New Frontiers for Discrete problems in Continuous Optimization

Project Summary

Overview

In the recent years, we have seen an exponential rise in the amount of data available for predictive tasks. Concurrently, there has been an almost equally explosive growth in available technologies for building larger, more complicated models in a single machine or in distributed setups. As such, there is an inherent need of selecting 'good' subsets from these large datasets, as well as from large models for speeding up training, load balancing and interpretability, as well as 'bad' subsets that should be removed or deferred during learning because of distribution shift or concerns about privacy-leakage. With these goals, in this proposal we aim to propose theoretically grounded algorithms for efficiently selecting *representative* subsets that can provide tangible engineering improvements with little degradation in predictive performance.

At the heart of these subset selection algorithms is novel theoretical studies that draws from the fields of discrete optimization and continuous optimization. Traditionally, these have largely evolved as disconnected silos with a few intermittent overlaps dubbed as sparse optimization. Most algorithms in sparse optimization can be broadly grouped under two umbrellas – convexification of the sparsity constraint, or use of thresholding operators on gradient descent steps to ensure sparsity. In this work, we intend to leverage techniques and results from discrete optimization studies to aid in development of new algorithms, analyses and applications for sparse continuous optimization.

For applications of these novel theoretical improvements, we plan to study selection of data and communication coresets in much more general settings than what has been considered before. These will lead to training speedups for single machine models and lower communication costs for distributed settings while ensuring a fairly balanced load. For large already trained models, the goal would be to identify meaningful subsets to interpret and understand the model's learning capabilities to sparsify the model with little degradation in predictive power. We seek to use interpretability-based techniques to identify elements of data and/or the (distributed) model that should be dropped for susceptibility of distribution shifts and privacy attacks.

Intellectual Merit

The project will develop new fundamental algorithms and analyses for sparse optimization by creating novel synergies in the fields of discrete and continuous optimization, drawing inspirations and advancing both the fields. Specifically, we will develop new provably convergent algorithms for the combinatorial support selection problem. We will leverage these findings to develop new techniques for faster and fair training through generalized data and communication coresets. For black-box interpretability, we will apply these methods for model compression and data-drop techniques.

Broader Impacts Of The Proposed Work

The PI intends to integrate the study of sparsity in machine learning into current and future course offerings at Purdue. Through various diversity promoting initiatives at Purdue, the PI will disseminate and encourage increased participation from underrepresented minorities.