Convex Hull-Based Ensemble Pruning for Computational Geometry course:

Researcher: Omid Feizi

Project Description:

This project is interested in combining geometry with ensemble machine learning to improve classifier performance by efficiently optimizing the process of eliminating redundant classifiers in ensemble methods towards a more efficient final decision making outcome through the use of the concept of convex hull—a fundamental construct of computational geometry—for testing and enhancing diversity, among ensemble classifiers

By projecting the voting results of each classifier into a space referred to as Euclidean space and then determining the convex figure around these points for identifying the classifiers indicating distinct and valuable information segments separately from others within this figure helps in eliminating redundant classifiers that lie within this figure boundary and therefore optimizing computation without compromising prediction accuracy or even potentially improving it. The project seeks to create algorithms which are able to accurately calculate the hull of vote vectors of classifiers and dynamically update ensembles through the use of measures of geometric diversity with efficiency being a priority area of concentration.

At the foundation of this project is the employment of current convex hull algorithms like Quickhull and mixing them with ensemble learning models together with the design of visual aids for demonstrating the vote space dynamics and convex boundaries in live pruning processes. In assessing the efficiency and effectiveness gains brought by convex hull based pruning on classification accuracy and speed in processing on various datasets will be instrumental measures, while in the course of measuring this project. Ultimately, the ultimate aim of this project is to produce a tool which will enhance the knowledge of ensemble learning and computational geometry by integrating it into real time data stream and resource scarce environments where high accuracy classification becomes the priority.

The project will be tested using known Python-based ensemble learning libraries such as Scikit-learn for model training and testing. Additional Python libraries such as NumPy and SciPy (for convex hull computation using scipy.spatial.ConvexHull) and Matplotlib for interactive plots will be used throughout the project.

1. Bonab, H. R., & Can, F. (2019). Less Is More: A Comprehensive Framework for the Number of Components of Ensemble Classifiers. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9), 2735-2745.
2. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
3. Bifet, A., & Gavaldà, R. (2009). Adaptive learning from evolving data streams. *Advances in Intelligent Data Analysis VIII*, 249-260.
4. Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4), 469-483.
5. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90-95.