Bagging: Introduction to Bagging

CS109A Introduction to Data Science

Pavlos Protopapas, Natesh Pillai and Chris Gumb

Steven Liu
Indonesia

Outline

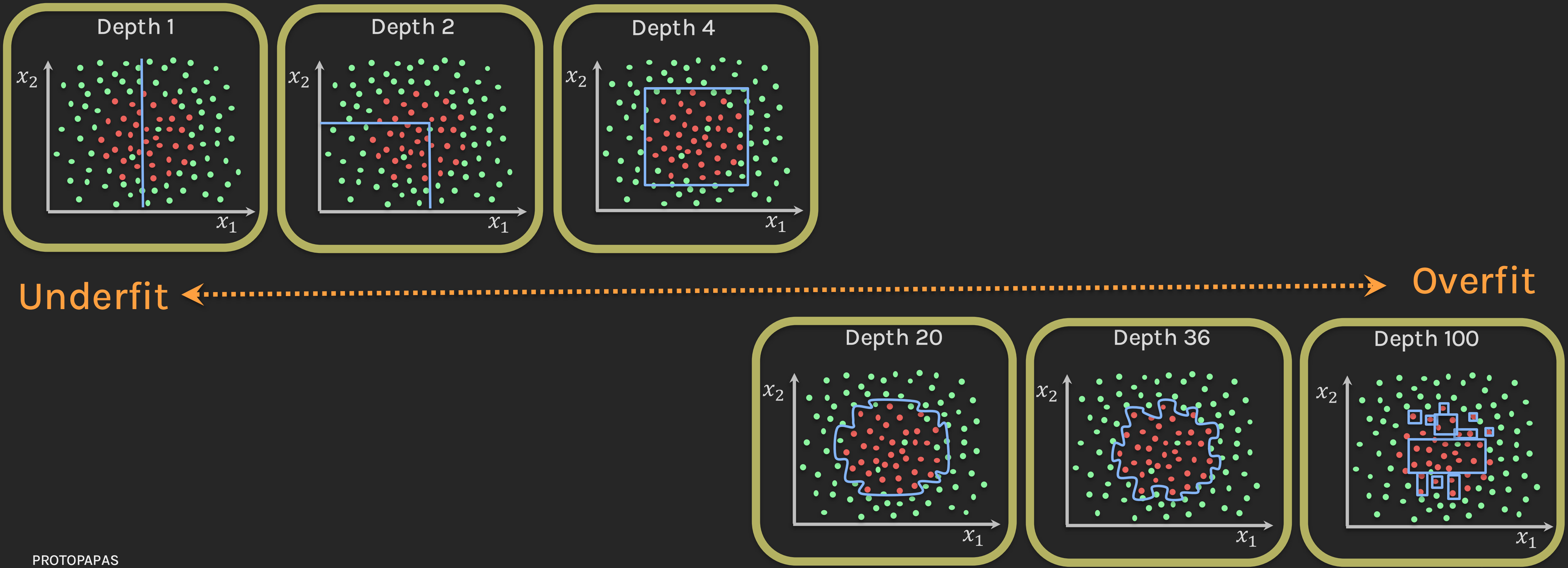
- Motivation
- Bagging
- Out-of-bag Error

Outline

- **Motivation**
- Bagging
- Out-of-bag Error

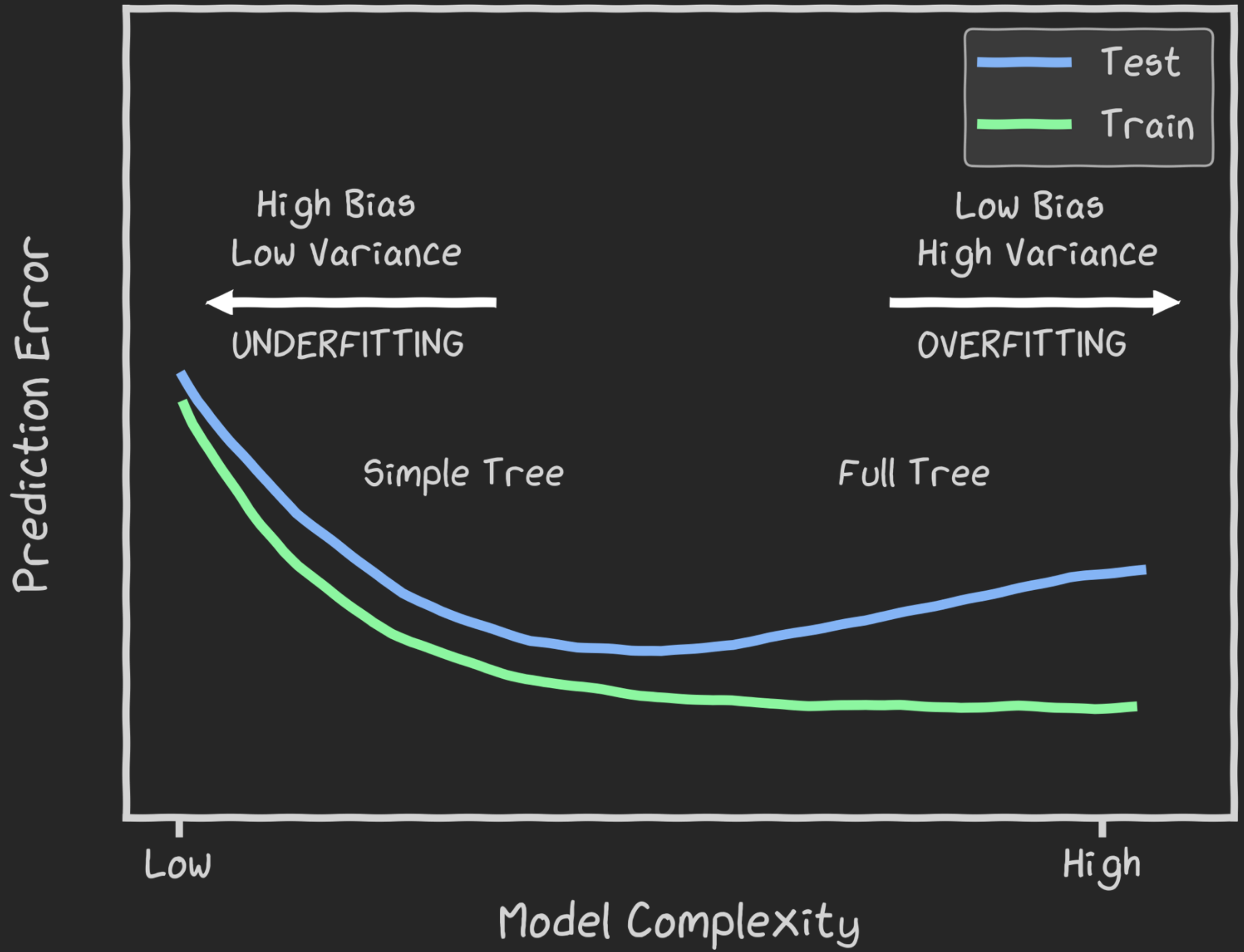
Underfitting and Overfitting

When a tree is too shallow, it cannot divide the input data into enough regions, so the model underfits. When the tree is too deep, it cuts the input space into too many regions and fits to the noise of the data, so it overfits.



