Machine learning for ocean exploration

Peter Gerstoft, UCSD NoiseLab
University of California San Diego

Website: noiselab.ucsd.edu

Machine learning is based on
- Statistical principles (Bayesian)
- Linear algebra
- **Data**
- Further, there is low-hanging fruit

- **Machine Learning (ML)** and **Artificial Intelligence (AI)** are closely related fields
- ML methods combine **statistics** and computer science principles to obtain models that describe phenomena, or accomplish specific tasks
- ML is based on theory for model identification, control, and estimation. Neural network approximations are solved rapidly on modern architecture, leading to great excitement
- E.g., I taught a class on data assimilation with 10 graduate students, In 2016 I changed it to

"Machine learning for physical applications" with 65 graduate students, 245 students in 2020.

=> to attract students, do some ML

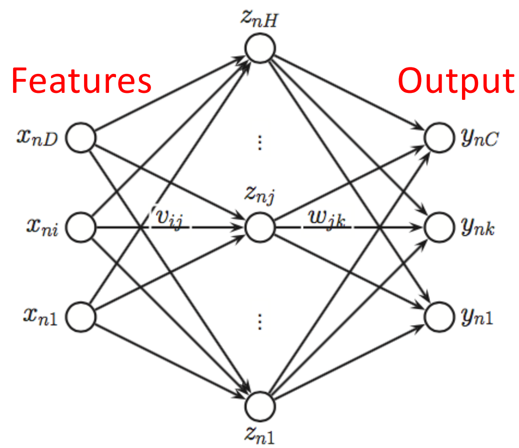
ML Principles: Learning from data

In ML, we are often interested in training a model to produce a desired output given inputs

y=f(x) $+\epsilon$

Input $\mathbf{x} \in \mathbb{R}^N$, N features
output $\mathbf{y} \in \mathbb{R}^P$, P outputs

- **Supervised learning**: the P outputs have labelled examples (response variables **y**)
- **Unsupervised learning**: there are no labels. The goal is to find interesting properties from **x,** as an autoencoder **x̃=f(x)**
- ….. and we *train* the model **on data**
- ….and test the model on new data.
- Testing measure the ability to **generalize**

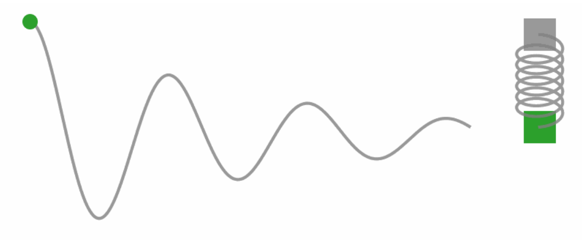


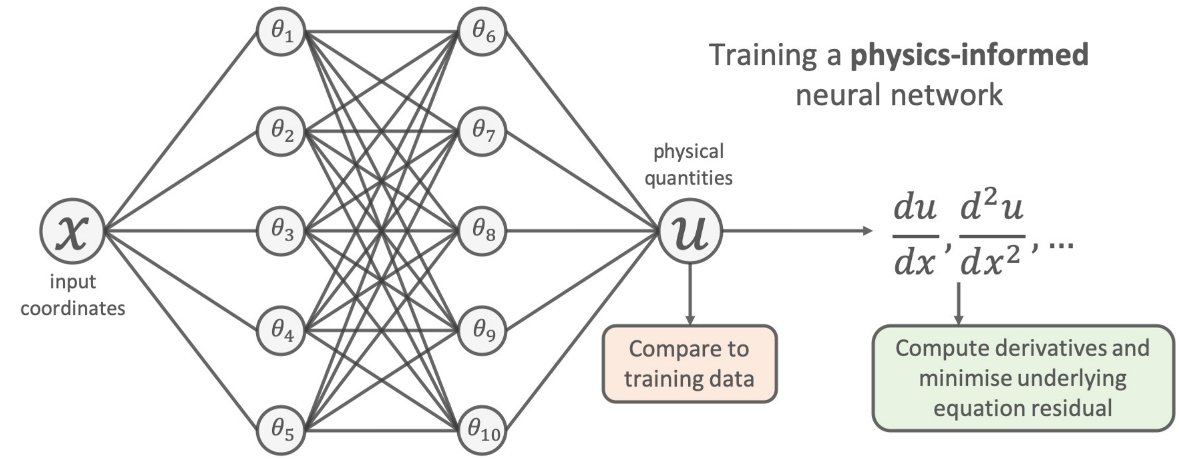
- Most relevant when data can't be explained by existing models

