AI-Driven Innovations in Power Grid Management

David Wozabal

Ksandr Live XL, 09/2024

Outline

Who am I?

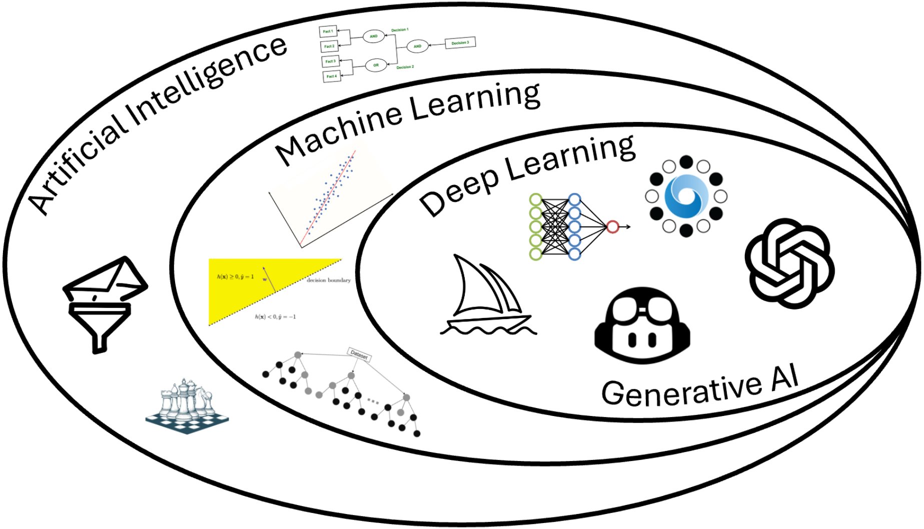
- Professor for Management Science and Machine Learning
  - At VU since 2023
  - Before: Professor at the Technical University of Munich.
- Background
  - Mathematics, Statistics & Machine Learning
  - Operations Research
- Research
  - Stochastic Programming & Reinforcement Learning
  - Optimal decisions for generation and flexibility
  - Stochastic thermal-electric power flow in low-temperature, low voltage grids
  - Smart charging of electric vehicles

Purpose of this talk

- Review state of the art AI methods relevant to energy systems modeling.
- Which methods to use for what?
- Discuss possible applications in the Dutch grid.

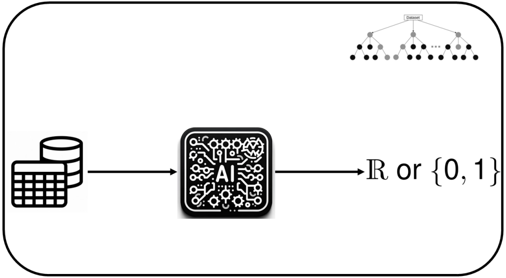
Outline

The Zoo of AI Methods

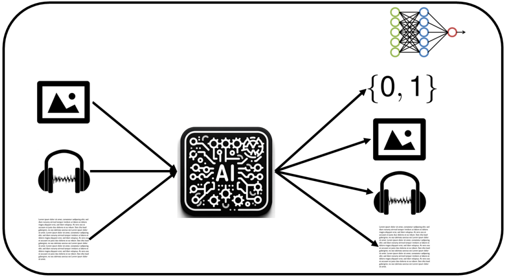


What is the difference?

Traditional Machine Learning



Deep Learning



Requirements

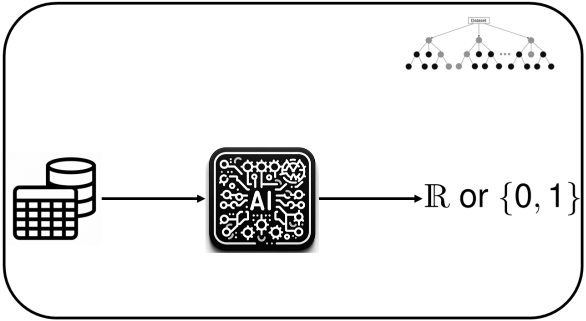
- Manageable data requirements
- Small teams
- Standard hardware

Requirements

- Large training sets
- Considerable expertise
- Massive compute for training

Outline

Forecasting with Tabular Data

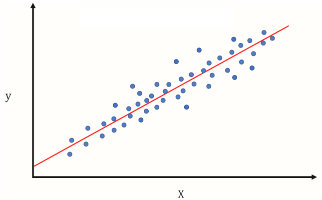


Classic ML methods outperform neural networks on tabular data.