Overfitting

Avoid overfitting by pruning or limiting the depth of the tree and using CV.



Limitations of Decision Tree Models

Using a greedy algorithm, decision trees models are highly interpretable and fast to train.

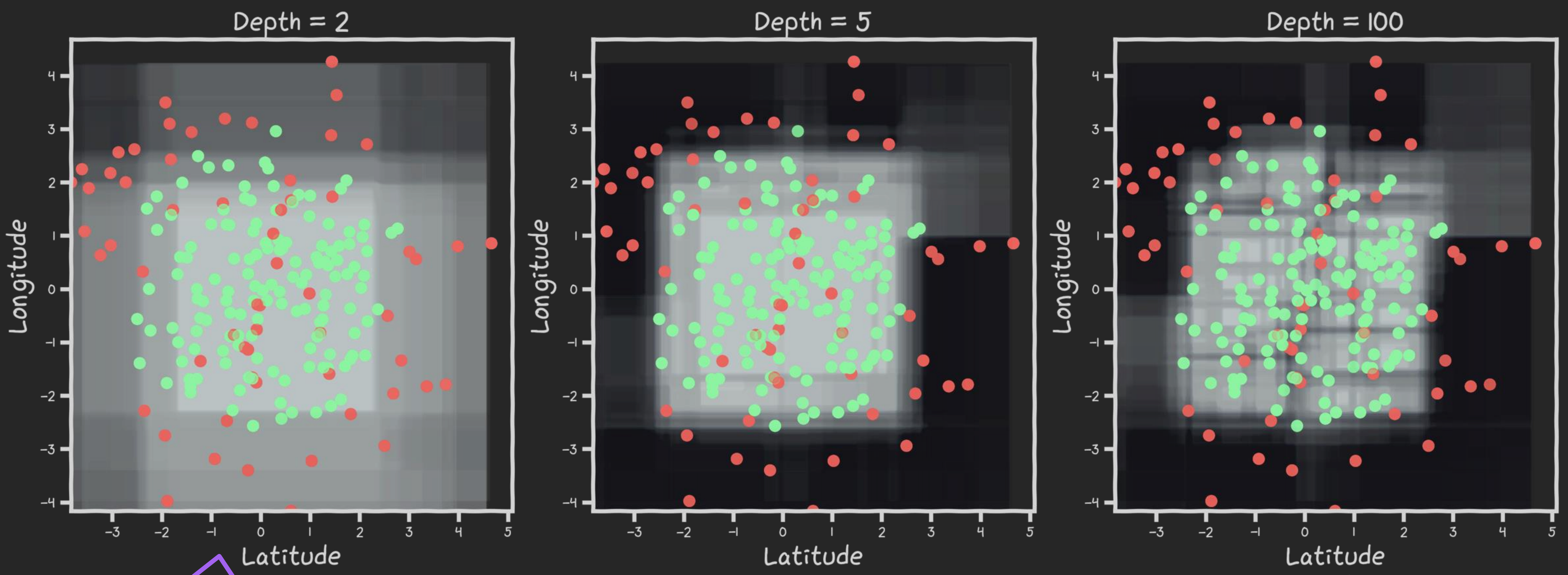
However, to **capture a complex decision boundary** (or approximate a complex function for regression), we must do axis-aligned splits. As a result, we need to use a large tree or deep tree.

Deep trees have high variance and are prone to overfitting.

For these reasons, in practice, decision tree models underperform in comparison to other classification or regression methods.

Motivation for Bagging

Decision tree models often underperform when compared to other classification or regression methods in situations of complex decision boundaries.



In certain cases it underfits.

… and in certain cases it overfits!

Outline

- Motivation
- **Bagging**
- Out-of-bag Error

Ensemble Learning - Intuition

The intuition of ensemble learning is to build a single model by training and aggregating multiple models.

Consider the case of MRI scans of patients and the goal is to find out if they have a brain tumor or not.

Ensemble learning is like consulting multiple doctors instead of just one.

Ensemble Learning - Intuition

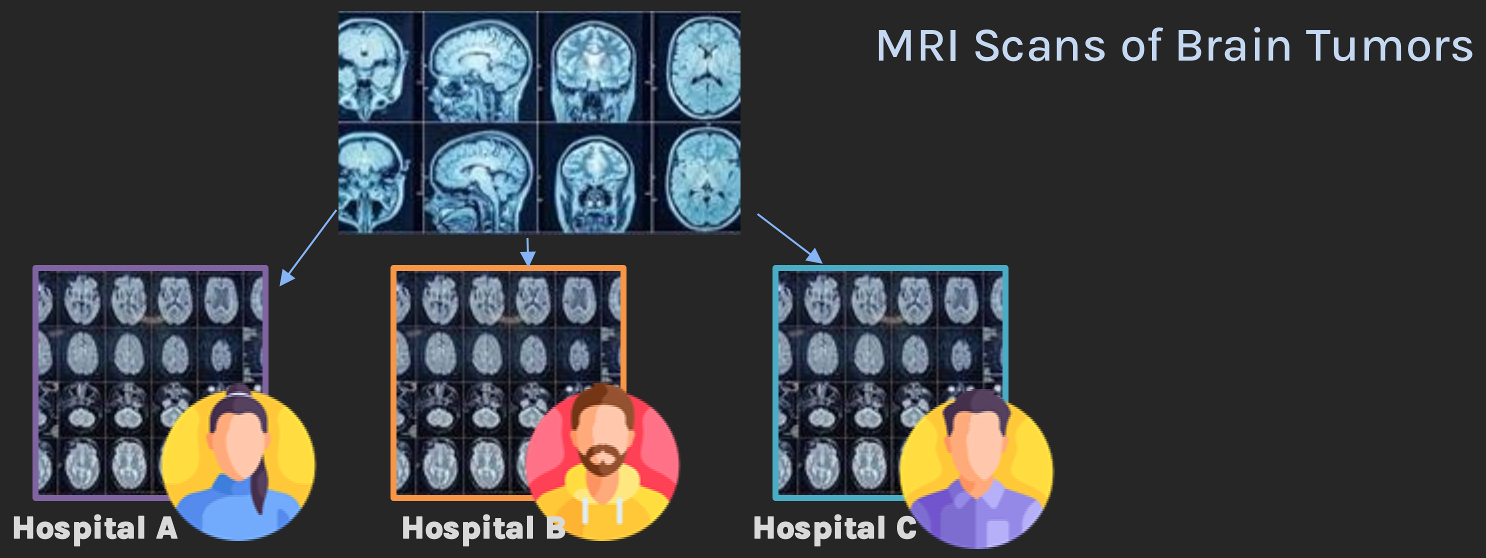
Instead of relying on the prediction from one specialist we consult multiple doctors.

We individually ask each of these doctors to predict whether the scan indicates the presence of brain tumor.

Each 'specialist' or model views the MRI scans and makes a prediction. Their collective decisions are then aggregated to form a final verdict. This approach helps in mitigating individual model errors, increasing the accuracy of the prediction.