Much more than neural networks
Dictionary leaning
Gaussian Processes
Uncertainty prediction
Graph processing

We need physics in the ML, e.g.   Physics-Informed NN

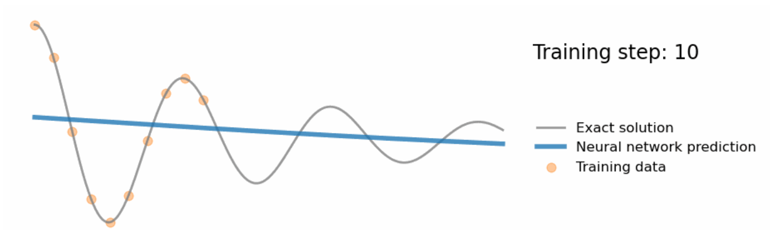
PHYSICS !



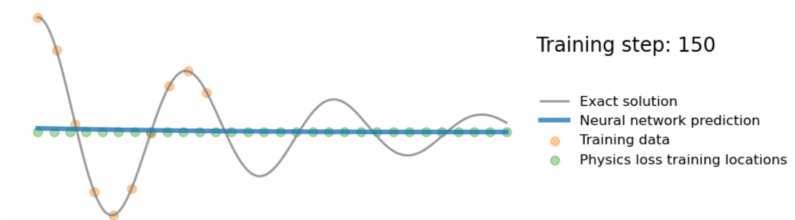


-**input** training locations
-**output** displacement
- **gradients** of NN output with respect to its input are computed (analytically available for NN).
- **Residual** of the PDE using these gradients is added to the loss function.

$$\min \frac{1}{N}\sum_{i}^{N}(u_{\mathrm{NN}}(x_i;\theta) - u_{\mathrm{true}}(x_i))^2$$
$$+\frac{1}{M}\sum_{j}^{M}\left(\left[m\frac{d^2}{dx^2} + \mu\frac{d}{dx} + k\right]u_{\mathrm{NN}}(x_j;\theta)\right)^2$$



Doesn’t generalize



Generalize well

Uncertainty quantification

noiselab.ucsd.edu

In predicting a value, the uncertainty is equally important

- Bayesian methods such as Markov Chain Monte Carlo can be slow, but give the full PDF

- Gaussian processes

- UQ often just characterize the uncertainty, not a PDF

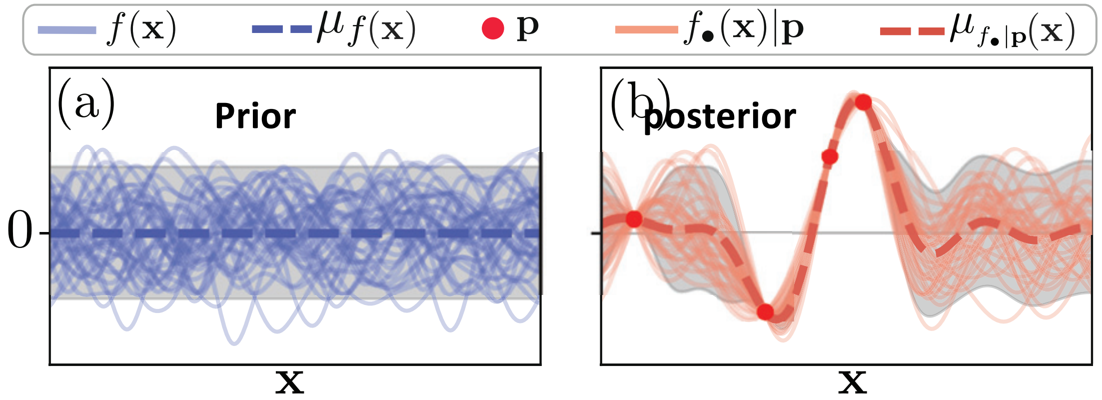
Conformal prediction guarantees a confidence interval

Gaussian process:

$$f(x) \sim GP(m(x), k(x, x')),$$

The kernel depend on distance between the points

$$\kappa(r_i, r_j) = \sigma_f \exp\left(\frac{-\|r_i - r_j\|^2}{2\ell^2}\right),$$