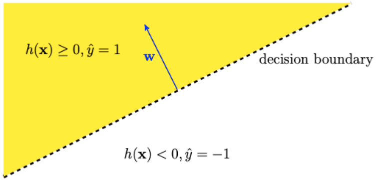
- Lack of locality
- Mixed feature types
- Lack of prior knowledge

Shwartz-Ziv, R., & Armon, A. *Tabular data: Deep learning is not all you need*. Information Fusion, 2022.

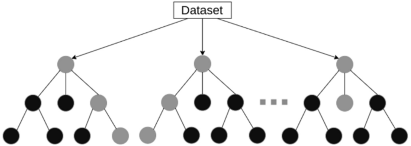
Methods



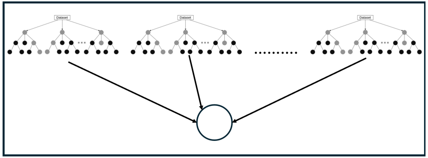
- Simple linear regression
- Polynomial regression
- LASSO



- Support vector machines (SVM)
- Non-linear transformations



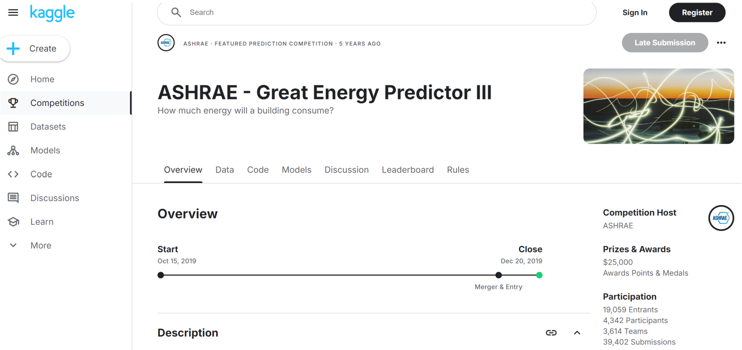
- Decision trees
- CART

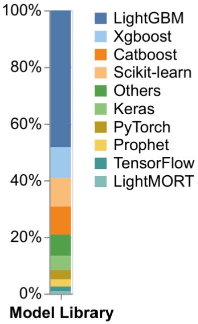


- Ensemble methods
- Random forest
- Boosted trees

Case: Great Energy Predictor III

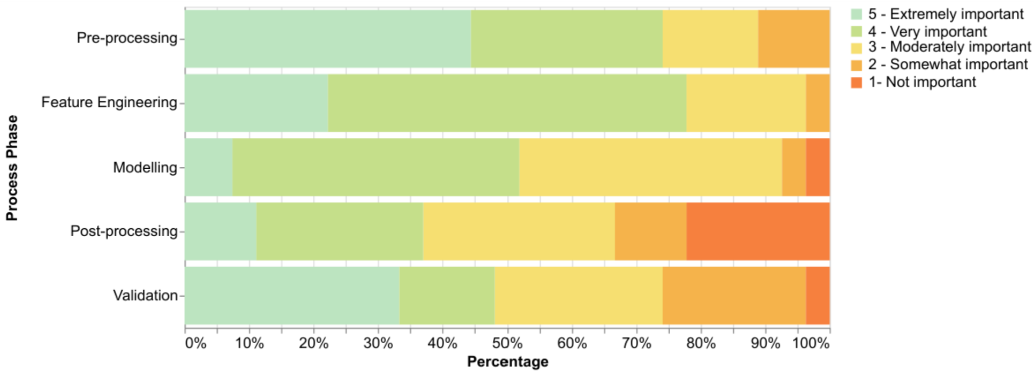
- Predict building energy usage
- Kaggle competition with more than 3600 teams competing
- Teams that scored in the top 5% mostly used gradient boosted trees





Miller, C., Hao, L., & Fu, C. *Gradient boosting machines and careful pre-processing work best: Ashrae great energy predictor III lessons learned*. arXiv 2022.

