UU Stats and ML Journal Club

This is the Machine learning and Statistics Journal Club. We gather roughly every third week.

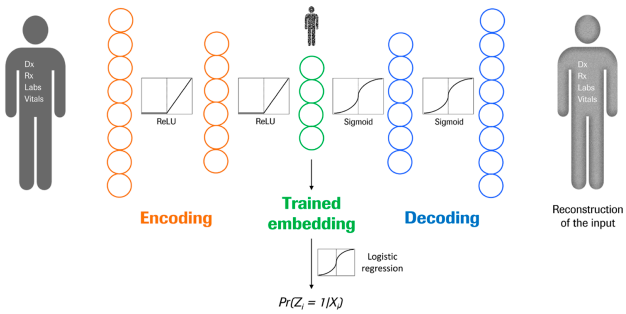
Upcoming meetings

2024-02-22 – Agnostic Bayesian Learning of Ensembles (Lacoste et al. 2014)

We propose a method for producing ensembles of predictors based on holdout estimations of their generalization performances. This approach uses a prior directly on the performance of predictors taken from a finite set of candidates and attempts to infer which one is best. Using Bayesian inference, we can thus obtain a posterior that represents our uncertainty about that choice and construct a weighted ensemble of predictors accordingly. This approach has the advantage of not requiring that the predictors be probabilistic themselves, can deal with arbitrary measures of performance and does not assume that the data was actually generated from any of the predictors in the ensemble. Since the problem of finding the best (as opposed to the true) predictor among a class is known as agnostic PAC-learning, we refer to our method as agnostic Bayesian learning. We also propose a method to address the case where the performance estimate is obtained from k-fold cross validation. While being efficient and easily adjustable to any loss function, our experiments confirm that the agnostic Bayes approach is state of the art compared to common baselines such as model selection based on k-fold cross-validation or a linear combination of predictor outputs.

Past meetings

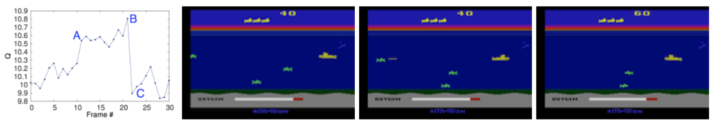
2024-01-18 – Deep Learning-based Propensity Scores (Weberpals et al. 2021)



Due to the non-randomized nature of real-world data, prognostic factors need to be balanced, which is often done by propensity scores (PSs). This study aimed to investigate whether autoencoders, which are unsupervised deep learning architectures, might be leveraged to compute PS.

Presenter: Chamika Porage

2023-12-07 – Playing Atari with Deep Reinforcement Learning (Mnih et al. 2013)



We present the first deep learning model to successfully learn control policies directly from high-dimensional sensory input using reinforcement learning. The model is a convolutional neural network, trained with a variant of Q-learning, whose input is raw pixels and whose output is a value function estimating future rewards. We apply our method to seven Atari 2600 games from the Arcade Learning Environment, with no adjustment of the architecture or learning algorithm. We find that it outperforms all previous approaches on six of the games and surpasses a human expert on three of them.

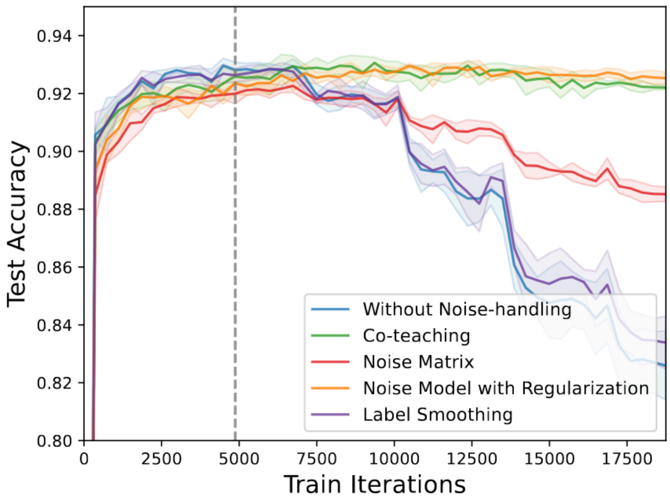
Presenter: Andreas Östling

2023-11-09 – Liquid Time-constant Networks (Hasani et al. 2021)