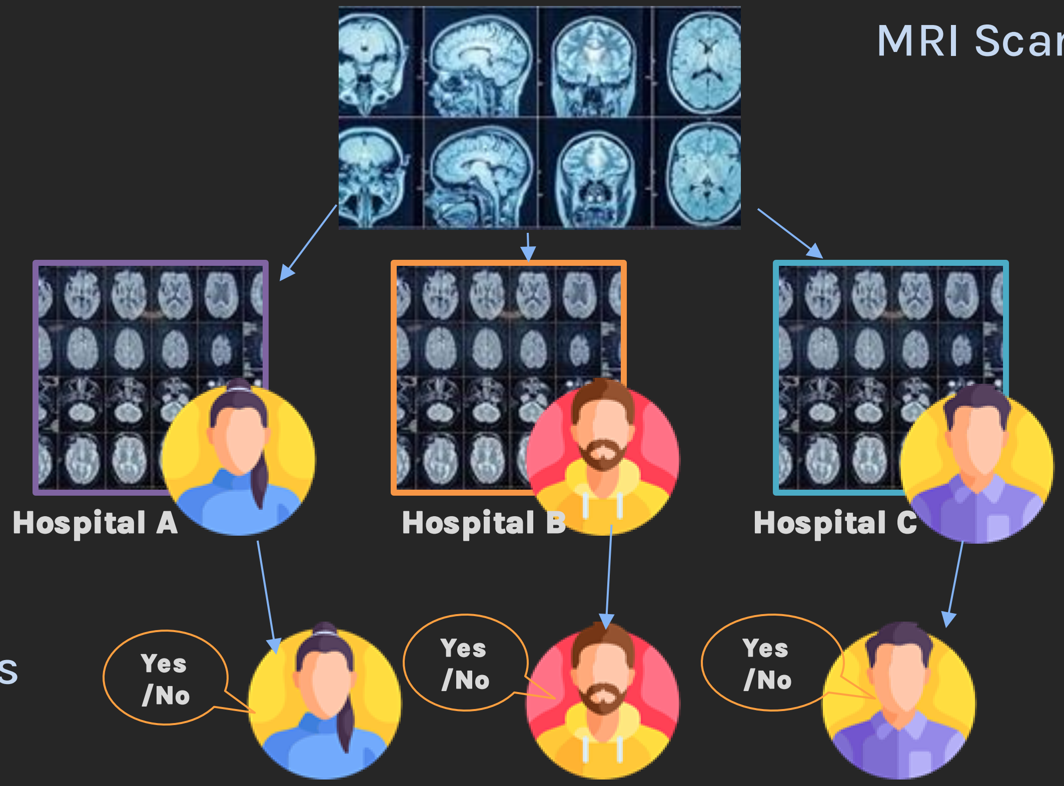
Ensemble Learning - Intuition

Doctors who are
- trained at different hospitals and
- trained on the examples available to them



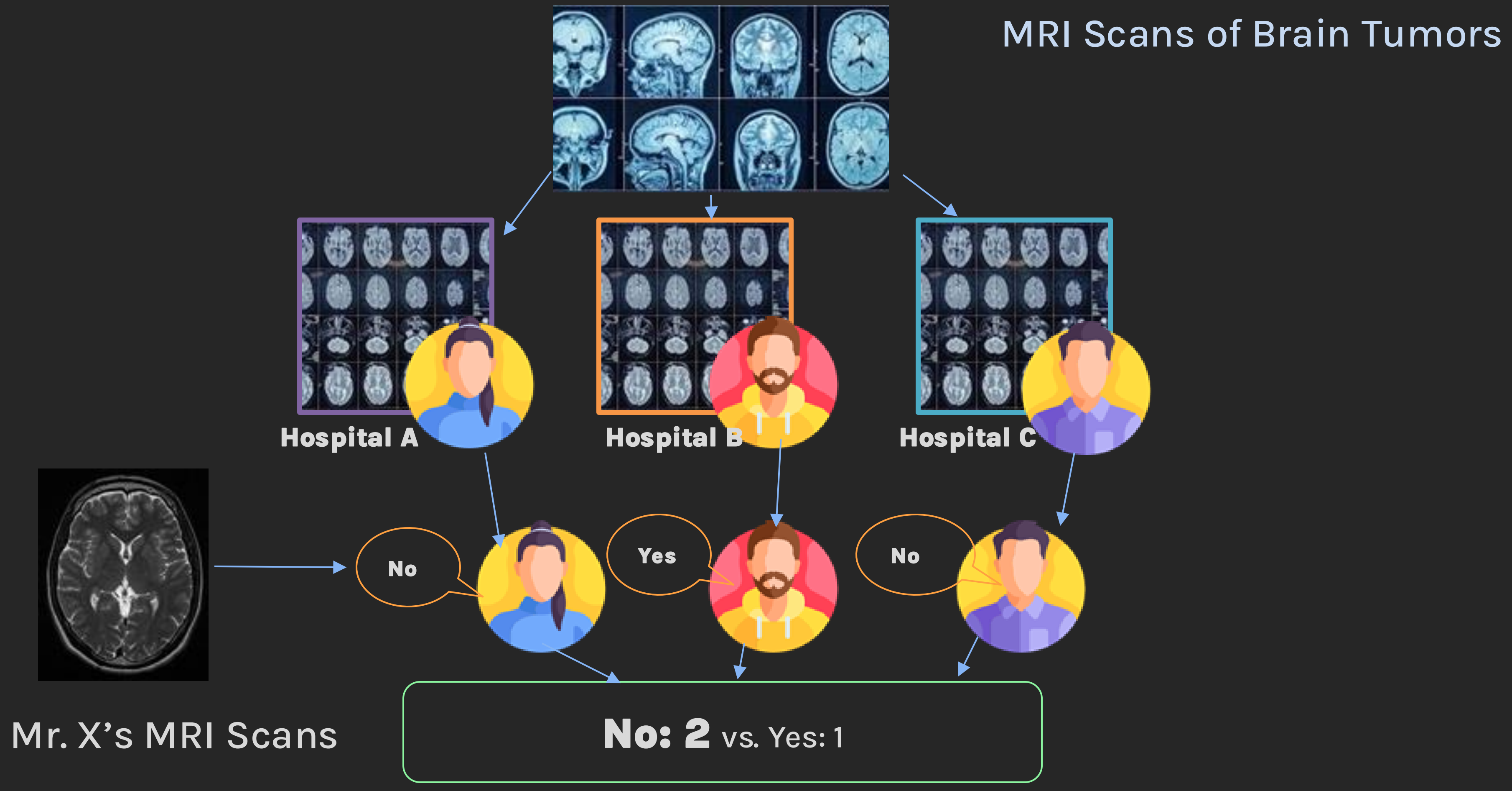
Ensemble Learning - Intuition

MRI Scans of Brain Tumors



Every doctor is “trained” to identify tumors

Ensemble Learning - Intuition



Hoping to increase the likelihood of an accurate diagnosis, Mr. X sends his MRI scans to the 3 doctors.

