Machine learning and its use as a tool in research

A very subjective journey, with objective implications

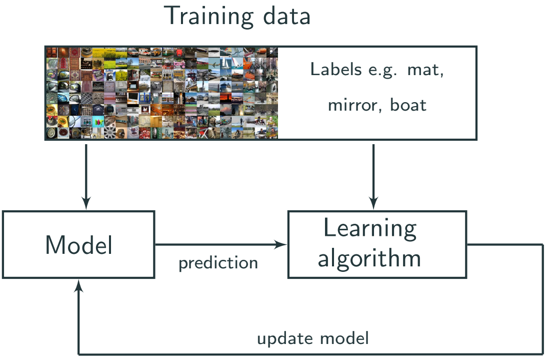
Thomas Schön
Uppsala University
Sweden

Swedish Academic Initiative for Nuclear Technology (SAINT) workshop
Online
January 13, 2020.

Machine learning (supervised)

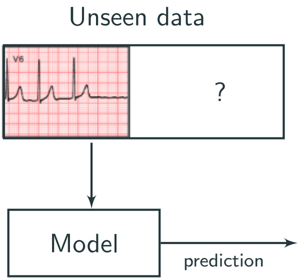
Data on its own is typically useless, it is only when we can extract knowledge from the data that it becomes useful.

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.

Machine Learning – the four cornerstones

1. Data Typically we need lots of it.

2. Mathematical model A compact representation of the data that in a precise mathematical form captures the key properties.

3. Learning algorithm Used to compute the unknown variables from the observed data using the model.

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

Key lesson from contemporary Machine Learning

Flexible models often give the best predictive 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.

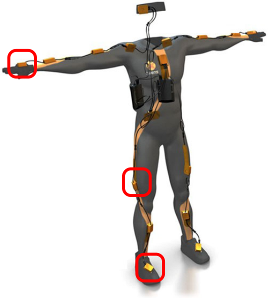
Be careful as flexible models can be deceptive!

Mindset – classical approach via motion estimation

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

Available sensors:

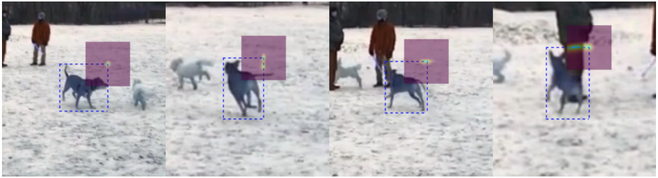
1. Accelerometers
2. Gyroscopes
3. Magnetometers (not used)
4. ultra-wideband



Show movie!

Mindset – Machine learning (deep probabilistic regression)

Aim: Estimate a bounding box of a target object in every frame of a video. The target object is defined by a given box in the first video frame.

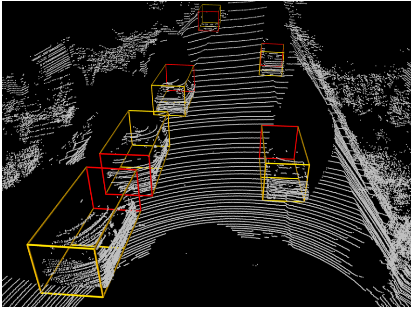


Key difference to classical approach: The model is **not** derived based on our ability to mathematically explain what we see in the image.

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

Fredrik K. Gustafsson, Martin Danelljan, Goutam Bhat, TS. **Energy-based models for deep probabilistic regression**. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Online, August, 2020.

Mindset – Machine learning (deep probabilistic regression)



Aim: Detect objects from sensor data (here laser), estimate their size and position in the 3D world.

Key perception task for self-driving vehicles and autonomous robots.

New probabilistic regression formulation based on deep neural networks.

The **combination** of **probabilistic models** and **deep neural networks** is very exciting and promising.

Fredrik K. Gustafsson, Martin Danelljan, and TS. **Accurate 3D object detection using energy-based models**. Submitted, October, 2020.

Aim and outline

Aim: Motivate (using three/four concrete examples) the use of ML as a tool in research.

Outline:

1. Introduction
2. **Gentle deep learning background**
3. Three concrete examples
   a. **Medicine** – human-level ECG diagnosis
   b. **Physics** – magnetic fields
   c. **Physics** – CT reconstruction
   d. **Physics** – Strain reconstruction (if there is time)

Note: Trying to convey the organic research journey our team has made over the past three years motivating the case for AI4Research.

