The Machine Learning ⇌ Science Colaboratory

Cluster of Excellence Machine Learning in Science, University of Tübingen
Seth Axen, Vladimir Starostin, Hanqi Zhou, and Álvaro Tejero-Cantero

Making ML accessible to scientists

research applications

Advice
We offer advice on your ML application idea or implementation and share helpful resources.

Guidance
We provide regular guidance to post-graduate researchers in applying machine learning to scientific problems.

Execution *
We develop an ML solution for a well-identified research gap for which there is data.

* We currently prioritize advice and guidance.

software and reproducibility

Ensuring best practices
We update ML code to adhere to best software practices.

Software stewardship
We provide fixed-term maintenance of select ML software.

Overcoming fragmentation
We upstream ML research code to popular software packages.

training and dissemination

Introductory courses
We teach domain scientists the main machine learning concepts and terminology they will need to formulate an ML project or discuss their problem with an ML expert.

Technical workshops
We train domain-science researchers in the practical use of the latest ML techniques as well as reproducibility, open science, and performance aspects of ML software engineering.

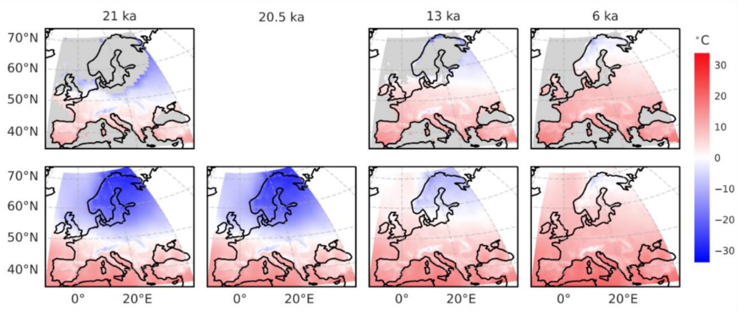
Learning materials and position pieces
We develop interactive learning materials and contribute our experience to community discussions.

Example projects

Paleoclimate reconstruction from fossil pollen and simulations

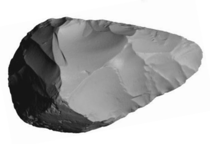
Domain #geosciences #climate science
ML #sparse spatiotemporal Gaussian processes
Collaborators: C. Sommer, N. Weitzel (Geosciences)

Build a consensus model of paleoclimate.



Characterization of lithic scars in Upper Paleolithic

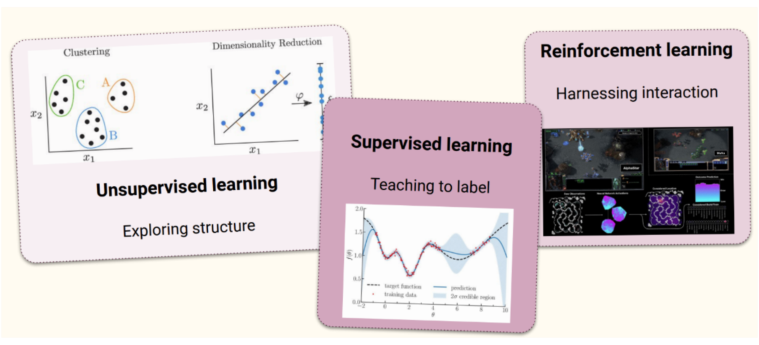
Domain #archeology
ML #GNNs #active learning
Collaborators: A. Falcucci (Geosciences)



Automate the analysis of lithic artifacts.

Introduction to machine learning

Runs twice a year
How does machine learning work, and what data, assumptions, and metrics do I need to think about for a successful application to my research?



Tracing neutrino paths from scintillation radiation

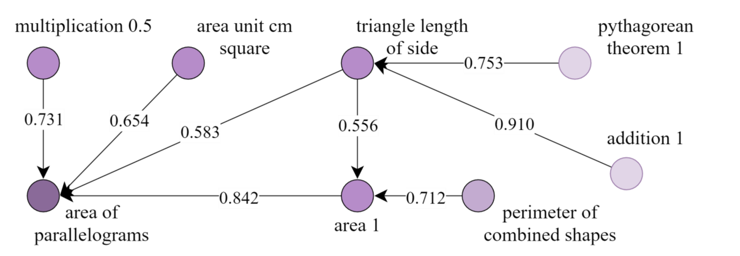
Domain #particle physics
ML #graph neural networks
Advised: L. Bieger, T. Laschenmaier (Physics)

Reconstruct neutrino trajectories from the light they generate, to evaluate the neutrino mass hierarchy.

Interpretable knowledge tracing for personalized education

Domain #education
ML #stochastic processes #variational inference #probabilistic modeling
Collaborators from education, computational linguistics and cognitive sciences.

Analyze learner cognitive traits and knowledge structure from online platform interaction data.



Extraction of concept dependencies from LLMs

Domain #education #data science
ML #data mining #LLM prompting
Student research project by D. Glandorf and A. Alekseeva (Computer science)

Map the relational structure of knowledge.

International workshop on simulation-based inference

mlcolab.org/resources/simulation-based-inference-for-scientific-discovery
How do I find parameters for my simulator that reproduce my data? We held a 3-day workshop focused on this question.