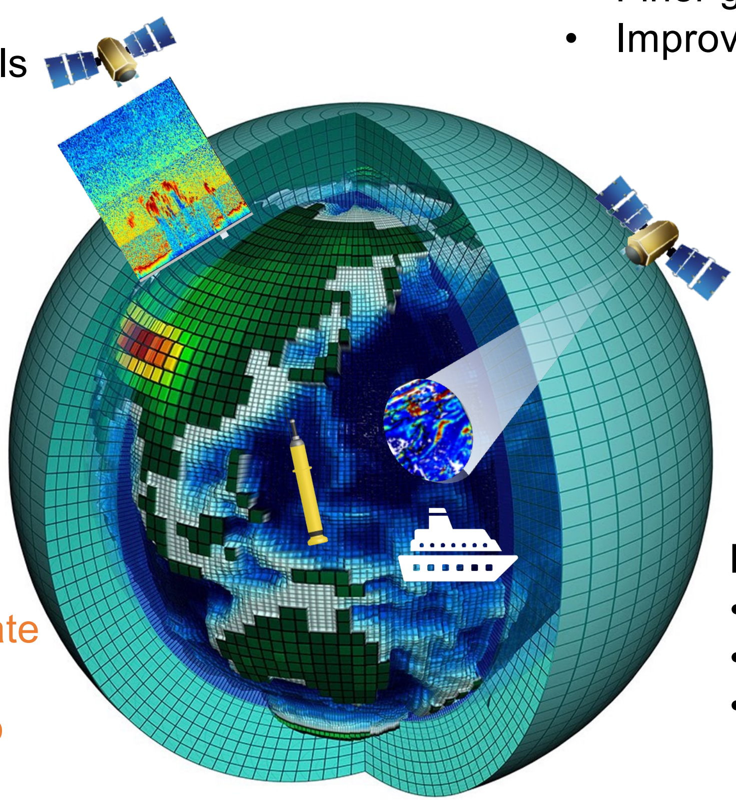


Related topics:
- Propagation of uncertainty
- Maybe by uncertainty for the latent variable
- Explainable models
- Pattern detection
- Unsupervised learning

ML for Ocean + Climate sciences

Modelling:
- **Learning** emulators and surrogates
- Both within and across models, for:
  - Higher resolution
  - Better physics (surrogate models for unresolved processes)
  - Faster compute (NN)

Observations:
- Feature detection
- Unsupervised clustering and outlier detection
- Finer grained classifications
- Improved trend analysis / event attribution



Classical Data assimilation:
- Produce reanalyses to **train** surrogate models
- Analysis residuals are input to ML to find patterns to improve the models

Data assimilation with ML:
- Can ingest more observational data
- Can learn observational operators
- Currently UQ needs improvement

Many of the things we already do, but faster and more precise!

ML in next decade

AI assistants for scientists to perform tasks that can be specified: Literature survey, gather data, write code to implement the analysis, create draft papers, review papers, …

- Goals: Human learning, problem solving (e.g., improved forecasts)
- Manual labelling could be reduced with unsupervised or semi-supervised learning
- Combine **multimodal sensing** with AI
  - Video and sound In observatories
  - Satellite could predict ocean background noise. Observing rain, waves and chlorophyll
  - Assimilation of SAR images with complicated physics
- Predict turbulence
- Hyperspectral sensing
- Climate change
- Ocean modelling
- Humans working closely with AI to vet the output and to learn

- Reduce the US ***Academic Research Fleet (ARF)*** with robots or wave gliders and AUVs all controlled by ML

- **Risk of AI**: Fake papers, Too many papers, replace apprenticeships (grad students), perhaps scientists

None of these people exist! With GAN

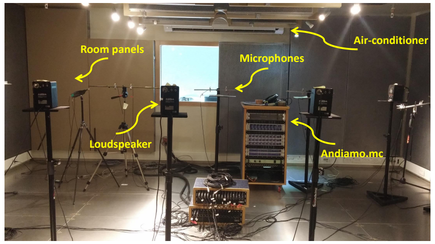
Machine learning in acoustics: Theory and applications

J Acoustical Soc. Am. Nov 2019

Freely available @
https://doi.org/10.1121/1.5133944
42 pages, 350 citations

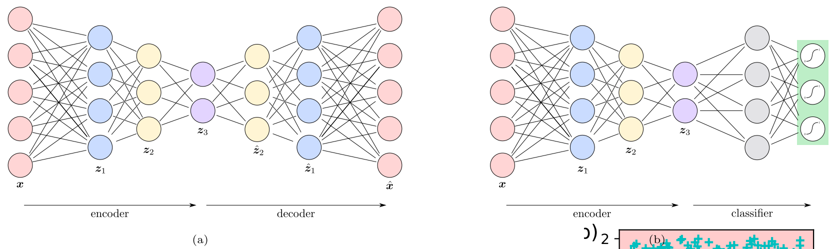
Michael J. Bianco,[1,a)] Peter Gerstoft,[1] James Traer,[2] Emma Ozanich,[1] Marie A. Roch,[3] Sharon Gannot,[4] and Charles-Alban Deledalle[5]
[1]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA*
[2]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*Department of Computer Science, San Diego State University, San Diego, California 92182, USA*
[4]*Faculty of Engineering, Bar-Ilan University, Ramat-Gan 5290002, Israel*
[5]*Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, California 92093, USA*