Deep learning – what is it?

The mathematical model has been around for 70 years, but over the past decade 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

Underlying idea: when representing a function, a deep, hierarchical model can be **exponentially more efficient** than a shallow model.

The functional representation has **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 $\theta^\mathsf{T} x$,

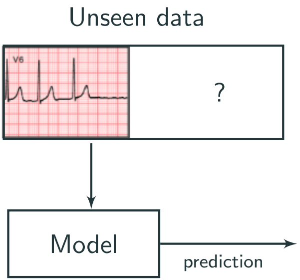
$$y = f(\theta^\mathsf{T} x).$$

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

Medicine – ECG diagnosis

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

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



Background: Joint work with cardiologists and ML engineers 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.

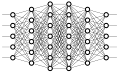
The existing telehealth network provides the data (more than 2 300 000 ECGs), implying some clinical relevance.

Medicine – Automatic human-level ECG diagnosis

ECG data

Input, x

Computer program



Model

$$y \in \begin{cases} \texttt{atrial fibrillation} \\ \texttt{sinus tachycardia} \\ \texttt{1st degree AV block} \\ \dots \end{cases}$$

Prediction

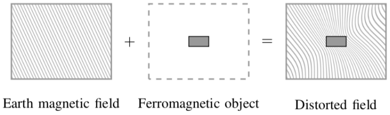
We are now reaching human level (medical doctor) performance on certain specific tasks.

Key difference to classical approach: 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.

Antonio H. Ribeiro, Manoel H. Ribeiro, Gabriela M.M. Paixao, Derick M. Oliveira, Paulo R. Gomes, Jessica A. Canazart, Milton P. S. Ferreira, Carl R. Andersson, Peter W. Macfarlane, Wagner Meira Jr., TS, Antonio Luiz P. Ribeiro. **Automatic diagnosis of the 12-lead ECG using a deep neural network**. *Nature Communications*, 11(1760), 2020.

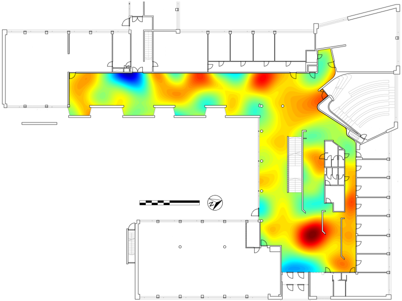
Physics – Ambient magnetic field map



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.



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 (NeurIPS)*, Long Beach, CA, USA, December, 2017.

Blending prior knowledge and data

Recall the two mindsets: While we can do a lot with our data and flexible black-box models, we have already understood a lot about nature.

Obvious idea: What if we could combine the two?!

Meaning that we start from small (rigid) models describing the phenomenon we are studying and augment them with flexible models driven by data.

Personal opinion: I believe that there are (massive) gains to be made in the simple combination of flexible data-driven models and solid widely available knowledge that we already have.

Vision

Create flexible model building blocks **containing** the basic knowledge we have about the phenomenon we are studying.

I stress the fact that the model should be **flexible** enough to allow for new knowledge to be gained.

The data complements our existing basic knowledge and adapts it to the specific situation we are studying.

Has the potential to also allow us to learn new basic knowledge.

Reflection: Quite obvious really, but surprisingly little has been done, but the idea is gaining traction.

I foresee such building blocks containing basic knowledge about physics, medicine, chemistry, biology, etc.

Blending prior knowledge and data

Resulting technical challenge: How can we use known structures and domain knowledge to design priors?

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)}$$

Once we have designed such a prior it will effectively **restrict the flexibility in a goal-oriented fashion**.

Question: What is the right blend of such priors and data?

Starting somewhere – linear functional constraints

Fact: Linear functional constraints and measurements are **useful** in describing nature and **simple** to work with.

Very specific examples:

1. The magnetic field H is curl-free (recall example from before)

$$\nabla \times H = 0.$$

2. Measurements are expressed as line integrals of the target function
   - X-ray computed tomography (CT)
   - Strain field reconstruction from neutron diffraction experiments

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

Johannes Hendriks, Carl Jidling, Adrian Wills, TS. **Linearly constrained neural networks**. *arXiv:2002.01600*, March, 2020.