We introduce a new class of time-continuous recurrent neural network models. Instead of declaring a learning system's dynamics by implicit nonlinearities, we construct networks of linear first-order dynamical systems modulated via nonlinear interlinked gates. The resulting models represent dynamical systems with varying (i.e., liquid) time-constants coupled to their hidden state, with outputs being computed by numerical differential equation solvers. These neural networks exhibit stable and bounded behavior, yield superior expressivity within the family of neural ordinary differential equations, and give rise to improved performance on time-series prediction tasks. To demonstrate these properties, we first take a theoretical approach to find bounds over their dynamics and compute their expressive power by the trajectory length measure in latent trajectory space. We then conduct a series of time-series prediction experiments to manifest the approximation capability of Liquid Time-Constant Networks (LTCs) compared to classical and modern RNNs.

Presenter: Jakob Torgander

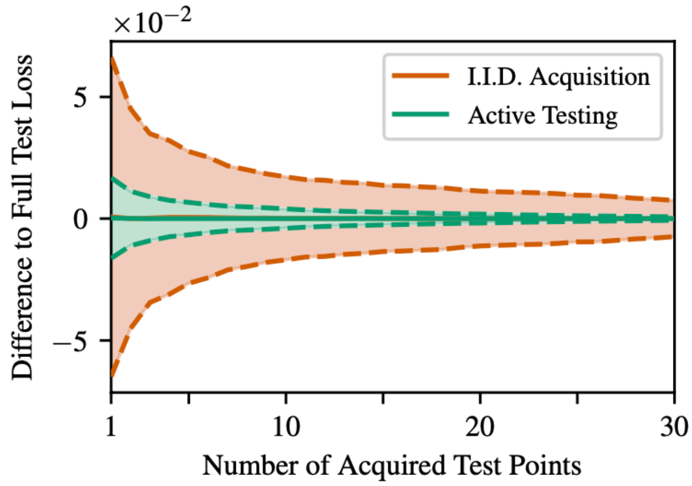
2023-10-12 – Is BERT Robust to Label Noise? (Zhu et al. 2022)



Abstract: Incorrect labels in training data occur when human annotators make mistakes or when the data is generated via weak or distant supervision. It has been shown that complex noise-handling techniques - by modeling, cleaning or filtering the noisy instances - are required to prevent models from fitting this label noise. However, we show in this work that, for text classification tasks with modern NLP models like BERT, over a variety of noise types, existing noisehandling methods do not always improve its performance, and may even deteriorate it, suggesting the need for further investigation. We also back our observations with a comprehensive analysis.

Presenter: Hannes Waldetoft

2023-09-14 – Active Testing: Sample–Efficient Model Evaluation (Kossen et al. 2021)



Abstract: We introduce a new framework for sample-efficient model evaluation that we call active testing. While approaches like active learning reduce the number of labels needed for model training, existing literature largely ignores the cost of labeling test data, typically unrealistically assuming large test sets for model evaluation. This creates a disconnect to real applications, where test labels are important and just as expensive, eg for optimizing hyperparameters. Active testing addresses this by carefully selecting the test points to label, ensuring model evaluation is sample-efficient. To this end, we derive theoretically-grounded and intuitive acquisition strategies that are specifically tailored to the goals of active testing, noting these are distinct to those of active learning. As actively selecting labels introduces a bias; we further show how to remove this bias while reducing the variance of the estimator at the same time. Active testing is easy to implement and can be applied to any supervised machine learning method. We demonstrate its effectiveness on models including WideResNets and Gaussian processes on datasets including Fashion-MNIST and CIFAR-100.

Presenter: Väinö Yrjänäinen

2023-06-08 – Using natural language and program abstractions to instill human inductive biases in machines (Kumar et al. 2022)

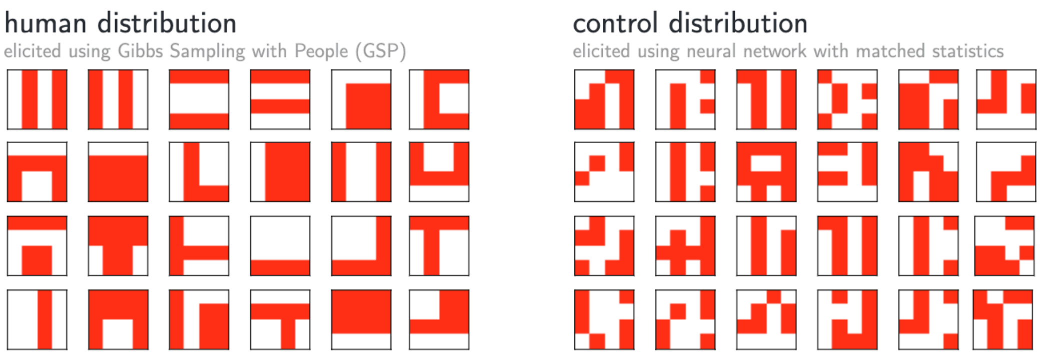


Figure 2: Example grids from human-elicited priors (left) and machine-generated priors (right).