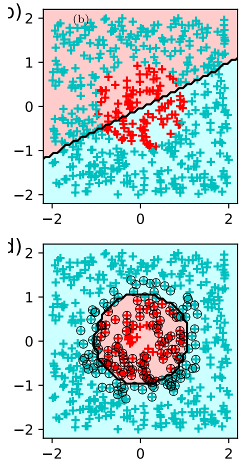
(Received 9 May 2019; revised 23 September 2019; accepted 14 October 2019; published online 27 November 2019)

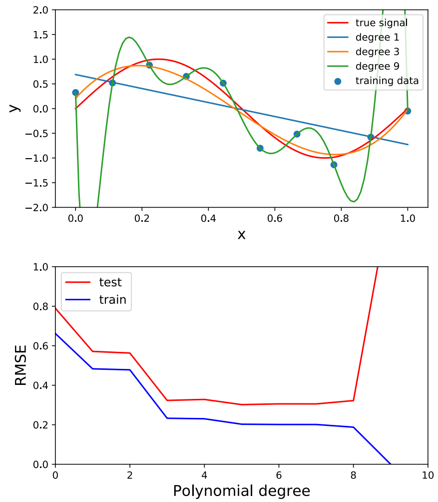


Comprehensive review, including:

- ***Introduction to ML theory***
- ***Deep learning***
- ***Source localization in speech processing***
- ***Source localization in ocean acoustics***
- ***Bioacoustics***
- ***Human perception of sound***







Much more than neural networks

Dictionary leaning
Gaussian Processes
Uncertainty prediction
Graph processing

Discussion (red text is questions posed

- *What sort of new capabilities are on the horizon and what might the next 10 years look like and beyond?*
  - a fleet of ocean robots
  - continuous high-resolution Climate and Ocean assimilation models
  - AI assistants for scientists to perform tasks that can be specified:
    Literature survey, gather data, write code to implement the analysis, create draft papers, review papers, …

- Discussion session with these questions:
- What new discoveries in ocean science does AI/ML enable?
  - See previous slides
  - Chat GPT for Climate Data, Ocean, no advisor needed!

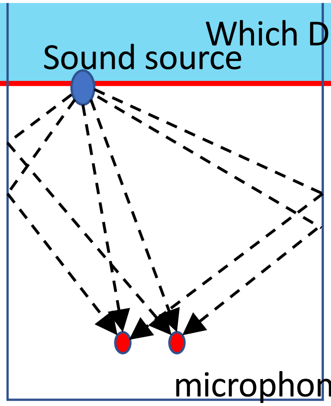
- What data are required for these discoveries?
  - UQ
  - Careful test-data / cross-validation for generalization
  - Explainability
  - Expertise in the pier-review community
  - Data server, similar to Iris. NSF or maybe some ONR data

- What are challenges for AI/ML research (computing, data aggregation, access, training)?
  A web portal for all data
  - seismic data are on iris.edu
  - Today each PI stores his own data. Ocean data (NSF) and acoustics (ONR)
  - we need all data stored together
  - Meta data too, but just start with any data
  - "Ocean Chat GPT" , might make this easier
  - Increase availability of ocean data, especially acoustic data

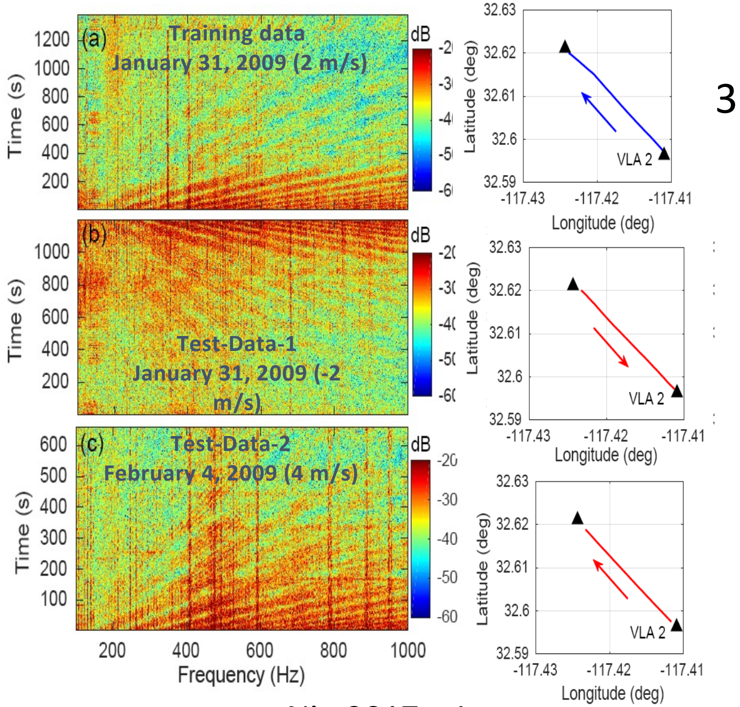
- Risk of AI
  - Larger critical mass
  - Too many papers,
  - Adversarial review!
  - Replace students and postdocs?

