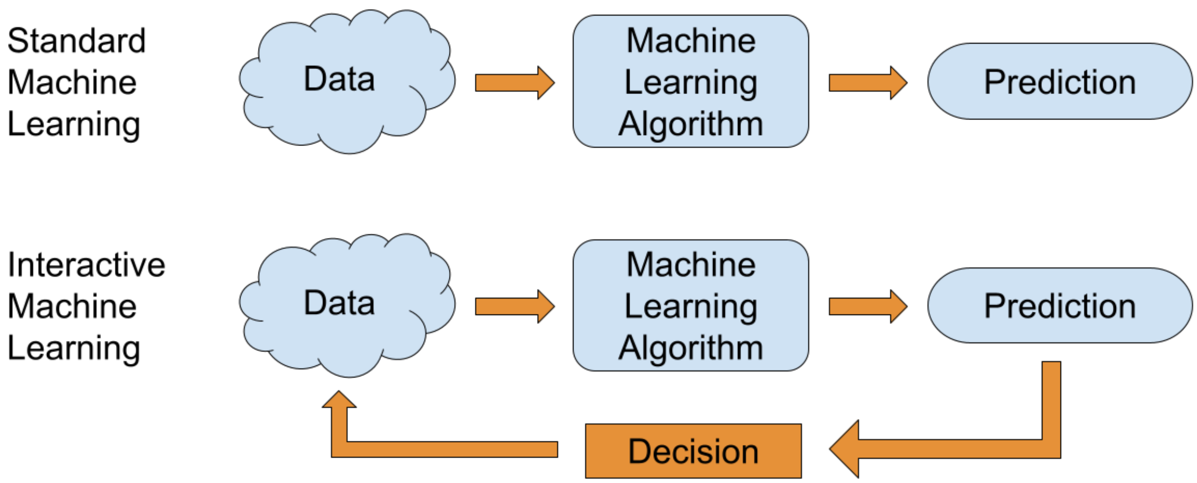
Mathematical Foundations of Interactive Machine Learning

Interactive machine learning (IML) is an emerging paradigm in the field of data science that integrates data collection with data analysis in an adaptive and iterative manner. In IML, inferences and models learned from previously collected data are used to optimize subsequent data selection, with the goal of maximizing the "informativeness" of these actions, leading to acquisition of new data that is most relevant to the modeling or inference task at hand. IML plays a critical role in situations where human expertise is needed to assign labels to data points or provide feedback, and when data is collected via costly scientific experiments. By using machine learning models to guide the solicitation of human inputs or the design of experiments, IML enables more accurate, efficient, and adaptable models that can handle the complexities and uncertainties of real-world scenarios. IFDS researchers are at the forefront of IML, developing new theory and methods to advance and optimize IML systems.



Interactive machine learning accelerates learning by focusing human effort and data collection resources where they will most aid prediction skill.

One significant IFDS contribution is to IML methods for neural network models (as opposed to simpler linear or kernel models used previously). Foundational ML theory has produced new methods capable of harnessing the representation power of neural networks even with limited labeled training data, making such models viable in a broad range of applications requiring human labeling of data and helping guide the design of new experiments.

Reinforcement learning is an important branch of IML in which autonomous agents learn and make decisions based on the feedback they receive from their environment. Most reinforcement learning algorithms are designed to minimize the time it takes to master the worst-case environment. This leads to unnecessarily slow convergence for typical or easy situations. To address this limitation, IFDS researchers have been developing innovative algorithms that can adapt to their *specific* environment, allowing agents to master easy environments faster than harder ones. These novel learning strategies have opened new possibilities for real-world applications, particularly in robotics, where adapting to the environment is crucial.

Publications.

[1] Y. Chen, K. Jamieson, S. Du. Active Multi-Task Representation Learning. International Conference on Machine Learning, 2022.

[2] Romain Camilleri, Zhihan Xiong, Maryam Fazel, Lalit Jain, and Kevin Jamieson. Selective Sampling for Online Best-arm Identification, *NeurIPS 2021*. PDF

[3] Andrew Wagenmaker, Max Simchowitz, Kevin Jamieson. Beyond No Regret: Instance-Dependent PAC Reinforcement Learning, *COLT 2022*. PDF

[4] Andrew Wagenmaker and Kevin Jamieson. Instance-Dependent Near-Optimal Policy Identification in Linear MDPs via Online Experiment Design, *NeurIPS 2022*. PDF

[5] Yinglun Zhu and Robert D. Nowak. "Active Learning with Neural Networks: Insights from Nonparametric Statistics." *Advances in Neural Information Processing Systems*. PDF

[6] Subhojyoti Mukherjee, Josiah P. Hanna, and Robert D. Nowak. "ReVar: Strengthening policy evaluation via reduced variance sampling." *Uncertainty in Artificial Intelligence*. PMLR, 2022. PDF

[7] Jifan Zhang, Julian Katz-Samuels, and Robert Nowak. "GALAXY: graph-based active learning at the extreme." *International Conference on Machine Learning*. PMLR, 2022. PDF

[8] Jifan Zhang, Yifang Chen, Gregory Canal, Stephen Mussmann, Yinglun Zhu, Simon Shaolei Du, Kevin Jamieson, Robert D Nowak, "LabelBench: A Comprehensive Framework for Benchmarking Label-Efficient Learning," arXiv:2306.09910.