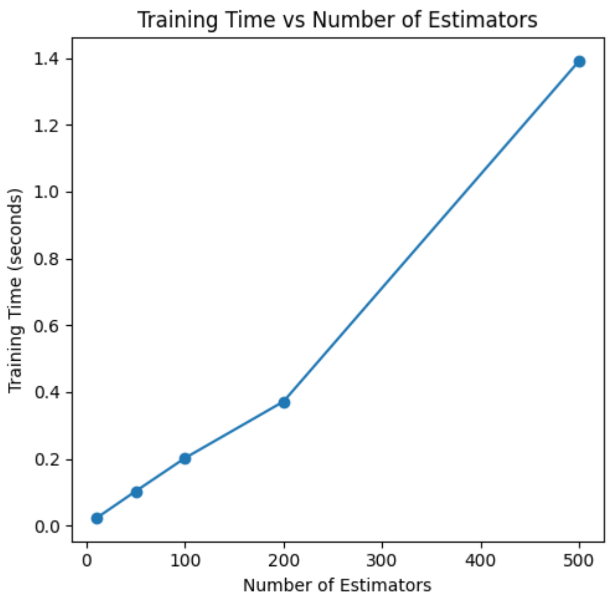
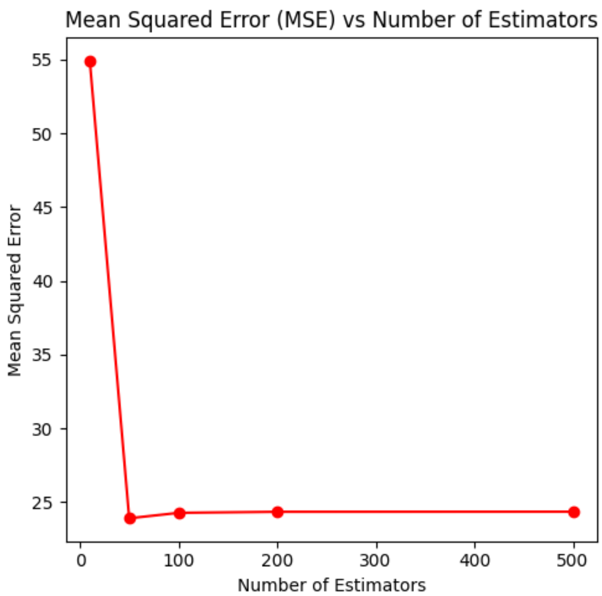
Report

Part 4:

