UPPSALA
UNIVERSITET

Machine learning overview with medical applications

Thomas Schön
Uppsala University

April 2, 2019.

"Machine learning gives computers the ability to **learn without being explicitly programmed** *for the task at hand."*

"Anyone making confident predictions about anything having to do with the future of artificial intelligence is either kidding you or kidding themselves."

Andrew McAfee, MIT

What we do in the team — who we are

We automate the extraction of knowledge and understanding from data.

Both basic research **and** applied research (with companies).




Create **probabilistic models** of dynamical systems and their surroundings.

Develop methods to **learn** models from data.

The models can then be used by machines (or humans) to **understand** or **make decisions** about what will happen next.

What do I hope to achieve today?

1. Briefly introduce the scientific field of Machine Learning.
2. Create an **awareness/interest** around this technology.
3. **A bit more specific:** Give a few examples from medicine and offer a (hopefully) intuitive understanding of deep learning.

What is machine learning all about?

Machine learning is about learning, reasoning and acting based on data.

Machine learning gives computers the ability to **learn without being explicitly programmed** for the task at hand.

"It is one of today's most rapidly growing technical fields, lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science."

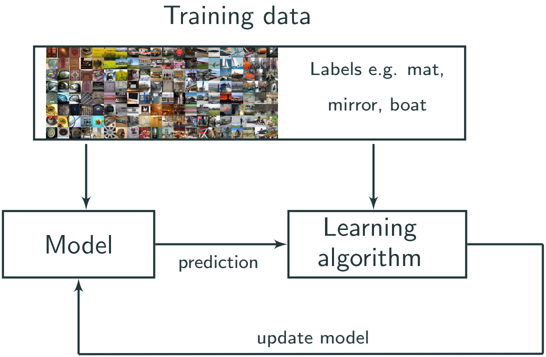
Ghahramani, Z. **Probabilistic machine learning and artificial intelligence**. *Nature* 521:452-459, 2015.

Jordan, M. I. and Mitchell, T. M. **Machine Learning: Trends, perspectives and prospects**. *Science*, 349(6245):255-260, 2015.

Machine Learning (supervised)

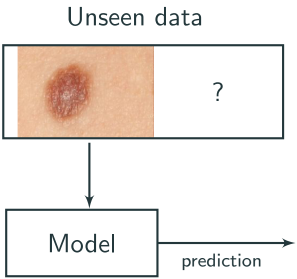
Data on its own is typically useless, it is only when we can extract knowledge from the data that it becomes useful.

Learning a model from labelled data.



Machine Learning (supervised)

Using the learned model on new previously unseen data.



The model must **generalize** to new unseen data.

Unsupervised, reinforcement and semi-supervised learning.

The four cornerstones

Cornerstone 1 **(Data)** Typically we need lots of it.

Cornerstone 2 **(Mathematical model)** A mathematical model is a compact representation of the data that in precise mathematical form captures the key properties of the underlying situation.

Cornerstone 3 **(Learning algorithm)** Used to compute the unknown variables from the observed data using the model.

Cornerstone 4 **(Decision/Control)** Use the understanding of the current situation to steer it into a desired state.

Mathematical models – representations

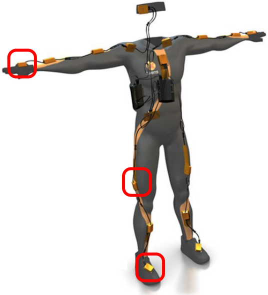
The performance of an algorithms typically depends on which representation (model) that is used for the data.

When solving a problem – start by thinking about **which model/representation to use**!

International Conference on Learning Representations (ICLR)
www.iclr.cc/

Ex (Classical Engineering) – Motion estimation

Aim: Compute the position and orientation of the different body segments of a person moving around indoors (motion capture).



What is a mathematical model?

Illustrate the use of three different models:

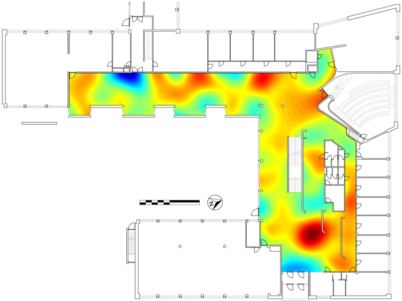
1. Integration of sensor observations.
2. Add a biomechanical model.
3. Add a world model.

Ex (Machine Learning) – Ambient magnetic field map

The Earth's magnetic field sets a background for the ambient magnetic field. Deviations make the field vary from point to point.

Aim: Build a map (i.e., a model) of the magnetic environment based on magnetometer measurements.

Solution: Customized Gaussian process that obeys Maxwell's equations.