Importance of Different Phases of Model Building



- Preprocessing, feature engineering, and validation the most important tasks
- Choice of concrete model often not that important

Case: Wind Power Forecasting

Sobolewski, R. A., Tchakorom, M., & Couturier, R. *Gradient boosting-based approach for short-and medium-term wind turbine output power prediction*. Renewable Energy, 2023.

- Short-and medium-term wind power forecasting (48 - 235 hours ahead)
- Based on weather related variables (humidity, pressure, . . .)
- Involved data preprocessing and feature engineering
- Ensemble methods outperform neural networks

Learning algorithm	RMSE [kW]	MAE [kW]
CatBoost	**76.18**	**54.87**
LightBoost	76.84	55.24
XGBoost	77.02	55.61
LSTM	78.73	57.85
DecisionTree	111.26	79.38
RandomForest	77.97	56.14

Outline

From Regression to Neural Networks

Goal: Estimate a functions $y = f(x)$ that maps features x to labels y.

Simplest Machine Learning Model: Linear regression

$$y = f_\phi(x) = b + wx$$

with parameters $\phi = (b, w)$

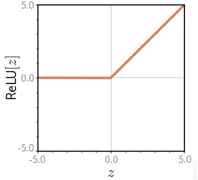
Introducing Non-Linearity: Add non-linear activation function a

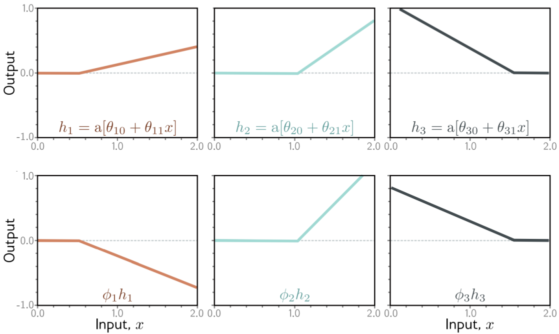
$$y = f_\phi(x) = b_1 + w_{10} a(b_{00} + w_{00} x) + w_{11} a(b_{01} + w_{01} x) + w_{12} a(b_{02} + w_{02} x)$$

Shallow neural network with 3 neurons.

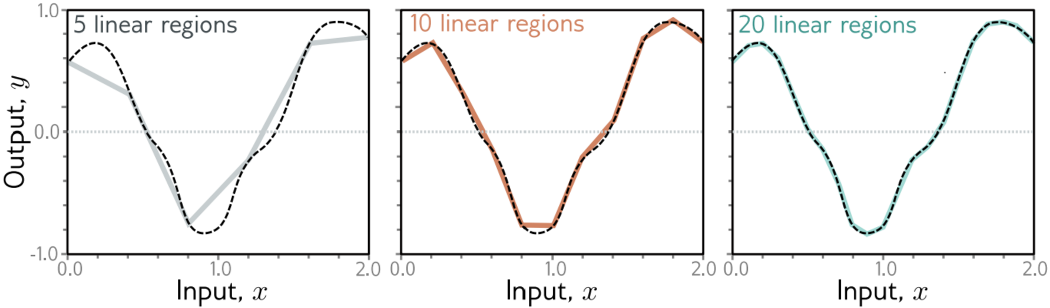
y ... labels   x ... features   w ... weights   b ... biases   a ... activation function

Simplest Activation: Rectified Linear Unit (ReLU)