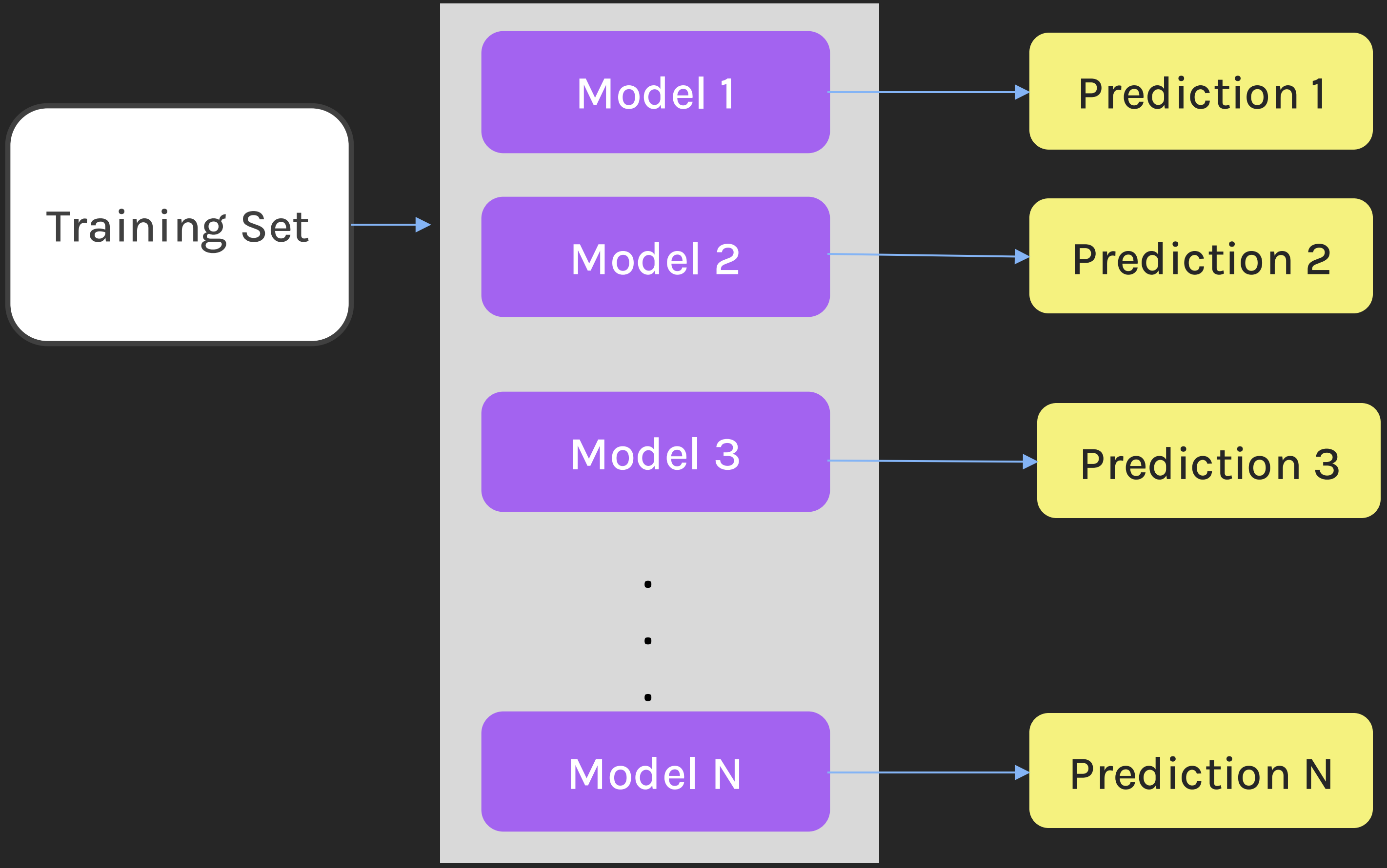
Ensemble Learning

Ensemble learning is a machine learning technique that combines several base models to produce one single optimal predictive model.

Training Set

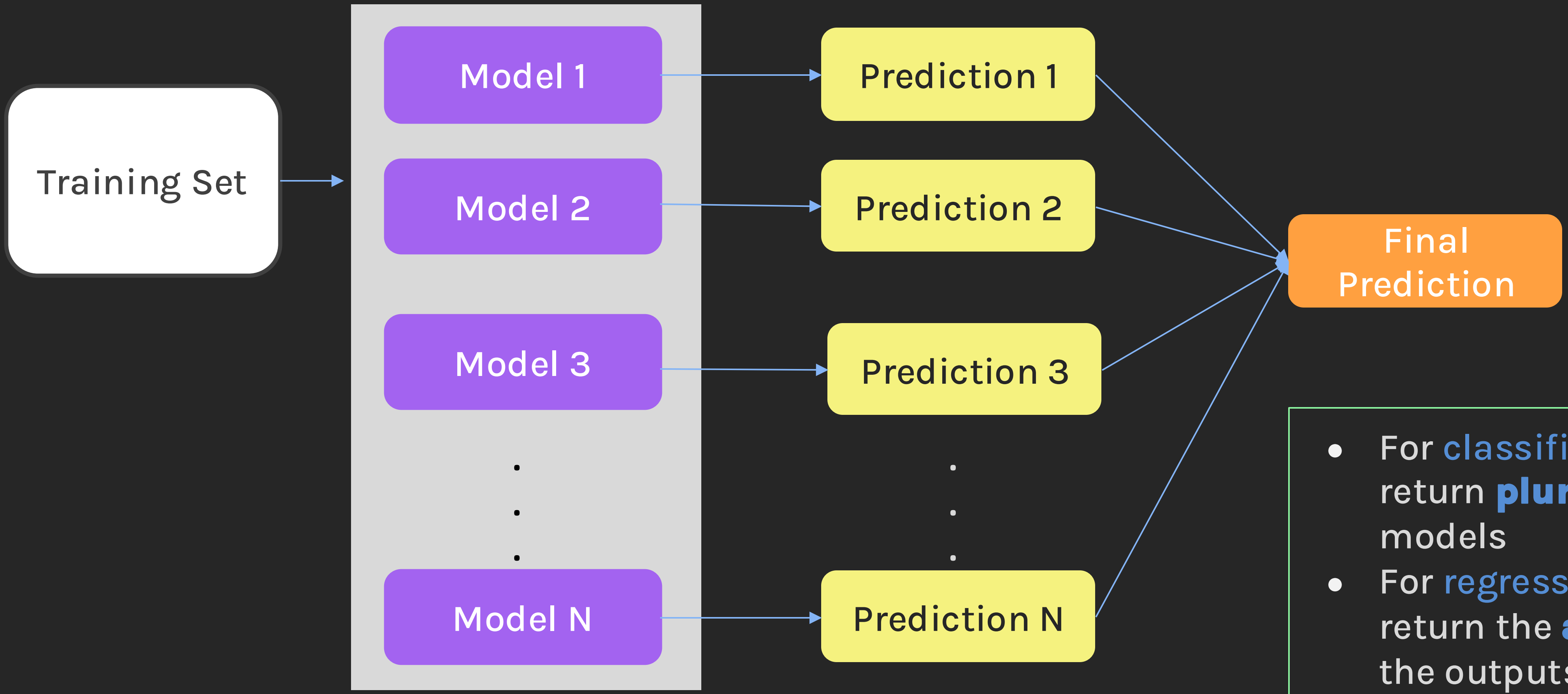
Ensemble Learning

Ensemble learning is a machine learning technique that combines several base models to produce one single optimal predictive model.



