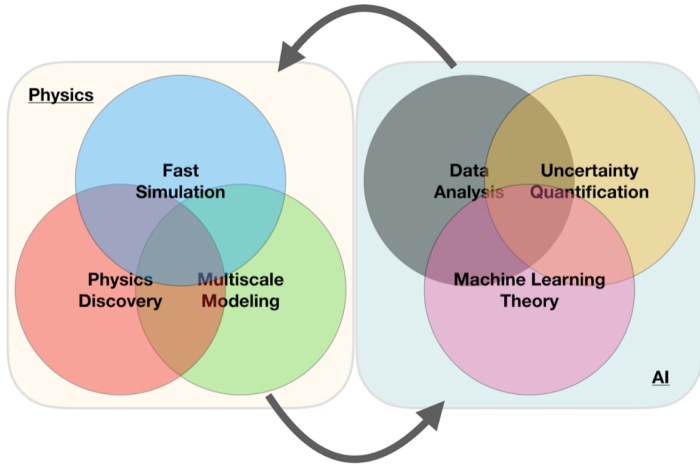
High performance computing demands (and immense power consumption)

In June 2023, just a few months after GPT-4 was released, the developers publicly explained that GPT-4 was comprised of roughly 1.8 trillion (trillion=1000 billions= one million millions) parameters. More specifically, the architecture consisted of eight models, with each internal model made up of 220 billion parameters.

Observations

- Need for AI/Machine Learning in science, lots of ongoing activities
- To solve many complex problems and facilitate discoveries, multidisciplinary efforts efforts are required involving scientists in physics, statistics, computational science, applied math and many many other fields.
- There is a need for focused AI/ML learning for specific disciplines

Possible start to raise awareness about ML in your own field

- Make an ML challenge in your own field a la Learning to discover: the Higgs boson machine learning challenge. Alternatively go to kaggle.com at
  `https://www.kaggle.com/c/higgs-boson`
- HEP@CERN and HEP in general have made significant impacts in the field of machine learning and AI. Something to learn from

Important questions

1. How do we incorporate these topics in our education?
2. More difficult: what are the consequences for universities and our educational mission?

Education

1. Incorporate elements of statistical data analysis and Machine Learning in undergraduate programs
2. Develop courses on Machine Learning and statistical data analysis
3. Build up a series of courses in Quantum Information Technologies (QIT)
4. Modifying contents of present Physics programs or new programs on Computational Physics and Quantum Technologies
   1. study direction/option in **quantum technologies**
   2. study direction/option in **Artificial Intelligence and Machine Learning**
   3. and more
5. Master of Science/PhD programs in Computational and Data Science
   1. UiO has already MSc programs in CS and DS
   2. Many other universities are developing or have similar programs

Important Issues to think of

1. Lots of conceptual learning: superposition, entanglement, QIT applications, etc.

2. Coding is indispensable.

3. Teamwork, project management, and communication are important and highly valued

4. Engagement with industry: guest lectures, virtual tours, co-ops, and/or internships.

Observations

1. People do not really know what QIT is.
2. ML/AI seen as black boxes/magic!

Future Needs/Problems

1. There are already great needs for specialized people (Ph. D. s, postdocs), but also needs of people with a broad overview of what is possible in ML/AI and/or QIT.
2. There are not enough potential employees in AI/ML and QIT . It is a supply gap, not a skills gap.
3. A BSc with specialization is a good place to start
4. It is tremendously important to get everyone speaking the same language. Facility with the vernacular of quantum mechanics is a big plus.