2	16	13	3
11	5	8	10
7	9	12	6
14	4	1	15

4	9	2
3	5	7
8	1	6

Computed tomography (CT)

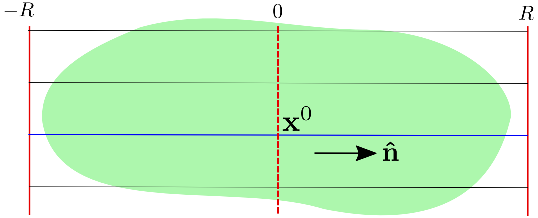
Tomographic reconstruction: Recover the internal structure

$$f(x), \qquad x = [x\ y]^{\mathsf{T}}$$

of an object from irradiation experiments.

Line integral measurements

$$y = \int_{-R}^{R} f(x^0 + s\hat{\mathbf{n}})ds + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$



Limited data (sparse projections) important.

Linear functional measurements in GPs (more general)

Model the target function $f(x)$ as a GP

$$f(x) \sim \mathcal{GP}(0, k(x, x'))$$

Fact: a GP is closed under linear transformations:

$$\mathcal{L}f(x) \sim \mathcal{GP}(0, \mathcal{L}\mathcal{L}'k(x, x'))$$

where for us (in the CT case)

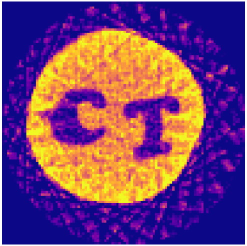
$$\mathcal{L}f(x) = \int_{-r}^{r} f(x^0 + s\hat{\mathbf{n}})ds,$$

Our CT and strain field reconstruction examples have measurements:

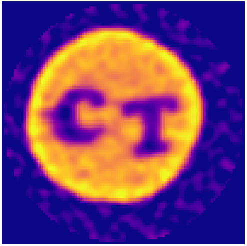
$$y = \int_{-r}^{r} f(x^0 + s\hat{\mathbf{n}})ds + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, Q)$$

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.

Ground truth



FBP



GP

Question: Why is the GP solution so blurry?

All details on this construction are available in

Zenith Purisha, Carl Jidling, Niklas Wahlström, TS and Simo Särkkä. **Probabilistic approach to limited-data computed tomography reconstruction**, *Inverse Problems*, 35(10):105004, 2019.

Extending the expressiveness to non-stationary behaviors

The covariance function $k(x, x')$, stipulates the basic behavior of the target function $f(x)$.

The selection of $k(x, x')$ is the most crucial part of GP modelling.

Extend the expressiveness of stationary covariance functions by transforming the inputs through a nonlinear mapping $u(\cdot)$ to form $k(u(x), u(x'))$, effectively opening up for non-stationary behaviors.

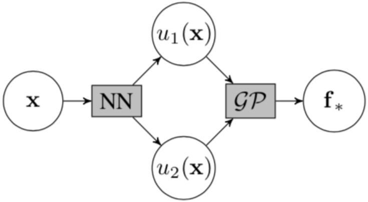
Question: Which mapping should we use?

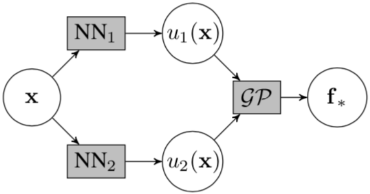
Let's try a deep neural network...

Roberto Calandra, Jan Peters, Carl E. Rasmussen, and Marc P. Deisenroth. **Manifold Gaussian processes for regression**. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2016.

Andrew G.Wilson, Zhiting Hu, Ruslan R. Salakhutdinov, and Eric P. Xing. **Deep kernel learning**. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

One useful way of combining deep learning with GPs

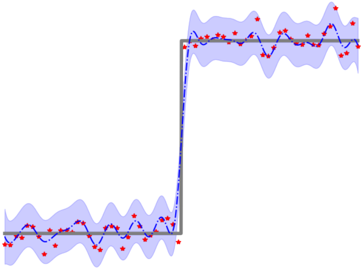




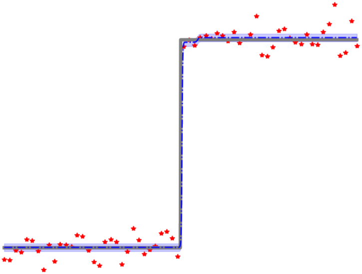
Intuition: The neural network does not have to learn the complete function $f(x)$, but only identify its discontinuities while for the remaining part the model can rely upon the regression capabilities of the GP.

