Summary of course

Morten Hjorth-Jensen Email morten.hjorth-jensen@fys.uio.no[1,2]

[1]Department of Physics and Center of Mathematics for Applications, University of Oslo
[2]National Superconducting Cyclotron Laboratory, Michigan State University

Nov 28, 2019

What? Me worry? No final exam in this course!

What did I learn in school this year?

Our ideal about knowledge on computational science
    Does that match the experiences you have made this semester?

Topics we have covered this year

The course has two central parts

1. Statistical analysis and optimization of data

2. Machine learning

Statistical analysis and optimization of data

The following topics will be covered

1. Basic concepts, expectation values, variance, covariance, correlation functions and errors;

2. Simpler models, binomial distribution, the Poisson distribution, simple and multivariate normal distributions;

3. Central elements from linear algebra

4. Gradient methods for data optimization

5. Estimation of errors using cross-validation, blocking, bootstrapping and jackknife methods;

6. Practical optimization using Singular-value decomposition and least squares for parameterizing data.

7. Principal Component Analysis.

Machine learning

The following topics will be covered

1. Linear methods for regression and classification;

2. Neural networks;

3. Decisions trees, random forests, boosting and bagging

4. Support vector machines

Learning outcomes and overarching aims of this course

The course introduces a variety of central algorithms and methods essential for studies of data analysis and machine learning. The course is project based and through the various projects, normally three, you will be exposed to fundamental research problems in these fields, with the aim to reproduce state of the art scientific results. The students will learn to develop and structure large codes for studying these systems, get acquainted with computing facilities and learn to handle large scientific projects. A good scientific and ethical conduct is emphasized throughout the course.

- Understand linear methods for regression and classification;

- Learn about neural network;

- Learn about baggin, boosting and trees

- Support vector machines

- Learn about basic data analysis;

- Be capable of extending the acquired knowledge to other systems and cases;

- Have an understanding of central algorithms used in data analysis and machine learning;

- Work on numerical projects to illustrate the theory. The projects play a central role and you are expected to know modern programming languages like Python or C++.

Perspective on Machine Learning

1. Rapidly emerging application area

2. Experiment AND theory are evolving in many many fields. Still many low-hanging fruits.

3. Requires education/retraining for more widespread adoption

4. A lot of "word-of-mouth" development methods

Huge amounts of data sets require automation, classical analysis tools often inadequate. High energy physics hit this wall in the 90's. In 2009 single top quark production was determined via Boosted decision trees, Bayesian Neural Networks, etc.

Machine Learning Research

Where to find recent results:

1. Conference proceedings, arXiv and blog posts!

2. **NIPS**: Neural Information Processing Systems

3. **ICLR**: International Conference on Learning Representations

4. **ICML**: International Conference on Machine Learning

5. Journal of Machine Learning Research

Starting your Machine Learning Project

1. Identify problem type: classification, generation, regression

2. Consider your data carefully

3. Choose a simple model that fits 1. and 2.

4. Consider your data carefully again... data representation

5. Based on results, feedback loop to earliest possible point

Choose a Model and Algorithm

1. Supervised?

2. Start with the simplest model that fits your problem

3. Start with minimal processing of data

Preparing Your Data

1. Shuffle your data

2. Mean center your data

   - Why?

3. Normalize the variance

   - Why?

4. **Whitening**

   - Decorrelates data
   - Can be hit or miss

5. When to do train/test split?

Which Activation and Weights to Choose in Neural Networks

1. RELU? ELU?

2. Sigmoid or Tanh?

3. Set all weights to 0?

   - Terrible idea

4. Set all weights to random values?

   - Small random values

Optimization Methods and Hyperparameters

1. Stochastic gradient descent

   (a) Stochastic gradient descent + momentum

2. State-of-the-art approaches:

   - RMSProp
   - Adam

Which regularization and hyperparameters? L_1 or L_2, soft classifiers, depths of trees and many other. Need to explore a large set of hyperparameters and regularization methods.

Resampling

When do we resample?

1. Bootstrap

2. Cross-validation

3. Jackknife and many other

Other courses on Data science and Machine Learning at UiO

The link here https://www.mn.uio.no/english/research/about/centre-focus/innovation/data-science/studies/ gives an excellent overview of courses on Machine learning at UiO.

1. STK2100 Machine learning and statistical methods for prediction and classification.

2. IN3050/IN4050 Introduction to Artificial Intelligence and Machine Learning. Introductory course in machine learning and AI with an algorithmic approach.