Subsurface flow and transport modeling

Domain #subsurface hydrology
ML #simulation-based inference
Advised: J. Allgeier, O. Cirpka (Geosciences)

Infer simulator parameters that reproduce hydraulic head data.

Advice across disciplines

Domains #archaeology #particle physics #philology #medicine #neuroscience #linguistics #sedimentology #paleoclimatology #transcriptomics
Advised: J. Lehmann (Romance studies), B. Starkovich, K. Fitzsimmons (Geosciences), J. Heyder (Art History), B. Goswami, K. Nieselt (Computer science), M. Inostroza (Medicine), A.-K. Schütz (Physics), F. Carcassi, M. Franke (Linguistics).

Time, space, and relations in Hadith literature

Domain #oriental and islamic studies
ML #NLP #LSTMs #transformers #semantic embeddings
Collaborators: M. Bednarkiewicz (Asian-Oriental Sciences)

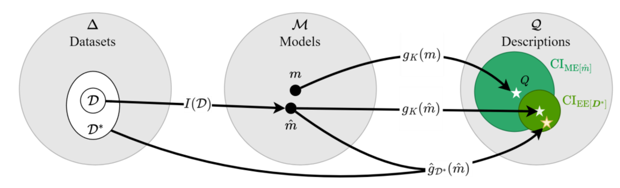

Image: British Library, public domain

Analyze chains of oral transmission to better understand islamic history.

Analyzing ML models to learn about real-world phenomena

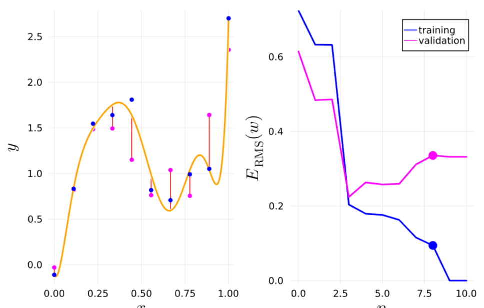
Domain #epistemology #interpretable machine learning
ML #statistical learning theory #interpretable ML
Collaborators: T. Freiesleben, G. König, C. Molnar (Computer science)

Determine if and how we can do science using a-priori weakly interpretable ML models.



Interactive notebook on supervised learning

mlcolab.github.io/IntroML.jl
Major concepts in supervised learning, from linear regression all the way to a feedforward neural network



Frequency effects in linear discriminative learning

Domain #computational linguistics
ML #linear algebra #statistics
Advised: M. Heitmeier, H. Baayen (Linguistics)

Account for word frequency in models of semantic learning.

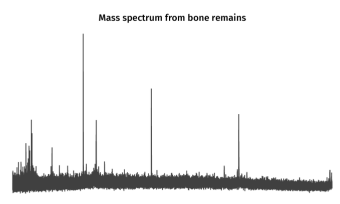
ML for classifying prime numbers

Domain #mathematics
ML #neural nets
Advised: R. Kahle (Philosophy)
Explore feasibility of classifying whether a number is prime.

Zooarchaeology by mass spectrometry

Domain #archaeology
ML #convolutional neural networks #interpretable ML
Collaborators: S. Brown (Geosciences), V. Borisov (Computer science)

Classify ancient animal samples into taxonomic groups using their mass spectra.



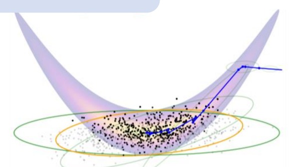
Climate drivers of leaf morphology

Domain #paleobotany
ML #logistic-binomial regression #Bayesian inference
Collaborators: C. Traiser, S. Roth-Nebelsick, J. Nebelsick (Geosciences)

Explore associations between fossilized leaf morphology and biodiversity.

Pathfinder.jl

Maintenance and development
Accelerating Bayesian model development and refinement



ArviZ

Maintenance and development
Exploratory analysis of Bayesian models.

CircStats.jl

Updating to follow best practices
Circular statistics in Julia.

ArviZ integration with SBI

Contributed feature
ArviZ diagnostics & visualizations for simulation-based inference.

Review of ML in Archaeology

Domain #archaeology #knowledge mapping
ML #ontologies #general ML
Advised: M. Bellat (Geosciences)

Survey the impact of ML in archaeology.

Seth Axen, ML research engineer
ML expertise: #Bayesian inference #probabilistic programming #Gaussian processes #automatic differentiation
Domain expertise: bioinformatics, structural biology

Hanqi Zhou, PhD student
ML expertise: #transfer learning (domain adaptation) #computer vision #3D scene understanding #adversarial learning
Domain expertise: cognitive sciences, personalized education

Vladimir Starostin, ML research engineer
ML expertise: #Bayesian inference #simulation-based inference #computer vision #object detection
Domain expertise: scattering physics, nonlinear optics, economics

Álvaro Tejero-Cantero, team lead
ML expertise: #time series #natural language #physical simulations #recommendation systems #graphs and networks
Domain expertise: physics, neuroscience

mlcolab@inf.uni-tuebingen.de | @mlcolab | fediscience.org/@mlcolab | Maria-von-Linden-str.6, Tübingen | https://mlcolab.org