Abstract: Strong inductive biases give humans the ability to quickly learn to perform a variety of tasks. Although meta-learning is a method to endow neural networks with useful inductive biases, agents trained by meta-learning may sometimes acquire very different strategies from humans. We show that co-training these agents on predicting representations from natural language task descriptions and programs induced to generate such tasks guides them toward more human-like

inductive biases. Human-generated language descriptions and program induction models that add new learned primitives both contain abstract concepts that can compress description length. Co-training on these representations result in more human-like behavior in downstream meta-reinforcement learning agents than less abstract controls (synthetic language descriptions, program induction without learned primitives), suggesting that the abstraction supported by these representations is key.

Presenter: Isac Boström

2023-05-04 – Forecasting the movements of Bitcoin prices: an application of machine learning algorithms (Pabuçcu, Ongan, and Ongan 2023)

Abstract: Cryptocurrencies, such as Bitcoin, are one of the most controversial and complex technological innovations in today's financial system. This study aims to forecast the movements of Bitcoin prices at a high degree of accuracy. To this aim, four different Machine Learning (ML) algorithms are applied, namely, the Support Vector Machines (SVM), the Artificial Neural Network (ANN), the Naive Bayes (NB) and the Random Forest (RF) besides the logistic regression (LR) as a benchmark model. In order to test these algorithms, besides existing continuous dataset, discrete dataset was also created and used. For the evaluations of algorithm performances, the F statistic, accuracy statistic, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Root Absolute Error (RAE) metrics were used. The t test was used to compare the performances of the SVM, ANN, NB and RF with the performance of the LR. Empirical findings reveal that, while the RF has the highest forecasting performance in the continuous dataset, the NB has the lowest. On the other hand, while the ANN has the highest and the NB the lowest performance in the discrete dataset. Furthermore, the discrete dataset improves the overall forecasting performance in all algorithms (models) estimated.

Presenter: Sahika Gökmen

Bibliography

Hasani, Ramin, Mathias Lechner, Alexander Amini, Daniela Rus, and Radu Grosu. 2021. "Liquid Time-Constant Networks". In *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:7657–66

Kossen, Jannik, Sebastian Farquhar, Yarin Gal, and Tom Rainforth. 2021. "Active Testing: Sample-Efficient Model Evaluation". In *International Conference on Machine Learning*, 5753–63

Kumar, Sreejan, Carlos G Correa, Ishita Dasgupta, Raja Marjieh, Michael Y Hu, Robert Hawkins, Jonathan D Cohen, Karthik Narasimhan, Tom Griffiths, and others. 2022. "Using Natural Language and Program Abstractions to Instill Human Inductive Biases in Machines". *Advances in Neural Information Processing Systems* 35: 167–80

Lacoste, Alexandre, Mario Marchand, François Laviolette, and Hugo Larochelle. 2014. "Agnostic Bayesian Learning of Ensembles". In *International Conference on Machine Learning*, 611–09

Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. "Playing Atari with Deep Reinforcement Learning". *Arxiv Preprint Arxiv:1312.5602*

Pabuçcu, Hakan, Serdar Ongan, and Ayse Ongan. 2023. "Forecasting the Movements of Bitcoin Prices: An Application of Machine Learning Algorithms". *Arxiv Preprint Arxiv:2303.04642*

Weberpals, Janick, Tim Becker, Jessica Davies, Fabian Schmich, Dominik Rüttinger, Fabian J Theis, and Anna Bauer-Mehren. 2021. "Deep Learning-Based Propensity Scores for Confounding

Control in Comparative Effectiveness Research: A Large-Scale, Real-World Data Study". *Epidemiology* 32 (3): 378–88

Zhu, Dawei, Michael A Hedderich, Fangzhou Zhai, David Ifeoluwa Adelani, and Dietrich Klakow. 2022. "Is BERT Robust to Label Noise? A Study on Learning with Noisy Labels in Text Classification". *Arxiv Preprint Arxiv:2204.09371*