Ensemble Learning

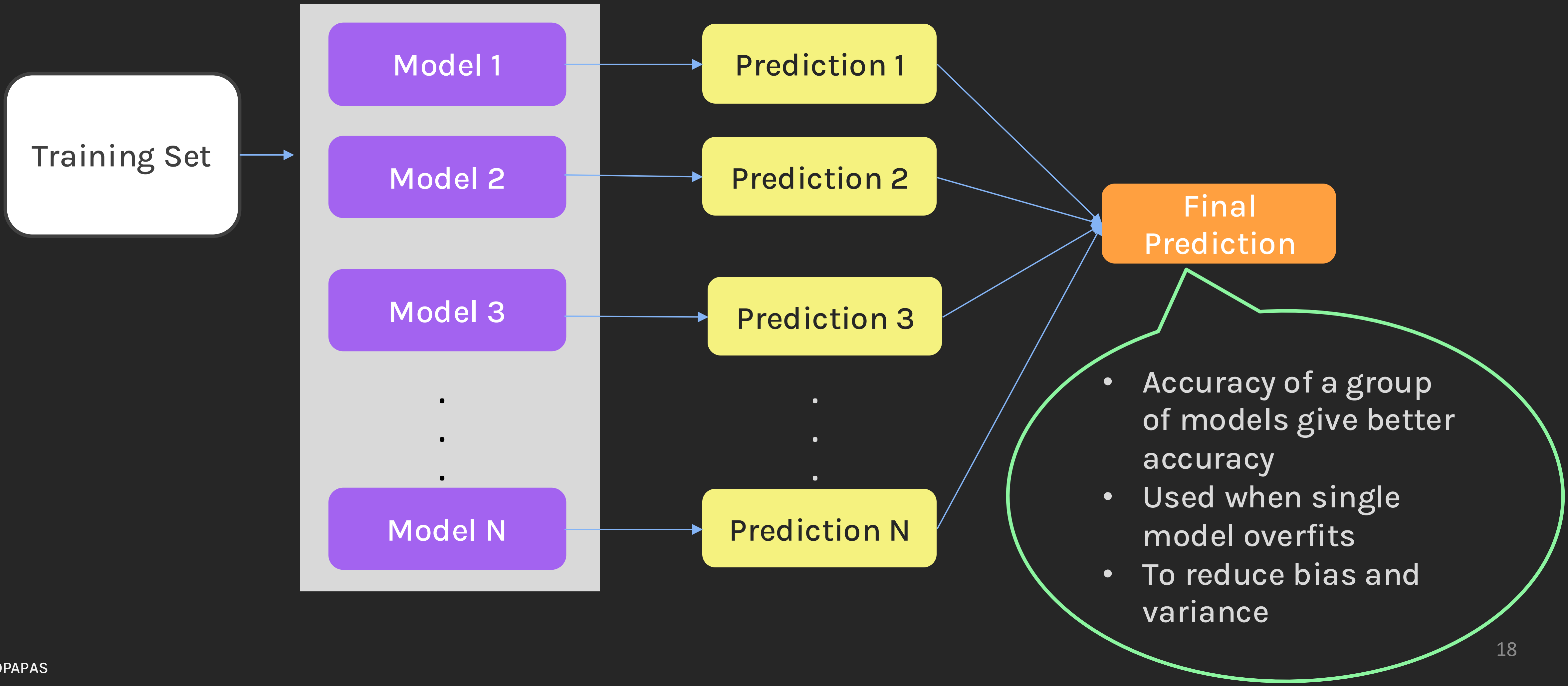
Ensemble learning is a machine learning technique that combines several base models to produce one single optimal predictive model.



- For classification, we return **plurality** of the models
- For regression, we return the **average** of the outputs from the models

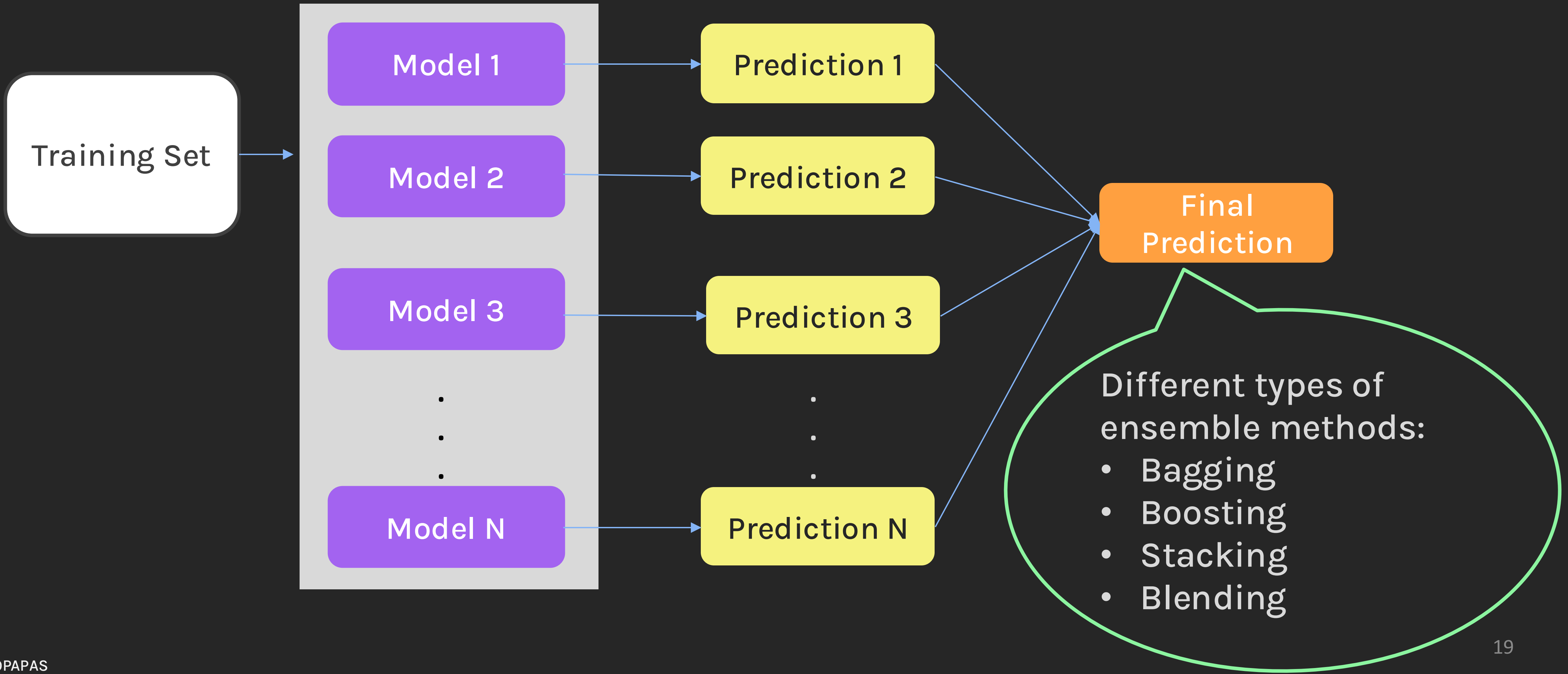
Ensemble Learning

Ensemble learning is a machine learning technique that combines several base models to produce one single optimal predictive model.



