Abstracts

Timo Freiesleben

What can post-hoc explanation methods tell us about our data?

In recent years, explanation methods such as SHAP, LIME, and partial dependence plots have gained popularity for interpreting trained machine learning models. But what if we applied these methods not to the models themselves, but directly to the data-generating process? Would this even be meaningful? In this talk, I argue that post-hoc explanation methods can, in some cases, be understood as estimators of certain aspects of the joint data distribution.

Rabanus Derr

Four Facets of Forecast Felicity - Revisited

Machine learning is about forecasting. Preferably, those forecasts are good. But what is a “good” forecast? Machine learning scholars have given answers of limited scope, mainly focusing on loss functions and calibration. Through the lens of game-theoretic probability and statistics I argue that a large part of prediction evaluation in machine learning can be subsumed under gambling with “available” gambles against the forecaster. Available gambles are gambles which the forecaster expects to incur a loss to the gambler. This perspective does not only show that calibration and regret expressed through loss functions are equivalently expressive. In addition, it provides an understanding of good predictions on outcomes as random outcomes on predictions.

Julian Rodemann

Towards Reciprocal Learning Theory: Generalizing From Self-Selected Samples

Reciprocal learning (as introduced at the 1st LuWSI Workshop) generalizes several learning paradigms, ranging from active learning over multi-armed bandits to self-training. These methods not only learn parameters from data, but also vice versa: They iteratively alter the training data as a function of previously learned parameters. In my talk at the 2nd LuWSI Workshop, I will address the elephant in the room: How well can these algorithms generalize from such self-selected samples?

Moritz Haas

Beyond μP: Scaling Insights from Infinite-width Theory for Non-standard Architectures and Learning Paradigms

Scaling is pivotal to the success of modern machine learning. However, upscaling model size also introduces new challenges, such as increased training instability. In this talk, I will discuss how infinite-width theory can be utilized to establish optimal scaling rules across architectures and learning paradigms. I will begin by discussing the scaling behaviour of standard neural networks trained with Sharpness-Aware Minimization—a min-max learning formulation observed to enhance generalization. As a second example, I will discuss the scaling behaviour of structured state space models (SSMs), which have emerged as efficient alternatives to transformers. Owing to the unique structure of their transition matrices, SSMs defy conventional scaling analyses and necessitate specialized approaches. Both examples show that achieving width-independent training dynamics and transfer of optimal hyperparameters across model scales is possible beyond the standard architectures and optimization algorithms, but they also highlight the need for specialized scaling strategies beyond standard μP.

Georg Schollmeyer

Size vs Capacity (or better Capacity vs Diversity?): On some subtle issues in highly nonparametric resampling-based tests that make use of VC-guarantees for regularization

In the proof of the sufficient condition for the uniform convergence of relative frequencies of events to their probabilities, Vapnik and Chervonenkis made use of the union bound applied to a projected finite family of sets. This is one of three main ideas in the proof of sufficiency, which can be seen as one cornerstone in the statistical analysis of empirical risk minimization.

One important aspect here is the cardinality of the projected family, which is growing polynomially under the sufficient condition, namely a finite VC dimension. More concretely: The VC dimension controls the growth function which itself controls the size of the projected family, which grows polynomially in the benign case, making uniform convergence establishable.

However, the union bound leaves a large possible gap. Concerning the VC dimension and (beyond) its role for the indirect control of the size of the projected family, in an interview with Lex Fridman 2018, Vapnik admitted that they '... called it capacity, but maybe it [is] better called diversity…'.

This talk is devoted to elaborating on the difference between the size and the diversity (i.e., the VC dimension) of a (finite) family of sets within the narrow bounds that relate both quantities within the mystical sandwich of the combinatorial factuality of the Sauer-Shelah lemma (which might have been better called Vapnik-Chervonenkis lemma?).

I will try to give some hints about possible implications of this difference for statistical hypothesis testing in high-dimensional situations, particularly I will discuss permutation tests and the bootstrap.

Rajeev Verma

On Calibration in Multi-Distribution Learning

Abstract: Modern challenges of robustness, fairness, and decision-making in machine learning have led to the formulation of multi-distribution learning (MDL) frameworks in which a predictor is optimized across multiple distributions. We study the calibration properties of MDL to better understand how the predictor performs uniformly across the multiple distributions. Through classical results on decomposing proper scoring losses, we first derive the Bayes optimal rule for MDL, demonstrating that it maximizes the generalized entropy of the associated loss function. Our analysis reveals that while this approach ensures minimal worst-case loss, it can lead to non-uniform calibration errors across the multiple distributions and there is an inherent calibration-refinement trade-off, even at Bayes optimality. Our results highlight a critical limitation: despite the promise of MDL, one must use caution when designing predictors tailored to multiple distributions so as to minimize disparity.