Plots from Prince (2024)



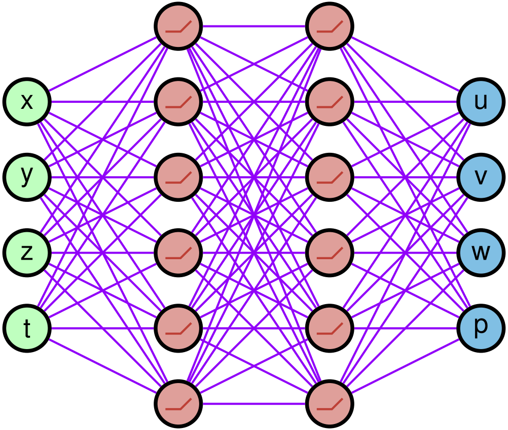
Plots from Prince (2024)

Universal Approximation Theorem

Every *sufficiently nice* function can be approximated to arbitrary precision by a shallow neural network with enough neurons.

- Not very surprising
- Curse of dimensionality

Deep Neural Networks



- Input layer: features (positions, time, velocities, loads, voltages, . . .)
- Hidden layers
- Output layer: predictions for the labels
- Connections: weights
- Activation function

Training a Neural Network

- Choose parameters (weights, biases, ...) to minimize a loss function
  - Measures average deviations of predictions from true labels
  - Example: Mean squared error

$$\mathcal{L}(w, b) = \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(u(x_i) - u^i)^2.$$

- Use stochastic gradient descent for this optimization
- Large networks with many parameters allow for expressiveness

Caveat

Large networks need a lot of data to train.

- Success depends on domain specific tricks
  - Image recognition: convolutional layers and pooling
  - Large language models: attention layers
  - Time series: long short-term memory layers

An Example form Fluid Dynamics

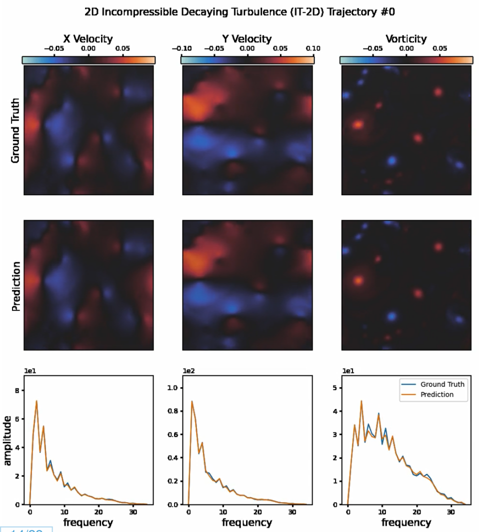
Stachenfeld et al. *Learned coarse models for efficient turbulence simulation*. ICLR(2022).

- Turbulent fluid dynamics
- Chaotic system evolving based on Navier-Stokes PDE
- Neural network trained on only 16 simulations
- Outperforms traditional methods

How is that possible?

An Example form Fluid Dynamics

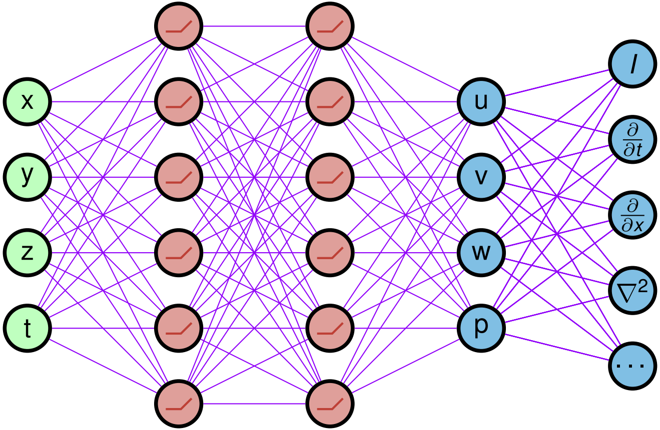
Stachenfeld et al. *Learned coarse models for efficient turbulence simulation*. ICLR(2022).