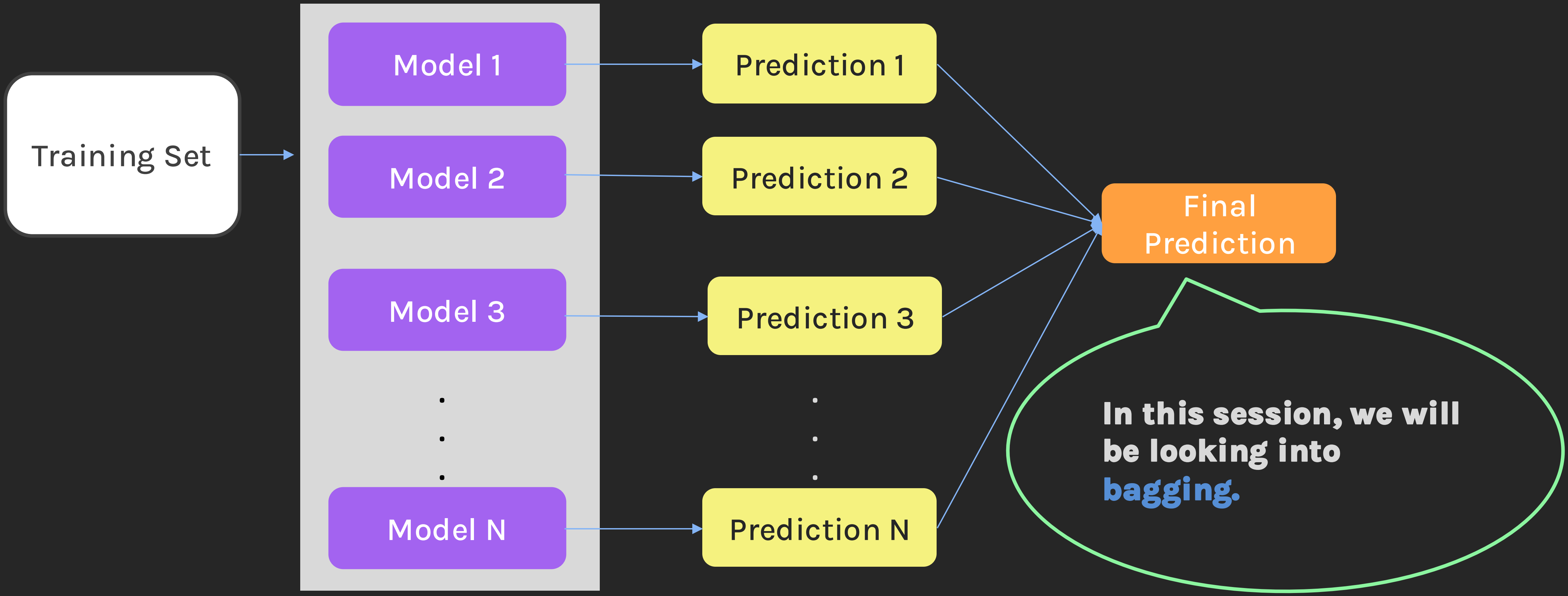
Ensemble Learning

Ensemble learning is a machine learning technique that combines several base models to produce one single optimal predictive model.



Ensemble Learning

Ensemble learning is a machine learning technique that combines several base models to produce one single optimal predictive model.



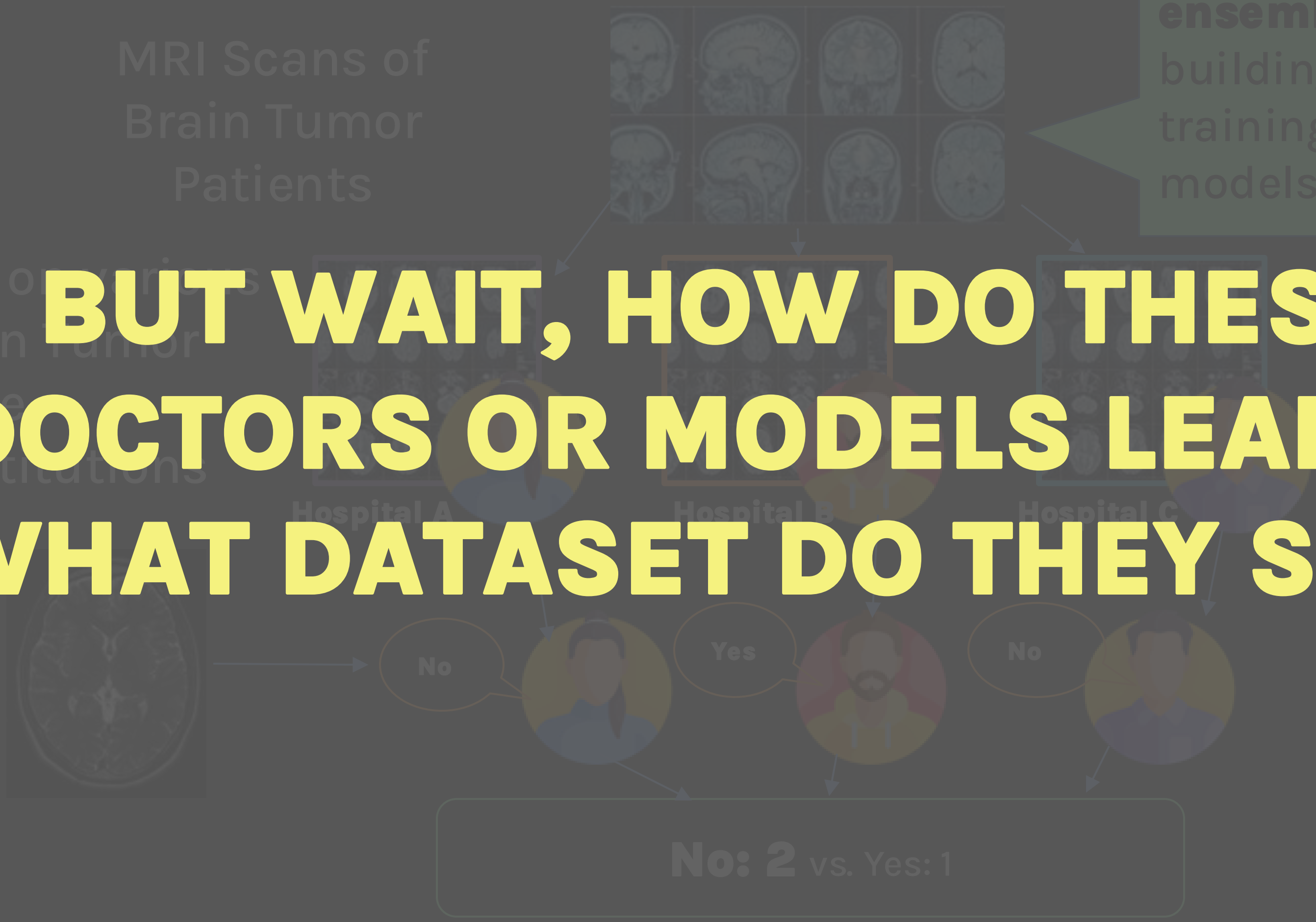
Ensemble Learning - Intuition

BUT WAIT, HOW DO THESE DOCTORS OR MODELS LEARN? WHAT DATASET DO THEY SEE?