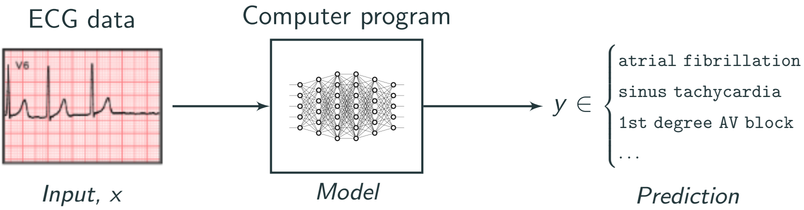


www.youtube.com/watch?v=enlMiUqPVJo

Arno Solin, Manon Kok, Niklas Wahlström, TS and Simo Särkkä. **Modeling and interpolation of the ambient magnetic field by Gaussian processes**. *IEEE Transactions on Robotics*, 34(4):1112–1127, 2018.

Carl Jidling, Niklas Wahlström, Adrian Wills and TS. **Linearly constrained Gaussian processes**. *Advances in Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, December, 2017.

Ex (Machine Learning) – Automatic ECG classification

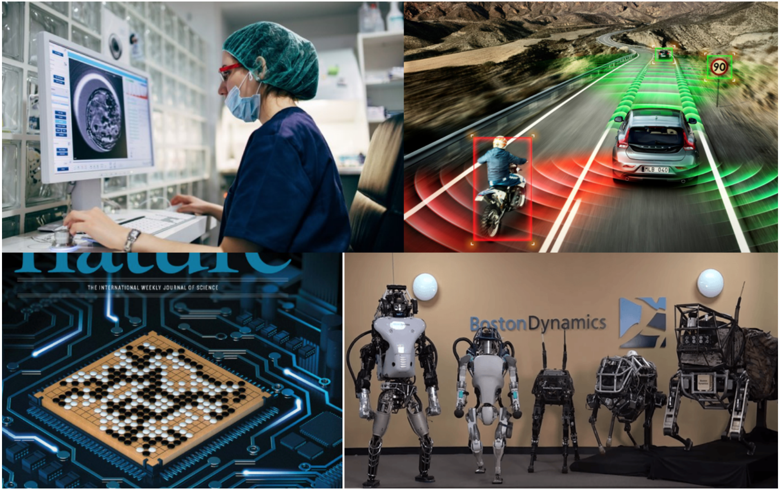


We are now reaching human level (medical doctor) performance on certain specific tasks.

Key difference to "classical engineering": The model is **not** derived based on our ability to mathematically explain what we see in an ECG.

Instead, a generic model is **automatically learned** based on data.

Machine Learning (ML) $\subset$ Artificial Intelligence (AI)



ML provides a **mathematical and algorithmic backbone** of AI.

The model – learning relationship

The problem of learning (estimating) a model based on data leads to computational challenges, both

- **Integration:** e.g. the HD integrals arising during marg. (averaging over all possible parameter values z):

$$p(D) = \int p(D \mid z) p(z) \mathrm{d}z.$$

- **Optimization:** e.g. when extracting point estimates, for example by maximizing the posterior or the likelihood

$$\widehat{z} = \arg\max_{z} \; p(D \mid z)$$

Typically impossible to compute exactly, use approximate methods

- Monte Carlo (MC), Markov chain MC (MCMC), and sequential MC (SMC).
- Variational inference (VI).
- Stochastic optimization.

Key lesson from contemporary Machine Learning

Flexible models often give the best performance.

How can we build and work with these flexible models?

1. Models that use a large (but fixed) number of parameters.
   (**parametric**, ex. deep learning)

   LeCun, Y., Bengio, Y., and Hinton, G. **Deep learning**, *Nature*, Vol 521, 436–444, 2015.

2. Models that use more parameters as we get access to more data.
   (**non-parametric**, ex. Gaussian process)

   Ghahramani, Z. **Bayesian nonparametrics and the probabilistic approach to modeling**. *Phil. Trans. R. Soc. A* 371, 2013.

   Ghahramani, Z. **Probabilistic machine learning and artificial intelligence**. *Nature* 521:452-459, 2015.

"With enough training data the machine can be trained to make very good predictions from previously unseen data."

Outline

Machine learning gives computers the ability to **learn without being explicitly programmed** for the task at hand.

Deep learning – what is it?

The mathematical model has been around for 70 years, but over the last $5 - 7$ years there has been a **revolution**. Key reasons:

1. Very large datasets
2. Better and faster computers
3. Enormous industrial interest (e.g. Google, Facebook, MS)
4. Some methodological breakthroughs

The underlying model is a big mathematical function with **multiple layers of abstraction**, commonly containing millions of parameters.

The parameter values are **automatically** determined based on a large amount of training data.

Constructing a neural network for regression

A **neural network (NN)** is a hierarchical nonlinear function $y = g_\theta(x)$ from an input variable x to an output variable y parameterized by θ.