Ex. – illustrating the idea

$$k(x, x') = \sigma_f^2 e^{-\frac{1}{2l^2}(x-x')^2}$$

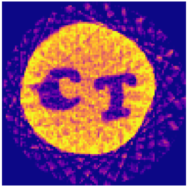


$$k(x, x') = \sigma_f^2 e^{-\frac{1}{2l^2}(u(x)-u(x'))^2}$$

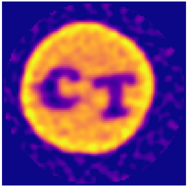


Using the idea together with integral measurements

Ground truth



FBP



GP

GP + DL

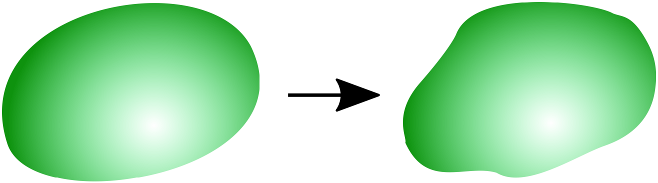
GP + DL: Deep learning to use the input mapping together with our taylored GP prior encoding our understanding of the underlying physics.

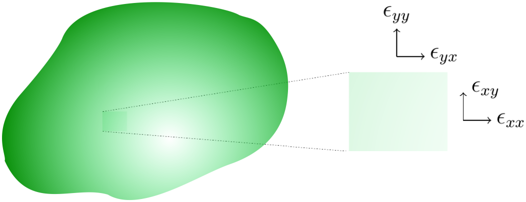
Recall our vision: Create flexible model building blocks containing the basic knowledge we have about the phenomenon we are studying.

Physics – strain field reconstruction

Tomographic reconstruction: Recover the internal structure of an object from irradiation experiments.

Deformed object

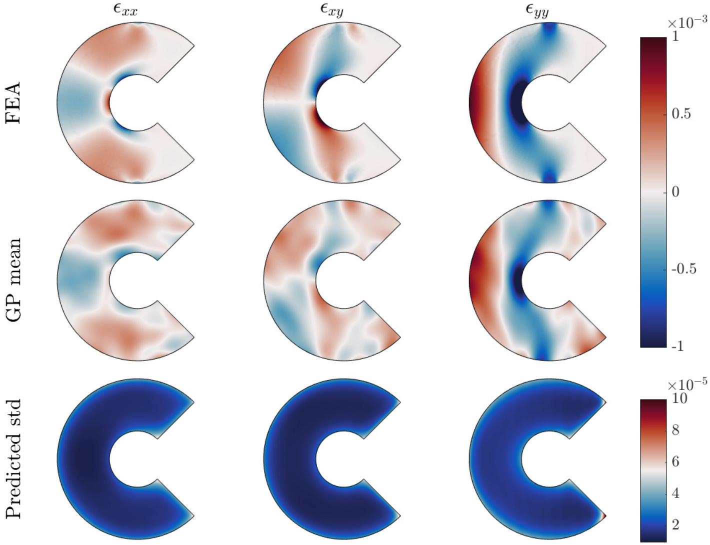




Reconstruct the **strain tensor**

$$\epsilon(x) = \begin{bmatrix} \epsilon_{xx}(x) & \epsilon_{xy}(x) \\ \epsilon_{xy}(x) & \epsilon_{yy}(x) \end{bmatrix}$$

Strain field reconstruction – experimental results



Carl Jidling, Johannes Hendriks, Niklas Wahlström, Alexander Gregg, TS, Chris Wensrich and Adrian Wills. **Probabilistic modelling and reconstruction of strain**. *Nuclear instruments and methods in physics research: section B*, 436:141-155, 2018.

AI4Research – new university-wide AI project

At Uppsala University we will **develop and make use of AI/ML for research**.

A **time-limited five year effort** consisting of an **antidisciplinary entity** from the entire university.

Located in newly refurbished premises at our main library Carolina Rediviva.

Key mechanism: **Internal AI sabbatical periods**

- Funded 50% by the entity and the rest by the department where the fellow remains employed/external grants.
- Duration: around 12 months.
- The fellows brings along 1-2 of their PhD students/post-docs.

AI4Research – new university-wide AI project

Read about the research from the project website

www.uu.se/forskning/ai4research

New positions for 2022 opens soon!

Conclusion

While ML techniques are used more and more in industry, scientists are—for good reasons—becoming aware of the potential in using ML in fundamental research.

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

Showed three (or four) concrete examples motivating **AI4Research**.

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