- Turbulent fluid dynamics
- Chaotic system evolving based on Navier-Stokes PDE
- Neural network trained on only 16 simulations
- Outperforms traditional methods

How is that possible?

Physics Informed Neural Networks (PINNs)



- $\mathcal{L}$ depends on both the outputs (u, v, w, p) as well as deviations from physical laws
- Use collocation points for training
- Work well with limited data
- Can be quickly evaluated

The Single Machine Infinite Bus Model

Misyris, G. S., Venzke, A., & Chatzivasileiadis, S. *Physics-informed neural networks for power systems*. IEEE power & energy society general meeting, 2020.

Swing Equation: $f_\delta(t, P) = M\frac{\partial^2\delta}{\delta t^2} + D\frac{\partial\delta}{\partial t} + BV_g V_e \sin(\delta) - P = 0$

δ	...	rotor angle
M	...	generator inertia constant
D	...	dampening constant
B	...	susceptance between generator and grid
V_g, V_e	...	grid and bus voltage magnitudes
P	...	mechanical power at the generator

Aim: Predict rotor angles after disturbances.

The Single Machine Infinite Bus Model

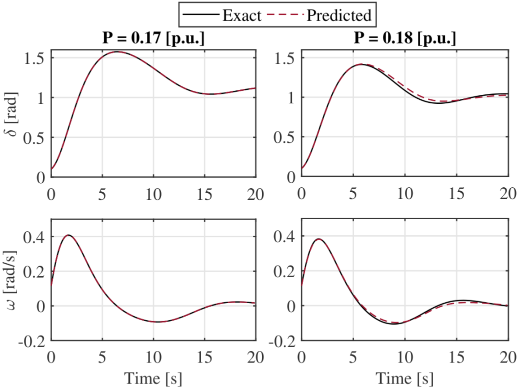
$$MSE = \underbrace{\frac{1}{N_u}\sum_{i}^{N_u} |u(t_u^i, x_u^i) - u^i|^2}_{MSE_u} + \underbrace{\frac{1}{N_f}\sum_{i}^{N_f} |f(t_f^i, x_f^i)|^2}_{MSE_f}$$

- N_u is the number of training data points for rotor angles
- N_f is the number of collocation points used for training
  - Used to assure compliance with swing equation
  - Arbitrarily sampled in spatio-temporal domain

Data: Simulate accurate 100 trajectories using *ode45* in 0.1s resolution. Setting $V_g = V_e = 1$ p.u. and $B = 0.2$ for T in $[0, 20]$

The Single Machine Infinite Bus Model

- Use $N_u = 40$ randomly sampled training points and $N_f = 8000$ collocation points.
- 5 layer neural network with 10 neurons per hidden layer
- $u = \delta$, $x = P$, ω by numeric differentiation
- Simulations
  - 28 times faster than with ODE solver
  - Small error
  - Need not be sampled sequentially