3. STK-INF3000/4000 Selected Topics in Data Science. The course provides insight into selected contemporary relevant topics within Data Science.

4. IN4080 Natural Language Processing. Probabilistic and machine learning techniques applied to natural language processing.

5. STK-IN4300 – Statistical learning methods in Data Science. An advanced introduction to statistical and machine learning. For students with a good mathematics and statistics background.

6. IN-STK5000 Adaptive Methods for Data-Based Decision Making. Methods for adaptive collection and processing of data based on machine learning techniques.

7. IN5400/INF5860 – Machine Learning for Image Analysis. An introduction to deep learning with particular emphasis on applications within Image analysis, but useful for other application areas too.

8. TEK5040 – Dyp læring for autonome systemer. The course addresses advanced algorithms and architectures for deep learning with neural networks. The course provides an introduction to how deep-learning techniques can be used in the construction of key parts of advanced autonomous systems that exist in physical environments and cyber environments.

Additional courses of interest

1. STK4051 Computational Statistics

2. STK4021 Applied Bayesian Analysis and Numerical Methods

What's the future like?

Based on multi-layer nonlinear neural networks, deep learning can learn directly from raw data, automatically extract and abstract features from layer to layer, and then achieve the goal of regression, classification, or ranking. Deep learning has made breakthroughs in computer vision, speech processing and natural language, and reached or even surpassed human level. The success of deep learning is mainly due to the three factors: big data, big model, and big computing.

In the past few decades, many different architectures of deep neural networks have been proposed, such as

1. Convolutional neural networks, which are mostly used in image and video data processing, and have also been applied to sequential data such as text processing;

2. Recurrent neural networks, which can process sequential data of variable length and have been widely used in natural language understanding and speech processing;

3. Encoder-decoder framework, which is mostly used for image or sequence generation, such as machine translation, text summarization, and image captioning.

Bayesian Machine Learning

This is an important topic if we aim at extracting a probability distribution. This gives us also a confidence interval and error estimates.

Bayesian machine learning allows us to encode our prior beliefs about what those models should look like, independent of what the data tells us. This is especially useful when we don't have a ton of data to confidently learn our model.

Reinforcement Learning

Reinforcement learning is a sub-area of machine learning. It studies how agents take actions based on trial and error, so as to maximize some notion of cumulative reward in a dynamic system or environment. Due to its generality, the problem has also been studied in many other disciplines, such as game theory, control theory, operations research, information theory, multi-agent systems, swarm intelligence, statistics, and genetic algorithms.

In March 2016, AlphaGo, a computer program that plays the board game Go, beat Lee Sedol in a five-game match. This was the first time a computer Go program had beaten a 9-dan (highest rank) professional without handicaps. AlphaGo is based on deep convolutional neural networks and reinforcement learning. AlphaGo's victory was a major milestone in artificial intelligence and it has also made reinforcement learning a hot research area in the field of machine learning.

Transfer learning

The goal of transfer learning is to transfer the model or knowledge obtained from a source task to the target task, in order to resolve the issues of insufficient training data in the target task. The rationality of doing so lies in that usually the source and target tasks have inter-correlations, and therefore either the features, samples, or models in the source task might provide useful information for us to better solve the target task. Transfer learning is a hot research topic in recent years, with many problems still waiting to be solved in this space.

Adversarial learning

The conventional deep generative model has a potential problem: the model tends to generate extreme instances to maximize the probabilistic likelihood, which will hurt its performance. Adversarial learning utilizes the adversarial behaviors (e.g., generating adversarial instances or training an adversarial model) to enhance the robustness of the model and improve the quality of the generated data. In recent years, one of the most promising unsupervised learning technologies, generative adversarial networks (GAN), has already been successfully applied to image, speech, and text.

Dual learning

Dual learning is a new learning paradigm, the basic idea of which is to use the primal-dual structure between machine learning tasks to obtain effective feedback/regularization, and guide and strengthen the learning process, thus reducing the requirement of large-scale labeled data for deep learning. The idea of dual learning has been applied to many problems in machine learning, including machine translation, image style conversion, question answering and generation, image classification and generation, text classification and generation, image-to-text, and text-to-image.

Distributed machine learning

Distributed computation will speed up machine learning algorithms, significantly improve their efficiency, and thus enlarge their application. When distributed meets machine learning, more than just implementing the machine learning algorithms in parallel is required.

Meta learning

Meta learning is an emerging research direction in machine learning. Roughly speaking, meta learning concerns learning how to learn, and focuses on the understanding and adaptation of the learning itself, instead of just completing a specific learning task. That is, a meta learner needs to be able to evaluate its own learning methods and adjust its own learning methods according to specific learning tasks.

