Abbas Tajik 04.11.2025

Activity – Explore Ensemble Methods

Week 5 - Ensemble Methods and Hyperparameter Tuning

Instructions:

- *With your choice of dataset, either one that you have explored already or one that's new to you, apply one or more of the ensemble methods that we have introduced in this session.*
- *Compare your ensemble model(s) to the other types of models that you have been using prior to this session. Do you see an immediate benefit (e.g., when evaluating on the test data) or do you maybe need to experiment with the hyperparameters before seeing a benefit?*

Introduction:

For this task I chose to work on a new dataset, which had piqued my interest through previous research. The '**Robot Execution Failures**' dataset found on the UC Irvine Machine Learning Repository website was donated on the 22nd of April 1999. This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterised by 15 force/torque samples collected at regular time intervals. The purpose of this dataset is to predict the impending failure for a lab robot arm.

Although this dataset originated from the late 90's, when robotics had already been implemented and deployed since the very early 60's in various automotive and CNC-based manufacturer industries (such as the use of Ford's rudimentary *Unimate* robot system developed by George Devol in the early 50's), as well as their more advanced cousins in the 80's, I would argue that the information found here is anything but outdated.

Reflection:

I selected this dataset because it contains multiple classification tasks (LP1 through LP5) with force and torque sensor data, making it suitable for exploring ensemble methods. The dataset has 6 features which provides enough complexity to see differences between models.

Models Compared:

- Baseline models: Logistic Regression and Decision Tree.
- Ensemble methods: Random Forest (bagging), Gradient Boosting, and AdaBoost (boosting).
- Stacking: Combined predictions from all ensemble models using a logistic regression meta-learner.

Results:

The ensemble methods generally outperformed the baseline models. Random Forest and Gradient Boosting showed strong performance, with Gradient Boosting achieving higher accuracy than the baseline models. Stacking provided significant improvement by combining the strengths of different ensemble approaches.

Hyperparameter Tuning:

I performed grid search cross-validation on both Random Forest and Gradient Boosting. The tuning improved performance compared to the default parameters, particularly for Random Forest where adjusting the number of estimators, max depth, and min samples split led to noticeable gains. The tuned models typically achieved better overall accuracy than their untuned counterparts.

Observations:

1. Tree-based models (Decision Tree, Random Forest, Gradient Boosting) handled the non-linear relationships in the sensor data better than Logistic Regression.
2. Ensemble methods reduced overfitting and improved generalisation compared to a single Decision Tree.
3. Gradient Boosting's sequential approach was particularly effective for this dataset.
4. Hyperparameter tuning is important - the default parameters don't always provide optimal performance.
5. Stacking can provide significant improvements by blending the predictions of diverse base models, demonstrating the power of combining multiple learning algorithms.