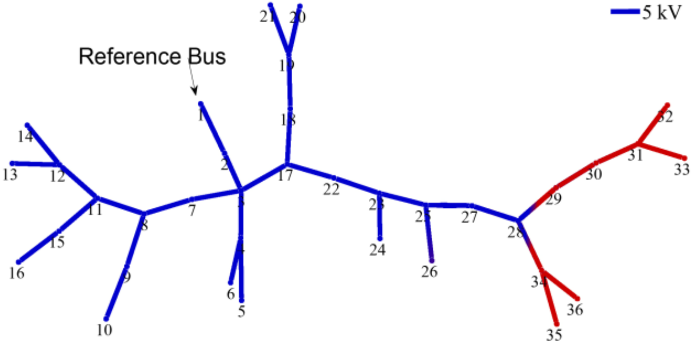


Estimate State of the Grid with Limited Measurements

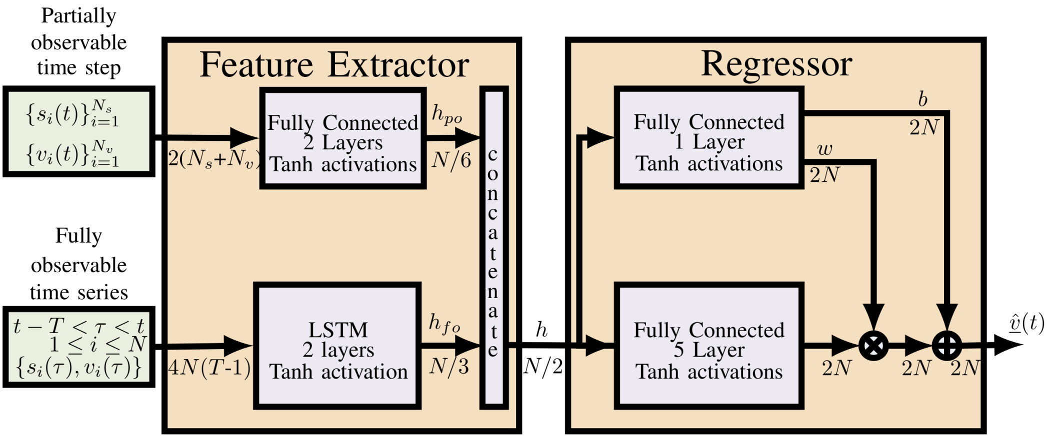
Ostrometzky, J., Berestizshevsky, K., Bernstein, A., & Zussman, G. *Physics-Informed Deep Neural Network Method for Limited Observability State Estimation*. 2020.

Input: Complete measurements up to point $t-1$, incomplete measurements at point t

Goal: Estimate state of the grid (voltages) at time t

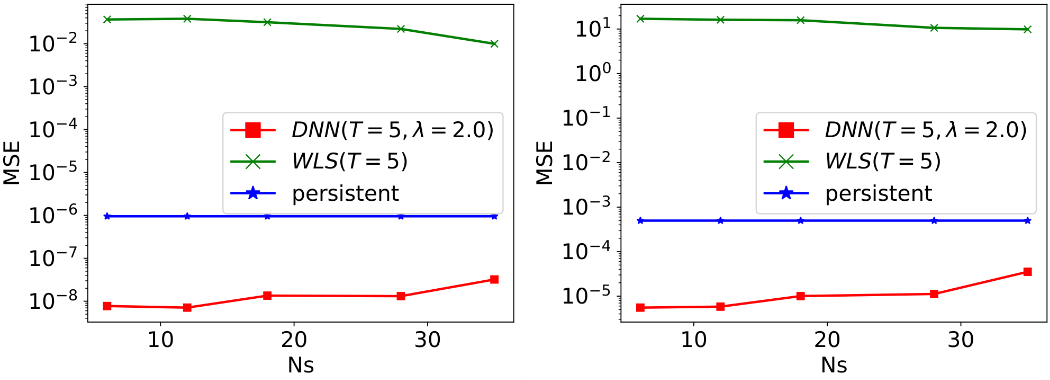


Estimate State of the Grid with Limited Measurements



$$\mathcal{L} = \mathcal{L}_{\text{data}} + \mathcal{L}_{\text{physics}} = \frac{1}{N}\sum_{i=1}^{N}\left(||v(x_i) - v^i||^2 + \lambda||\,\text{diag}(v(x_i))\,Y^* v(x_i)^* - s^i||^2\right)$$

Estimate State of the Grid with Limited Measurements



(a) Magnitude$\{\underline{v}(t)\}$

(b) Angle$\{\underline{v}(t)\}$

- PINN outperforms traditional methods

Outline

Discussion

1. What are the greatest challenges you face in your work?

2. Did you already work with Machine Learning?

3. What do you think Machine Learning could do for you?

4. What do you think Machine Learning cannot do?

david wozabal
d.wozabal@vu.nl