**Gradient Boosting and Ada boosting, a comparative study of machine learning Techniques**

A progress report submitted for Info7017 Postgraduate Project B in partial fulfilment of the requirements for the degree of Master of Data Science

Supervisor: Dr Laurence Park

School of Computer, Data and Mathematical Sciences

Western Sydney University

April 2025

**Introduction**

Over the decades as technology advances, traditional programming has shown that they cannot handle modern datasets whether it be due to sheer variety, volume or complexity. This led to the development of machine learning by necessity, since then machine learning has become invaluable in various fields such as cloud computing and e-commerce, with every modern business and organizations participating in its use or development in some manner with continued technological advancements in computational power meaning it will stay relevant for decades to come.

Initially machine learning started off as a method to train artificial intelligence before splitting off into its own field and gained prominence with the rise of the internet, allowing it both to gain access to digital data and to spread its use across the world.

Unlike rigid rule-based systems there is room in machine learning models to continuously improve itself as they interact with more data, allowing them to become more accurate and effective over time, allowing them to take on problems that are too difficult to program explicitly.

Of course, machine learning itself is just a broad term to describe several self-learning methods that attempt to categorize, study and predict data such as decision trees, support vector machines and neural networks etc. While effective for a time, they were just the precursors to future techniques known as ensemble methods which aim to combine the strength of multiple simpler models to improve performance and accuracy. These developments in particular AdaBoost and Gradient Boost which will be covered in this report, decisively proved that machine learning models are outperforming traditional methods in classification and regression tasks. This is achieved by building models sequentially, where each new model attempts to correct errors made by previous models, leading to a powerful final model.

The first model covered in our report, AdaBoost can be traced back to the early 1990s, AdaBoost was one of the first algorithms to demonstrate that weak learners (models slightly better than random guessing) could be combined into a powerful ensemble through a principled iterative procedure by building models sequentially, where each new model attempts to correct errors made by previous models, leading to a powerful final model (This method of combining multiple weak learners to create a single accurate model became known as boosting).

Adaboost had demonstrated a remarkable ability to reduce bias and variance while being relatively resistant to overfitting. Its relative simplicity and elegant formula garnered it great attention. A key feature of AdaBoost is how sample weights are updated to focus on hard-to-classify instances, showcasing a novel way of dealing with the limitations of weak classifiers.

Gradient Boosting, developed in the late 90s and early 2000s, extended the idea of boosting into the realm of numerical optimization. Instead of simply reweighting samples, Gradient Boosting interprets boosting as a gradient descent procedure in function space, aimed at minimizing a chosen loss function. This change extended its use into the realms of regression rather than just binary classification. Gradient Boost’s flexibility to choose different loss functions, paired with the strength of decision trees as base learners, made it a significant leap forward in the world of Data science

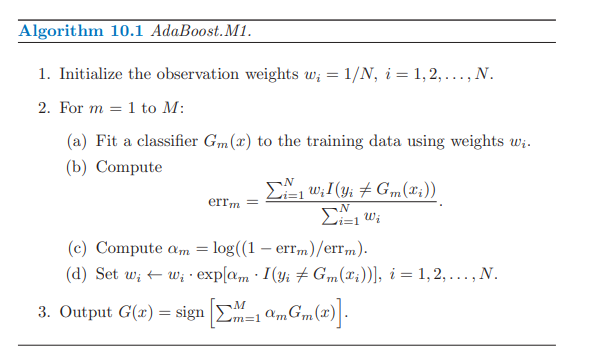
**Objective**

The primary objectives of this project are:

1. To explain the theoretical foundations and differences between AdaBoost and Gradient Boosting algorithms.
2. To implement both algorithms from scratch using Python and apply them to multiple tasks in classification and regression. For this report we will just be using the ‘wine’ dataset sourced from the UCI Machine Learning Repository
3. To analyze the performance of both models, including their accuracy, computational complexity, and ability to handle real-world data.
4. To contextualize both methods using Chapter 10 of *The Elements of Statistical Learning* and other sources.

**Methodology**

For this project to succeed, we must first gain a deeper understanding of algorithms Adaboost and Gradient boost relies on. As such we will present an in-depth examination of the two methods' theoretical foundations. As mentioned, we will heavily rely on Chapter 10 of *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Second Edition) by Hastie, Tibshirani, and Friedman.



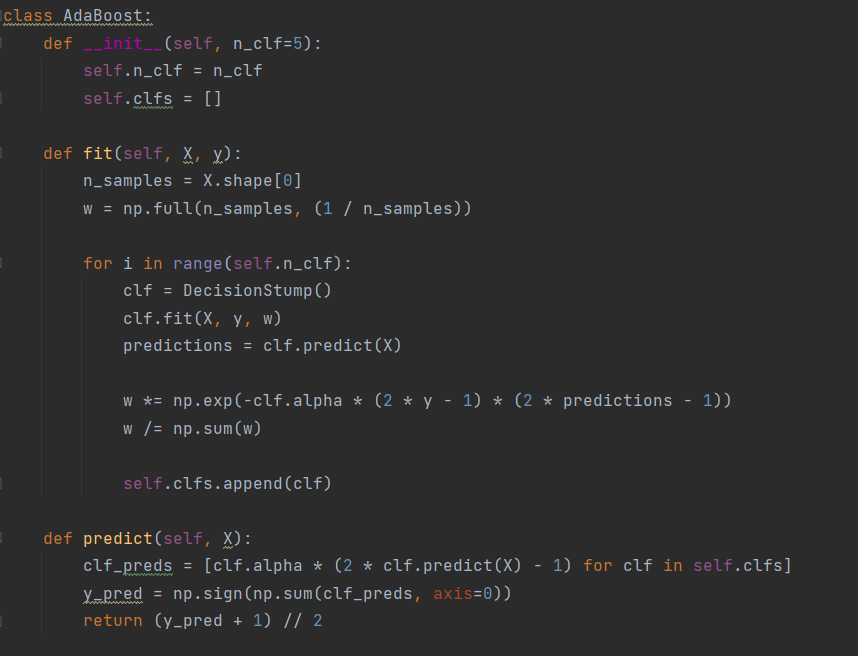
This algorithm can be found in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* it captures a faithful summary of how Adaboost functions.

