Stochastic Machine Learning
01 - Introduction

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

WS 2020/21

Literature (incomplete, but growing):

- I. Goodfellow, Y. Bengio, and A. Courville (2016). **Deep Learning**. http://www.deeplearningbook.org. MIT Press
- D. Barber (2012). **Bayesian Reasoning and Machine Learning**. Cambridge University Press
- R. S. Sutton and A. G. Barto (1998). **Reinforcement Learning : An Introduction**. MIT Press
- G. James et al. (2014). **An Introduction to Statistical Learning: With Applications in R**. Springer Publishing Company, Incorporated. ISBN: 1461471370, 9781461471370
- T. Hastie, R. Tibshirani, and J. Friedman (2009). **The Elements of Statistical Learning**. Springer Series in Statistics. Springer New York Inc. URL: https://statweb.stanford.edu/~tibs/ElemStatLearn/
- K. P. Murphy (2012). **Machine Learning: A Probabilistic Perspective**. MIT Press
- CRAN Task View: Machine Learning, available at https://cran.r-project.org/web/views/MachineLearning.html
- UCI ML Repository: http://archive.ics.uci.edu/ml/ (371 datasets)
- Warren B Powell (2011). **Approximate Dynamic Programming: Solving the curses of dimensionality**. Vol. 703. John Wiley & Sons
- A nice resourse is https://github.com/aikorea/awesome-rl

Motivation

- **Machine Learning** is nowadays used at many places (Google, Amazon, etc.) with a great variety of applications.
- It is a great job opportunity ! It needs maths and probability !
- Many applications are surprisingly successful (speech / face recognition) and currently people are seeking further applications
- Here we want to learn about the foundations, discuss implications and what can be done by ML and what not.

Organization

The lecture will be available online. We meet every Monday 10.15 on ZOOM and discuss the lectures from last week. I expect active participation from your side, the videos should be switched on.

- The lectures will mix python implementation with theory - so it is now a good time for you to start learning python.
- Every lecture ends with a short set of questions. These questions will be discussed in our session on Monday.
- Homework will be done by projects. You can choose a topic which interests you, and we will provide topics. Groups up to 5 people work on a project, more than one group can also work on the same project.
- We provide a shared bitbucket repository for all projects, such that you share your current work and can profit form the work from others.
- **Lars Niemann** is organizing the projects - please contact him for questions.

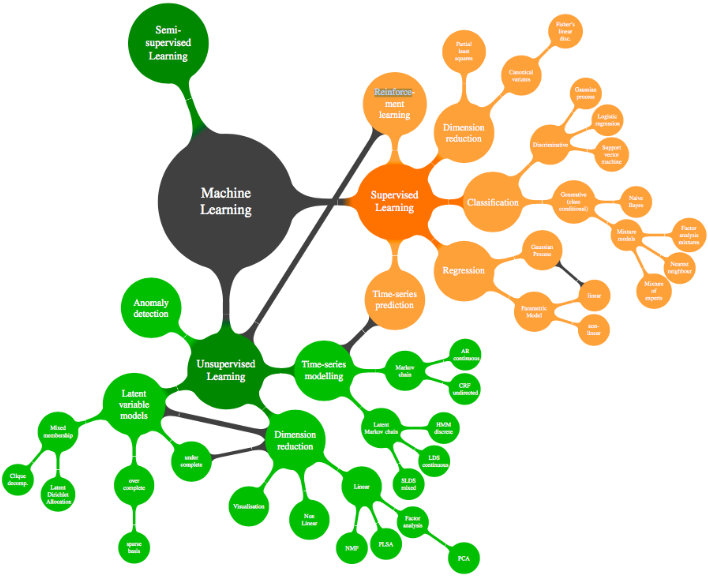
Overview[1]

- Artificial intelligence is the field where computers solve problems.
- It is easy for a computer to solve tasks which can be described formally (Chess, Tic-Tac-Toe). The challenge is to solve a tasks which are hard to describe formally (but are easy for humans: walk, drive a car, speak, recognize people ...)
- The solution is to allow computers to learn from experience and to understand the world by a hierarchy of concepts, each concept defined in terms of its relation to simpler concepts.
- A fixed knowledge-base would be somehow limiting such that we are interested in such attempts where the systems acquire their own knowledge, which we call **Machine Learning**.