MRI Scans of Brain Tumor Patients

Doctors trained o[illegible] samples of Brain Tumor MRI Scans in [illegible] educational i[illegible] they studied i[illegible]

Mr. X sends his MRI Scan forward to get the diagnosis from 3 doctors just to make sure that the results are more confident.



This is the intuition behind **ensemble method**, a method of building a single model by training and aggregating multiple models.

Bootstrap - Motivation

In practice, we do not have different datasets or different doctors.

We want multiple models (doctors) to train on different datasets.

However, we have only one dataset.

How can we generate more datasets?

To address the data scarcity for model fitting, we generate new datasets using …

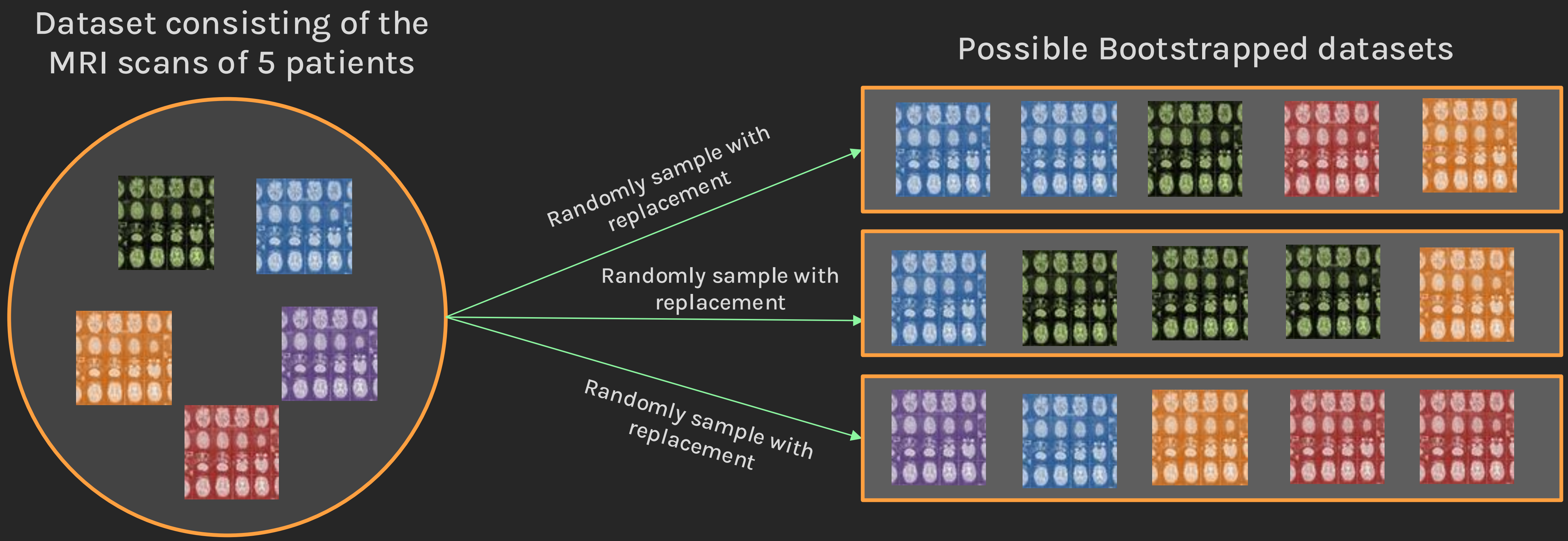
Bootstrapping

But what is Bootstrapping ?

Bootstrapped datasets!

Bootstrapping

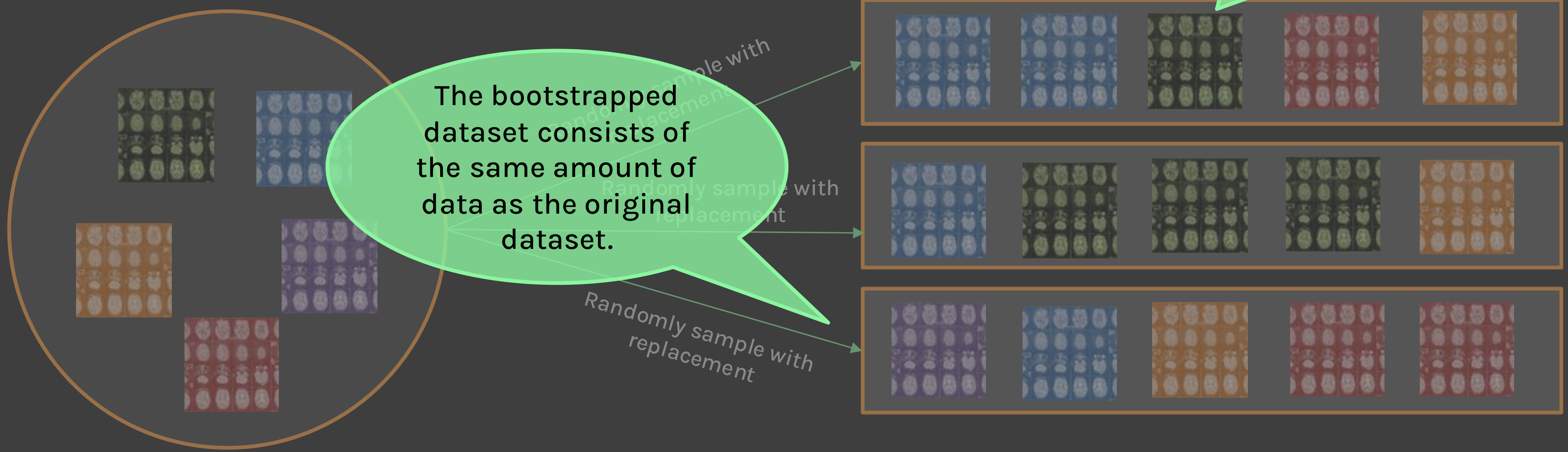
Bootstrapping is the process of **sampling with replacement** from a dataset and performing calculations on such multiple bootstrapped datasets to get an overall aggregate inference.



Bootstrapping [TWO KEY IDEAS WE WANT TO EMPHASIZE)

Bootstrapping is the process of **sampling with replacement** from a dataset and performing calculations on such multiple bootstrapped datasets to get an overall aggregate inference.

Randomly choosing data and allowing for duplication is called **Sampling with replacement!**



The bootstrapped dataset consists of the same amount of data as the original dataset.