The Challenges Facing Machine Learning

While there has been much progress in machine learning, there are also challenges.

For example, the mainstream machine learning technologies are black-box approaches, making us concerned about their potential risks. To tackle this challenge, we may want to make machine learning more explainable and controllable. As another example, the computational complexity of machine learning algorithms is usually very high and we may want to invent lightweight algorithms or implementations. Furthermore, in many domains such as physics, chemistry, biology, and social sciences, people usually seek elegantly simple equations (e.g., the Schrödinger equation) to uncover the underlying laws behind various phenomena. In the field of machine learning, can we reveal simple laws instead of designing more complex models for data fitting? Although there are many challenges, we are still very optimistic about the future of machine learning. As we look forward to the future, here are what we think the research hotspots in the next ten years will be.

Explainable machine learning

Machine learning, especially deep learning, evolves rapidly. The ability gap between machine and human on many complex cognitive tasks becomes narrower and narrower. However, we are still in the very early stage in terms of explaining why those effective models work and how they work.

What is missing: the gap between correlation and causation Most machine learning techniques, especially the statistical ones, depend highly on data correlation to make predictions and analyses. In contrast, rational humans tend to reply on clear and trustworthy causality relations obtained via logical reasoning on real and clear facts. It is one of the core goals of explainable machine learning to transition from solving problems by data correlation to solving problems by logical reasoning.

Quantum machine learning

Quantum machine learning is an emerging interdisciplinary research area at the intersection of quantum computing and machine learning.

Quantum computers use effects such as quantum coherence and quantum entanglement to process information, which is fundamentally different from classical computers. Quantum algorithms have surpassed the best classical algorithms in several problems (e.g., searching for an unsorted database, inverting a sparse matrix), which we call quantum acceleration.

When quantum computing meets machine learning, it can be a mutually beneficial and reinforcing process, as it allows us to take advantage of quantum computing to improve the performance of classical machine learning algorithms. In addition, we can also use the machine learning algorithms (on classic computers) to analyze and improve quantum computing systems.

Quantum machine learning algorithms based on linear algebra

Many quantum machine learning algorithms are based on variants of quantum algorithms for solving linear equations, which can efficiently solve N-variable linear equations with complexity of O(log2 N) under certain conditions. The quantum matrix inversion algorithm can accelerate many machine learning methods, such as least square linear regression, least square version of support vector machine, Gaussian process, and more. The training of these algorithms can be simplified to solve linear equations. The key bottleneck of this type of quantum machine learning algorithms is data input—that is, how to initialize the quantum system with the entire data set. Although efficient data-input algorithms exist for certain situations, how to efficiently input data into a quantum system is as yet unknown for most cases.

Quantum reinforcement learning

In quantum reinforcement learning, a quantum agent interacts with the classical environment to obtain rewards from the environment, so as to adjust and improve its behavioral strategies. In some cases, it achieves quantum acceleration by the quantum processing capabilities of the agent or the possibility of exploring the environment through quantum superposition. Such algorithms have been proposed in superconducting circuits and systems of trapped ions.

Quantum deep learning

Dedicated quantum information processors, such as quantum annealers and programmable photonic circuits, are well suited for building deep quantum networks. The simplest deep quantum network is the Boltzmann machine. The classical Boltzmann machine consists of bits with tunable interactions and is trained by adjusting the interaction of these bits so that the distribution of its expression conforms to the statistics of the data. To quantize the Boltzmann machine, the neural network can simply be represented as a set of interacting quantum spins that correspond to an adjustable Ising model. Then, by initializing the input neurons in the Boltzmann machine to a fixed state and allowing the system to heat up, we can read out the output qubits to get the result.

Social machine learning

Machine learning aims to imitate how humans learn. While we have developed successful machine learning algorithms, until now we have ignored one important fact: humans are social. Each of us is one part of the total society and it is difficult for us to live, learn, and improve ourselves, alone and isolated. Therefore, we should design machines with social properties. Can we let machines evolve by imitating human society so as to achieve more effective, intelligent, interpretable "social machine learning"?

And much more.

The last words?

Early computer scientist Alan Kay said, **The best way to predict the future is to create it**. Therefore, all machine learning practitioners, whether scholars or engineers, professors or students, need to work together to advance these important research topics. Together, we will not just predict the future, but create it.

Best wishes to you all and thanks so much for your heroic efforts this semester