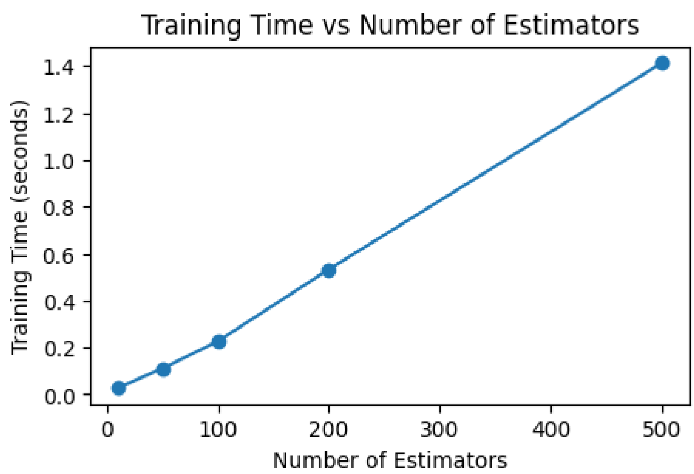
Graphs

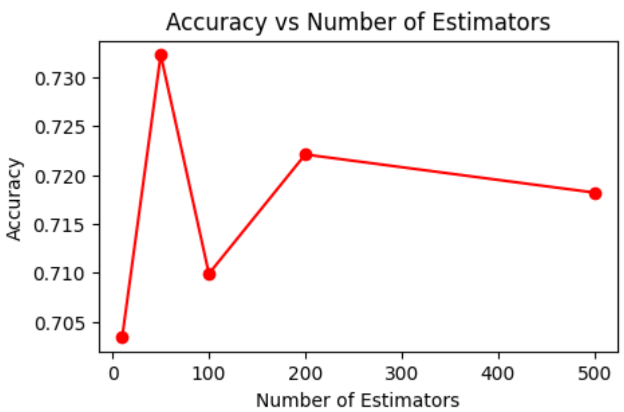
1. Gradient boosted regressors



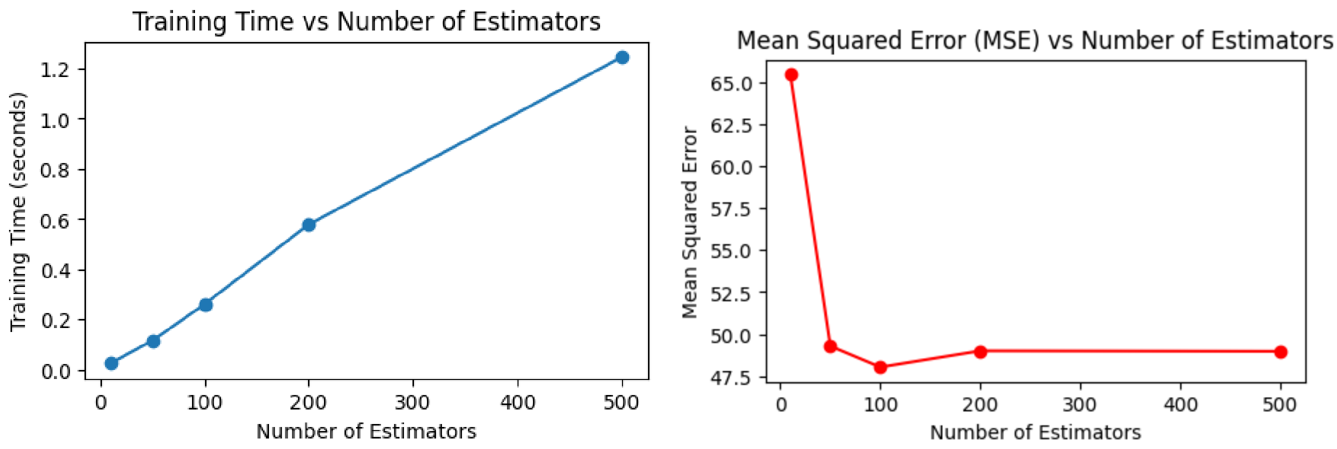


2. Adaboost classifier

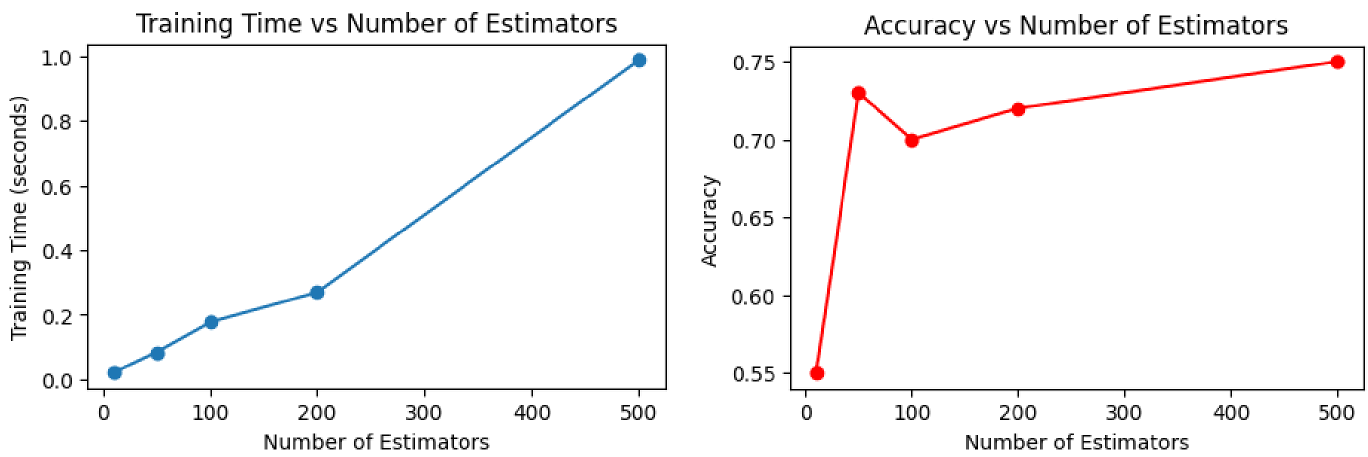




3. Adaboost Regressor



4. Gradient Boost Classifiers



Analysis

- ## Random Forest

While Random Forest models possess great power, they do have limitations. When there are too many trees, overfitting can happen and the model's generalizability will suffer. Finding a balance between complexity and performance can be facilitated by adjusting hyperparameters or by applying strategies like cross-validation.