Linear regression models the relationship between a continuous output variable y and an input variable x,

$$y = \sum_{i=1}^{n} \theta_i x_i + \theta_0 + \varepsilon = \theta^\mathsf{T} x + \varepsilon,$$

where θ is the parameters composed by the "weights" θ_i and the offset ("bias") term θ_0,

$$\theta = \begin{pmatrix} \theta_0 & \theta_1 & \theta_2 & \cdots & \theta_n \end{pmatrix}^\mathsf{T},$$
$$x = \begin{pmatrix} 1 & x_1 & x_2 & \cdots & x_n \end{pmatrix}^\mathsf{T}.$$

Generalized linear regression and NNs

We can generalize this by introducing nonlinear transformations of the predictor $\theta^\mathsf{T} x$,

$$y = f(\theta^\mathsf{T} x).$$

We can think of the neural network as a **sequential construction** of several generalized linear regressions.

Deep neural networks

Let the computer **learn from experience** and understand the situation in terms of a **hierarchy of concepts**, where each concepts is defined in terms of its relation to simpler concepts.

If we draw a graph showing these concepts of top of each other, the graph is **deep**, hence the name deep learning.

It is accomplished by using **multiple levels of representation**. Each level transforms the representation at the previous level into a new and more abstract representation,

$$z^{(l+1)} = f\left(\Theta^{(l+1)} z^{(l)} + \theta_0^{(l+1)}\right),$$

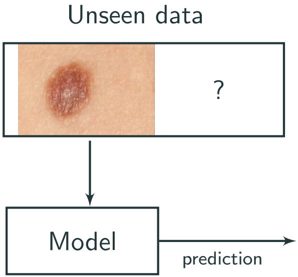
starting from the input (raw data) $z^{(0)} = x$.

Key aspect: The layers are **not** designed by human engineers, they are generated from (typically lots of) data using a learning procedure and lots of computations.

Deep learning – example (skin cancer)

Start from a mathematical model trained on 1.28 million images (**transfer learning**). Make minor modifications of it, specializing to present situation.

Learn new model parameters using 129 450 clinical images ($\sim$ 100 times more images than any previous study).



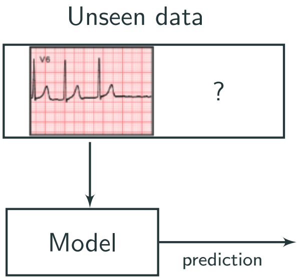
The results are on par with professional dermatologists on specific tasks. Still, far from being clinically useful, but at least they give us “valid reasons to remain cautiously optimistic” as someone said.

Andre Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. and Thrun, S. **Dermatologist-level classification of skin cancer with deep neural networks**. *Nature*, 542, 115–118, February, 2017.

ECG classification – the CODE study

Aim: Predict abnormalities based on a short-duration 12-lead electrocardiogram (ECG) recording.

Current situation: The automated diagnosis that is currently available is not useful.



Background: Joint work with medical doctors from Brazil with an urgent need for automated analysis due to the **vast distances** between the patient and a cardiologist with full expertise in ECG diagnosis.

The existing telehealth network provides the data (more than 2 300 000 ECGs), implying some clinical relevance.

CODE = Clinical Outcomes in Digital Electrocardiology

CODE – Contribution

Contribution: An end-to-end deep neural network that recognize 6 types of ECG abnormalities in standard 12-lead short-duration ECGs with a diagnostic performance that is at least as good as medical residents and students.

The 6 abnormalities are:

1. 1st degree AV block (1dAVb),
2. right bundle branch block (RBBB),
3. left bundle branch block (LBBB),
4. sinus bradycardia (SB),
5. atrial fibrillation (AF),
6. and sinus tachycardia (ST).

CODE – Some results

	true label	DNN not present	DNN present	cardio. not present	cardio. present	emerg. not present	emerg. present	stud. not present	stud. present
1dAVb	not present	796	3	797	2	786	13	782	17
	present	3	25	9	19	5	23	2	26
RBBB	not present	788	5	788	5	792	1	790	3
	present	0	34	1	33	8	26	2	32
LBBB	not present	796	1	797	0	796	1	795	2
	present	0	30	3	27	4	26	3	27
SB	not present	808	3	808	3	808	3	807	4
	present	1	15	1	15	2	14	4	12
AF	not present	811	3	811	3	812	2	805	9
	present	1	12	3	10	5	8	1	12
ST	not present	787	4	790	1	788	3	787	4
	present	1	35	6	30	2	34	6	30

(predicted label)

F1 Score			
DNN	cardio.	emerg.	stud.
0.893	0.776	0.719	0.732
0.932	0.917	0.852	0.928
0.984	0.947	0.912	0.915
0.882	**0.882**	0.848	0.750
0.857	0.769	0.696	0.706
0.933	0.896	0.932	0.857

$$F_1 = 2\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$$

CODE – Limitations and future work

This is a **proof-of-concept** study, which has **limitations:**

- We **cannot** (with statistical significance) show that our algorithm is better than the humans we compared against.
- We have not tested the algorithm on other classes of abnormalities.
- The real clinical setting is clearly much more complex...