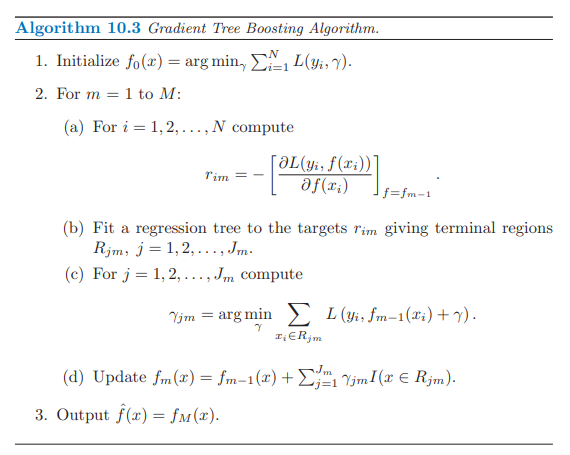
- We initialize each weak classifier with weight 1/N as a starting point

- We train a weak classifier with the goal to minimize weighted error, samples that hard harder to classify are assigned higher weights which cause them to influence the training more

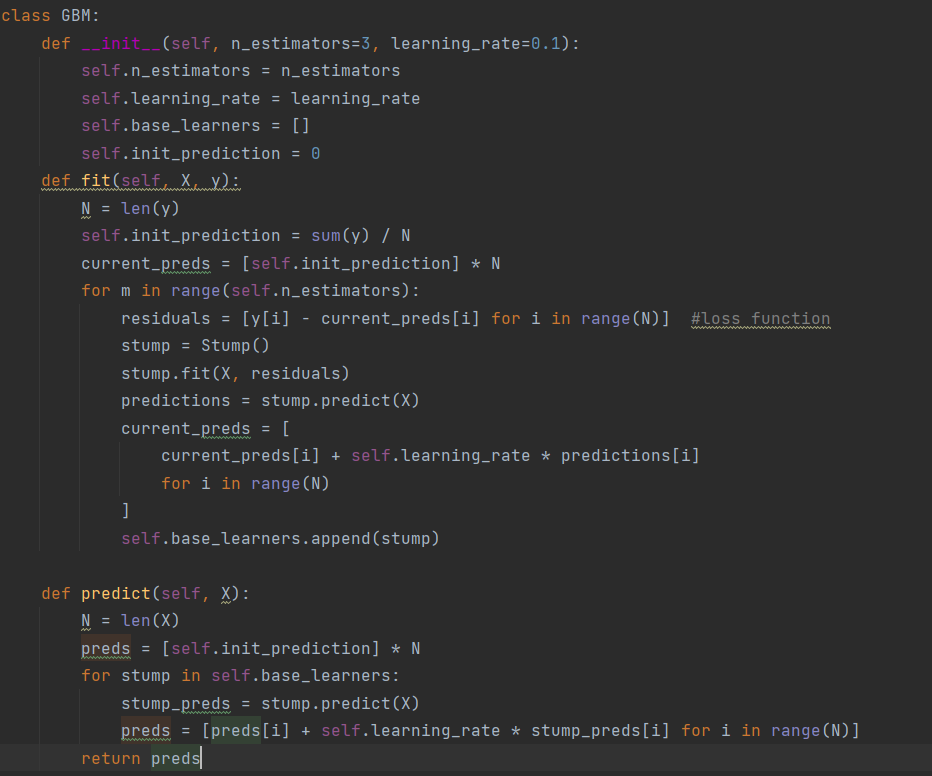
- We compute alpha, if there is no error, alpha approaches infinity (perfect classifier), if error is 0.5, alpha is 0 (random, flip of a coin basically) and higher erro means worse than random in the case of binary classification.

- We use alpha to assign weights to our classifiers so the worse the alpha for a classifier is, the more we focus on them





Code



**Sample Dataset**

For this report a simple dataset is chosen, The Wine dataset is a multivariate dataset that can be found in the UCI Machine Learning Repository and the sklearn dataset library. It is frequently used for benchmarking classification algorithms due to its moderate size, clean structure, and interpretability. The dataset contains the results of a chemical analysis of wines grown in the same region derived from three different varieties of grapes.

There are 178 samples and 13 features in the dataset such as alcohol content, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, color intensity, hue etc. The target variable is a categorical class label (0, 1, or 2) which shows what variety of grape was used to create the wine. To simplify the problem to binary classification, only classes 0 and 1 were retained, excluding all samples belonging to class 2.

**Results**

Adaboost: 0.9538461538461539

GradientBoost: 0.9230769230769231

**Conclusion**

The adoption of both algorithms accelerated in recent years with the availability of practical implementations and validation in competitions and real-world applications. For example, gradient boosting has led to further evolution in form of XGBoost, LightGBM, and CatBoost which consistently Kaggle competitions. Meanwhile, AdaBoost remains an instructive and efficient method for understanding ensemble behavior and is still used in certain low-noise, high-bias contexts.

Reference

[A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S002200009791504X) ((**Freund and Schapire** (1997))

[Greedy function approximation: A gradient boosting machine.](https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full) (**Friedman (2001)**,)

[A Brief History of Machine Learning - DATAVERSITY](https://www.dataversity.net/a-brief-history-of-machine-learning/)

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.