An additional challenge is managing outliers. Outliers may not be well handled by Random Forests, particularly in regression tasks. Because the majority voting mechanism treats all data equally, outliers can have a significant impact on predictions. Model robustness can be increased by investigating alternative ensemble techniques, such as robust regression techniques, that are less susceptible to outliers or by scaling features.

Random Forests are not perfect for capturing intricate nonlinear relationships. Even though they are adaptable, they might have trouble with complex patterns. Examining more intricate models may be helpful. A Random Forest's capacity to capture nonlinearities can also be improved by feature engineering or variable transformation.

Random Forests rely heavily on ensemble learning for both regression and classification. They increase robustness against different aspects of the data and decrease the risk of overfitting by utilising the diversity of individual models to improve overall performance. When compared to individual models, this collective decision-making approach frequently produces predictions that are more accurate and stable.

Decision trees serve as the foundational learners for Random Forests. Decision trees are sometimes referred to as "weak learners" because of their propensity to overfit, but Random Forests excel at aggregating these weak learners. This makes up for the weaknesses of individual trees and results in a more reliable and accurate model as a whole. The ensemble's ability to capture intricate data relationships is aided by this combination of weak learners.

- ## AdaBoost

AdaBoost models do have advantages and disadvantages when it comes to combining decision trees. If the base learners—in this case, individual decision trees—become unduly complex, overfitting may result. It's critical to adjust hyperparameters, like limiting the depth of the trees, to avoid overfitting. In addition to improving generalisation, this constraint can keep the model from becoming overly tailored to the training set.

AdaBoost is typically less sensitive than individual decision trees when handling outliers. Outliers can still affect performance, though, particularly because AdaBoost gives misclassified points a higher weight. Outliers may become more influential as a result, so managing their effect on the model's predictions may require the use of robust models or preprocessing methods.

AdaBoost, which uses decision trees to capture complex relationships, may still have difficulties with highly nonlinear data. To better capture complex patterns in the data, it might be necessary to investigate nonlinear models or alternative ensemble methods.

The combination of Decision Trees and AdaBoost is powerful because it is an ensemble model that utilises the capabilities of several weak learners. Through a collaborative learning approach, AdaBoost iteratively modifies the weights of misclassified samples, improving the model's overall performance and increasing its ability to handle complex datasets.

Indeed, decision trees are weak learners in the context of AdaBoost. On the other hand, AdaBoost's sequential learning mechanism is its main strength. By making up for the shortcomings of individual trees, each new tree strengthens the model as a whole by fixing the mistakes made by the ones before it. The secret to AdaBoost's ability to produce a reliable and accurate ensemble model is its iterative correction process.

- ## Gradient Boosted Decision Trees

Even though they are very effective, gradient-boosted decision trees (GBDT) also have certain drawbacks. Concerns about overfitting arise, particularly when the ensemble's individual decision trees get unduly complicated. It's critical to adjust hyperparameters like tree depth or learning rates to maintain a balance between model complexity and performance in order to reduce overfitting and improve generalisation.

Compared to individual decision trees, GBDT models are less susceptible to outliers, although outliers can still have an impact on the model. Because of their boosting nature, they might be given more weight, which could affect projections. Preprocessing methods or robust models can be used to lessen their impact on the performance of the model.

Even with its skill at capturing intricate relationships, GBDT can still be problematic with highly nonlinear data. It may be required to investigate more precisely customised nonlinear models or alternative ensemble techniques in order to capture complex patterns in these kinds of datasets.

GBDT is built on the foundation of ensemble learning. Iterative learning improves the model's overall performance and robustness by building decision trees one after the other and having each one correct the errors of the previous ones. This greatly increases the model's ability to handle complex datasets.

Indeed, decision trees are regarded as weak learners in the context of GBDT. The strength of GBDT, however, is found in its sequential learning process, where each tree makes up for the shortcomings of the ones before it. By gradually correcting errors, this strengthens the model as a whole and makes up for the shortcomings of the individual trees.