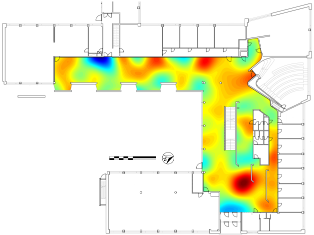
Future work:

- Develop algorithms that can diagnose multiple and complex abn. (myocardial infarction, cardiac chamber enlargement, etc.), less common forms of arrhythmia and to recognize a normal ECG.
- Test the algorithm in a controlled real-life situation, showing that accurate diagnosis could be achieved in real-time. This can have big impact in improving healthcare in low and middle-income countries.

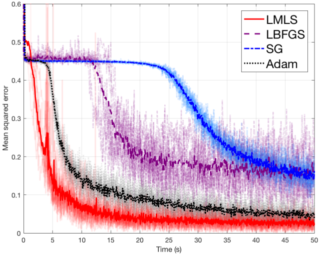
Ribeiro, A. H., Ribeiro, M. H., Paixao, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., Andersson, C. R., Macfarlane, P. W., Meira, W., Schön, T. B. and Ribeiro, A. L. P. **Automatic diagnosis of the short-duration 12-lead ECG using a deep neural network: the CODE study**. Submitted, 2019.

Active research topics – slightly technical

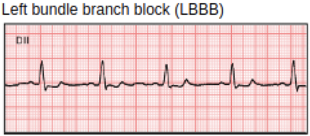
1. **Probabilistic modelling**
   a) General: Flexible models, in particular the Gaussian process (GP), deep GPs.
   b) Specific: **Dynamical** phenomena and their surroundings.



2. **New relevant algorithms**
   a) Large-scale optimization
   b) Approximate integration/inference
      i) Sequential Monte Carlo
      ii) Variational inference
      iii) Markov chain Monte Carlo



Active research topics – slightly technical

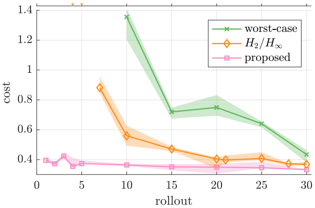


3. **Deep learning (DL)**
   a) Deep probabilistic constructions
   b) Representing (and understanding) uncertainty within deep learning (including Bayesian DL)
   c) Mathematical understanding of DL

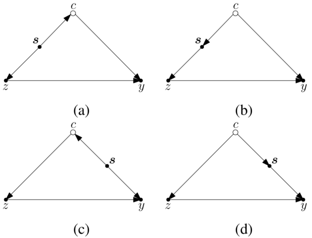
4. **Probabilistic programming**
   a) Potential to automate modelling!
   b) Developing our own probabilistic programming language (Birch)

Active research topics – slightly technical

5. **Reinforcement learning/control**
   a) Learn how to control
   b) Mathematical guarantees



6. **Causality** (new topic)
   a) Aiming to learn causal relationships (not just associations/correlations)
   b) Naturally leads to the need for combining human knowledge **and** data.



7. **Self-supervised** learning (new topic)
   a) Use small amount of labeled data and large amounts of unlabeled data.

New AI initiative at Uppsala — AI for the sciences

It is important for us to **start controlling this development** rather than having the development controlling us.

The university has an important possibility/duty in helping society on this important topic.

Use of our rather unique position:

1. a broad university
2. very good standing within several classical branches of science.

AI for the sciences – A temporary team

The **heart** of this plan consists in a **time-limited five year effort** consisting of an **antidisciplinary entity**.

We want a **real team that is collocated**:

- work together on a regular basis during a longer time span.
- create a reality where **trust** can be built.
- Important mechanism: AI internal sabbatical periods lasting around 12 months.
- The fellows are expected to bring along one or several of their PhD students/post-docs to develop the ideas formed within the entity.
- Post-doc's hired jointly by the entity and by the fellows.

What did I hope to achieve today?

1. Briefly introduce the scientific field of Machine Learning.
2. Create an **awareness/interest** around this technology.
3. **A bit more specific:** Give a few examples from medicine and offer a (hopefully) intuitive understanding of deep learning.

Conclusion

Machine learning gives computers the ability to **learn without being explicitly programmed** for the task at hand.

Uncertainty is a key concept!

The best predictive performance is currently obtained from **highly flexible** learning systems, e.g.

1. Deep learning
2. Gaussian processes

Remember to talk to people who work on **different problems** with **different tools**!! (Visit other fields!)