In this talk I'll first briefly review MDL, especially the learning-theoretic foundations of it, and then will talk about the calibration and refinement, ending the presentation with some open questions.

Bob Williamson

The Rhetorics of ML
TBA

Polina Gordienko

Consensus in Motion: A Case of Dynamic Rationality of Sequential Learning in Probability Aggregation

We propose a framework for probability aggregation based on propositional probability logic. Unlike conventional judgment aggregation, which focuses on static rationality, our model addresses dynamic rationality by ensuring collective beliefs update consistently with new information. We provide sufficient conditions for fair learning, where individuals agree on a shared subset of propositions (common ground), ensuring Bayesian updates—before or after aggregation—yield the same belief.

Christoph Jansen

Empirical Decision Problems

Analyzing decision problems (DPs) often relies on idealizing assumptions on the describability of the world. Many approaches fail if the states of the world are inaccessible. We avoid states as a primitive by addressing DPs empirically: we assume our DP to consist only of a protocol of observed act-consequence pairs. We show how optimality in empirical DPs can be addressed by protocol-based choice functions and discuss three approaches for deriving inferential guarantees for their choice sets.

Sebastian Zezulka

Prediction, Potential Outcomes, and Performativity

When predictions inform algorithmic policies, they are not just forecasts but have a causal impact on outcomes. Performativity entangles pragmatic and epistemic issues, making accuracy ambiguous. Proposals to address performativity, such as “endogenizing” performative effects or steering outcomes, are misguided. By distinguishing actual and counterfactual predictions, pragmatic and epistemic issues can be separated, enabling machine learning models to be decision- and discourse-supportive.

Laura Iacovissi

Data Processing Inequalities with Constrained Model Class

The classical Data Processing Inequality (DPI) for mutual information and Shannon entropy imposes a fundamental constraint on information propagation in learning systems. This result has been extended to Bayes Risk (BR) and f-divergences — interpreted as generalized entropy and information of a statistical experiment — in the unconstrained model class setting.

However, modern machine learning models, particularly those utilizing deep representation learning, appear to violate this theoretical framework. The most popular explanation for this phenomenon is that the inequality actually holds as an equality under BR, and learned representations aid convergence to an optimal solution, mitigating data limitations.

A less explored, largely informal perspective suggests that these violations arise due to the constrained nature of the model class. We align with this view, arguing that deep models, despite their large capacity, operate within a finite hypothesis space shaped by architecture, data availability, and the learning paradigm (e.g., unsupervised, semi-supervised, self-supervised). This induces an *effective model class*, fundamentally altering the applicability of DPI.

I will present our approach to formalizing the Constrained DPI, along with preliminary results on its validity. This is a work in progress, carried out jointly with Rabanus Derr and Bob Williamson.

Ana-Andreea Stoica

The Fairness-Quality Trade-off in Clustering

In this talk, we introduce novel algorithms for tracing the complete trade-off curve, or Pareto front, between quality and fairness in clustering problems. Unlike previous work that deals with specific objectives for quality and fairness, we deal with general classes encompassing most of the special cases addressed in previous work. This shows the challenges in learning with weakly structured data and opens up new questions for multi-objective optimization in the context of unsupervised learning.

Gerrit Bauch

Correlation Uncertainty

Research frequently tackles the understanding of complex situations by investigating their constituent aspects, e.g., climate sensitivity and albedo in climate change. But what uncertainty remains if these aspects might be correlated? This talk provides a decision-theoretic foundation for aggregating statistical evidence from subspaces. We characterize the set of possible correlations between subspaces as a convex polytope and identify its extreme as the local maxima of mutual information.

Mohammad Amin Charusaie

The Fundamental Lemma of Neyman and Pearson: Generalizations and Applications in Multi-Objective Learning

The Neyman-Pearson lemma was originally introduced to identify the uniformly most powerful statistical test in a binary hypothesis testing setting—i.e., a test that minimizes the false negative error rate while keeping the false positive error rate bounded. In this talk, we discuss various generalizations of this lemma to multi-hypothesis and cost-sensitive settings. We further demonstrate how these generalizations enable the recovery of the Bayes-optimal solution in a range of multi-objective learning problems, including fair multiclass classification, selective classification, and out-of-distribution detection.

Nan Lu

A Problem Reduction Perspective on Machine Learning under Data Corruption

Problem reductions provide a way to transfer theory and algorithms across different machine learning problems. In this talk, we will explore how several machine learning challenges can be reduced to the classical cost-sensitive classification framework, with examples including learning under data corruption, fairness-aware classification, and classification with rejections. We will discuss the implications of these reductions, considering how they help us understand the difficulty and relationships between the original problems. Additionally, we will raise open questions about which problems are unrelated in the context of these reductions, when such reductions are possible, and when they fail.