[1] This introduction follows closely Goodfellow et.al. (2016).

- First examples of machine learning are **logistic regression** or **naive Bayes** $\rightarrow$ standard statistical procedures (Cesarean delivery / Recognition of Spam, more examples to follow)
- Problems become simpler with a nice representation. Of course it would be nice if the system itself could find such a representation, which we call **representation learning**.
- An example is the so-called **auto-encoder**. This is a combination of an encoder and a decoder. The encoder converts the input to a certain representation and the decoder converts it back again, such that the result has nice properties.
- Speech for example might be influenced by many factors of variation (age, sex, origin, ...) and it needs nearly human understanding to disentangle the variation from the content we are interested in.
- **Deep Learning** solves this problem by introducing hierarchical representations.

- This leads to the following hiearchy:
- AI $\rightarrow$ machine learning $\rightarrow$ representation learning $\rightarrow$ deep learning.



Source: Barber (2012).

Examples of Machine Learning

Some of the most prominent examples:

- LeCun et.al.[2] recognition of handwritten digits. The MNIST Database[3] provides 60.000 samples for testing algorithms. The NIST database is of increased size[4]
- The Viola & Jones face recognition,[5]. This path-breaking work proposed a procedure to combine existing tools with machine-learning algorithms. One key is the use of approx. 5000 learning pictures to train the routine. We will revisit this procedure shortly.
- Imagenet is an image database containing many images classified (cats, cars, etc.)[6]
- Various twitter datasets are available, for example for learning to detect hate speach.
- `Kaggle`[7] is a platfrom where computational competitions are hosted. It also provides many many data examples with it.
- Datasets for machine-learning research on Wikipedia[8].

---

[2] Y. LeCun et al. (1998). „Gradient-based learning applied to document recognition“. In: **Proceedings of the IEEE** 86.11, pp. 2278–2324.

[3] `http://yann.lecun.com/exdb/mnist/`

[4] `https://www.nist.gov/srd/nist-special-database-19`

[5] P. Viola and M. Jones (2001). „Robust Real-time Object Detection“. In: **International Journal of Computer Vision**. Vol. 4. 34–47.

[6] `http://image-net.org`

[7] www.kaggle.com

[8] `https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research`

- Speech recognition has long been a difficult problem for computers (first works date to the 50's) and only recently been solved with high computer power. It may seem surprising, that mathematical tools are at the core of these solutions. Let us quote Hinton et.al.[9]

  > *Most current speech recognition systems use hidden Markov models (HMMs) to deal with the temporal variability of speech and Gaussian mixture models (GMMs) to deter- mine how well each state of each HMM fits a frame or a short window of frames of coefficients that repre- sents the acoustic input. (...)*
  > *Deep neural networks (DNNs) that have many hidden layers and are trained using new methods have been shown to outperform GMMs on a variety of speech recognition benchmarks, sometimes by a large margin*

  So, one of our tasks will be to develop a little bit of mathematical tools which we will need later. Most notably, some of the mathematical parts can be replaced by deep learning, which will be of high interest to us.

[9] Geoffrey Hinton et al. (2012). „Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups“. In: **IEEE Signal Processing Magazine** 29.6, pp. 82–97.

Questions

- Was is artificial intelligence ?
- Was is machine learning ?
- Do you know what a neural network is (look for the history in the internet)?
- What are shallow / deep networks ?
- What are the applications which you find most exciting ?
- What are the applications that you think will have the largest impact on our future?
- Research a bit yourself: look for datasets, look for latest applications etc.