Research Statement

The goal of my research is to develop a better understanding of how and when machine learning succeeds and the broader societal implications of data-driven systems. My work lies at the intersection of machine learning, optimization, and statistics, including topics such as high dimensional learning and algorithmic explainability.

The remarkable and continuing empirical success of modern machine learning, especially deep learning, has opened up several foundational questions about our understanding of why learning models succeed. Answering these questions have the potential to expand the theoretical and algorithmic frontiers of learning and optimization. At the same time, as machine learning proliferates into the real world, it is of profound importance to critically evaluate the broad societal impacts of these systems. I am interested in developing a theoretical understanding of learning models that will lead to principled methods for building systems with a positive social impact.

My recent research spans two lines of work that contribute to these broad themes. My primary line of work develops new theories towards understanding the role of optimization in the success of modern machine learning models, especially deep neural networks. My second line of work is on formulating algorithms that aid from theoretical understanding. Hence paves the way for more theoretically grounded techniques that could potentially help solve the barriers of current deep learning algorithms, to name a few - Generalization and Catastrophic forgetting.

An open-ended research question that I tried addressing recently is the problem of catastrophic forgetting in Continual learning algorithms and deep learning in general. Catastrophic forgetting is the problem of the models forgetting the previously trained data, which is very unlikely of a human brain. In other words, human brains are very good at effectively forgetting unnecessary data. As straight forward it may seem, it is not very easy to translate to machine learning models, where catastrophic forgetting can be extremely severe to a point it almost forgets everything.

Few works that I found interesting recently -

- Emergent properties of the local geometry of neural loss landscapes - Stanislav Fort
- Large Scale Structure of Neural Network Loss Landscapes - Stanislav Fort
- Kernel and Deep Regimes in Overparametrized Models - Suriya Gunashekar
- Loss Surfaces, Mode Connectivity, and Fast Ensembling of DNNs - Timur Garipov
- Fluctuation-dissipation relations for stochastic gradient descent - Sho Yaida
- The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks - Jonathan Frankle
- Deconstructing Lottery Tickets: Zeros, Signs, and the Supermask - Jason Yosinski
- LCA: Loss Change Allocation for Neural Network Training - Jason Yosinski
- Reconciling modern machine learning practice and the bias-variance trade-off - Mikhail Belkin