- Not Used

ML source range classification



Array Data: 300–950Hz with 10Hz increment, i.e., 66 frequencies.
16 hydrophones with 1 m spacing



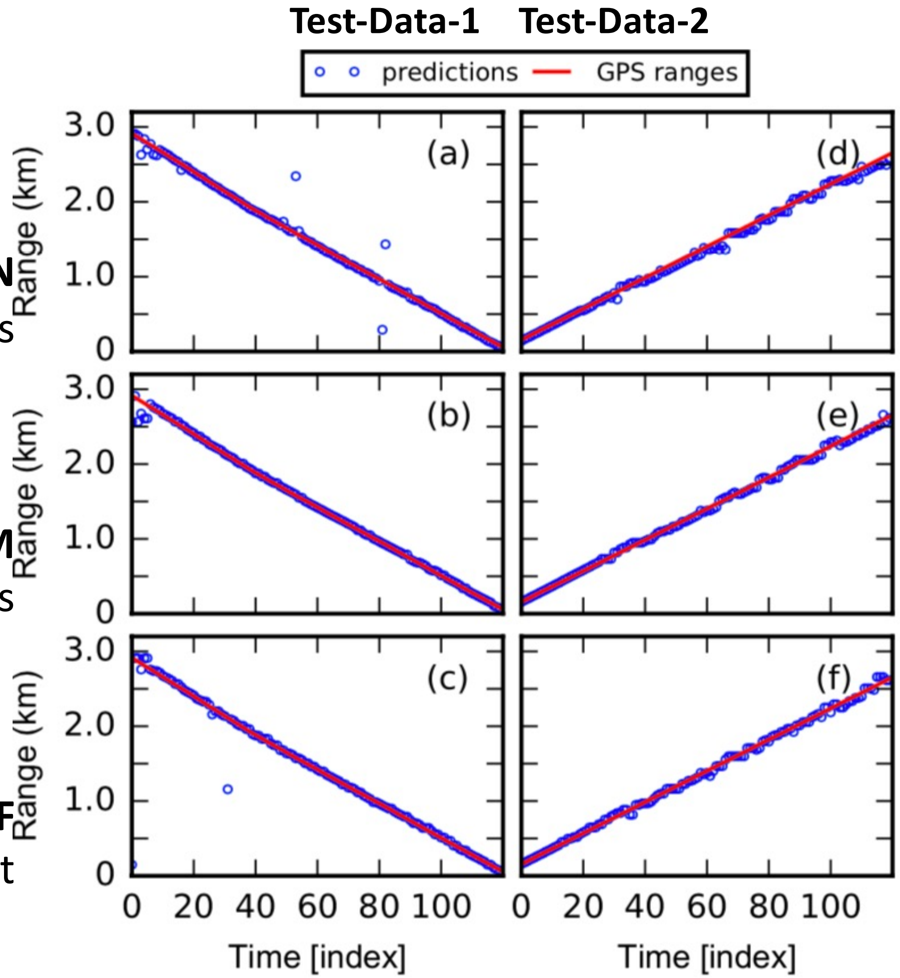
Niu 2017a, JASA

First NN is trained with one source

FNN
3 hidden layers with 512 nodes

SVM
Radial basis Functions

RF
Random forest



Summary

- Machine learning, big data, data science, artificial intelligence are similar.
- We need explainable artificial intelligence. We want the ML algorithm to provide a line of reasoning together with the calculated result / fit / decision.

Can ML (Peter's 2019 list)

- Replace CTBTO processing chain?
- Discover PDE (Partial differential equation) in video?
- Find sea mines?
- Design metamaterials?
- Predict earthquakes?
- Replace 50 years of array processing?
- Source location in the ocean waveguide w/o training.