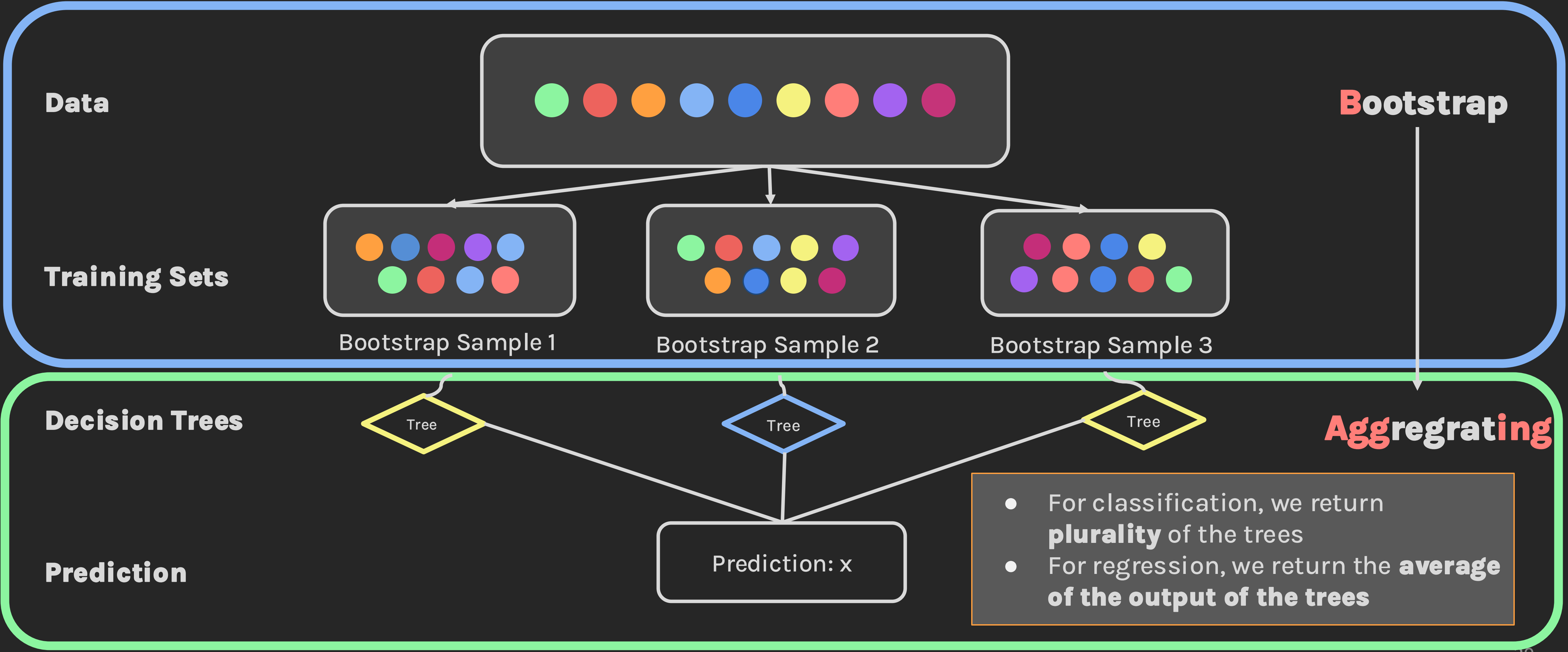
Combining Ensemble and Bootstrapping together …

Bootstrap + aggregating

Bagging

We get Bagging!

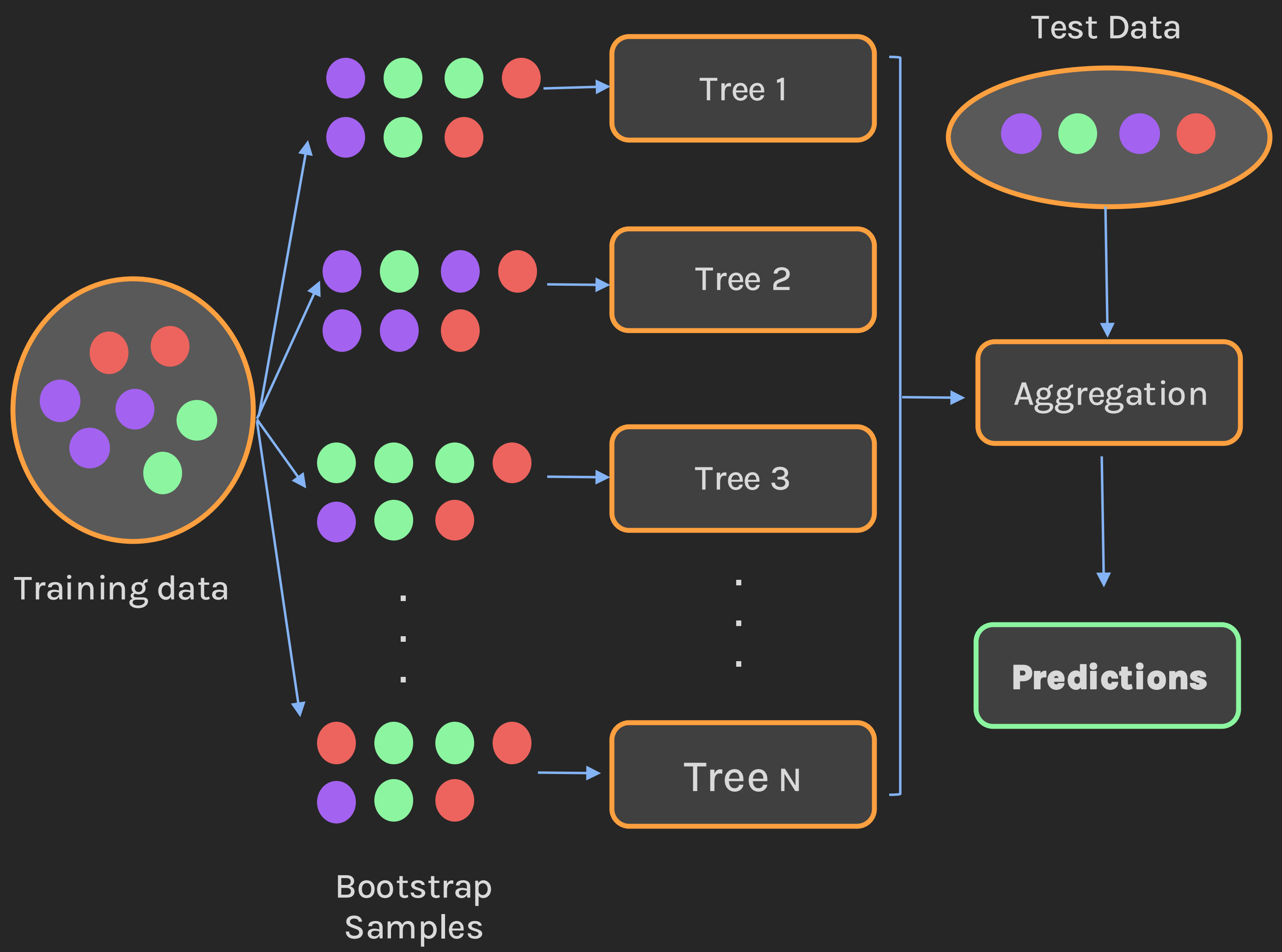
This method is called **Bagging** (Breiman, 1996), short for, of course, **B**ootstrap **Agg**regat**ing**.



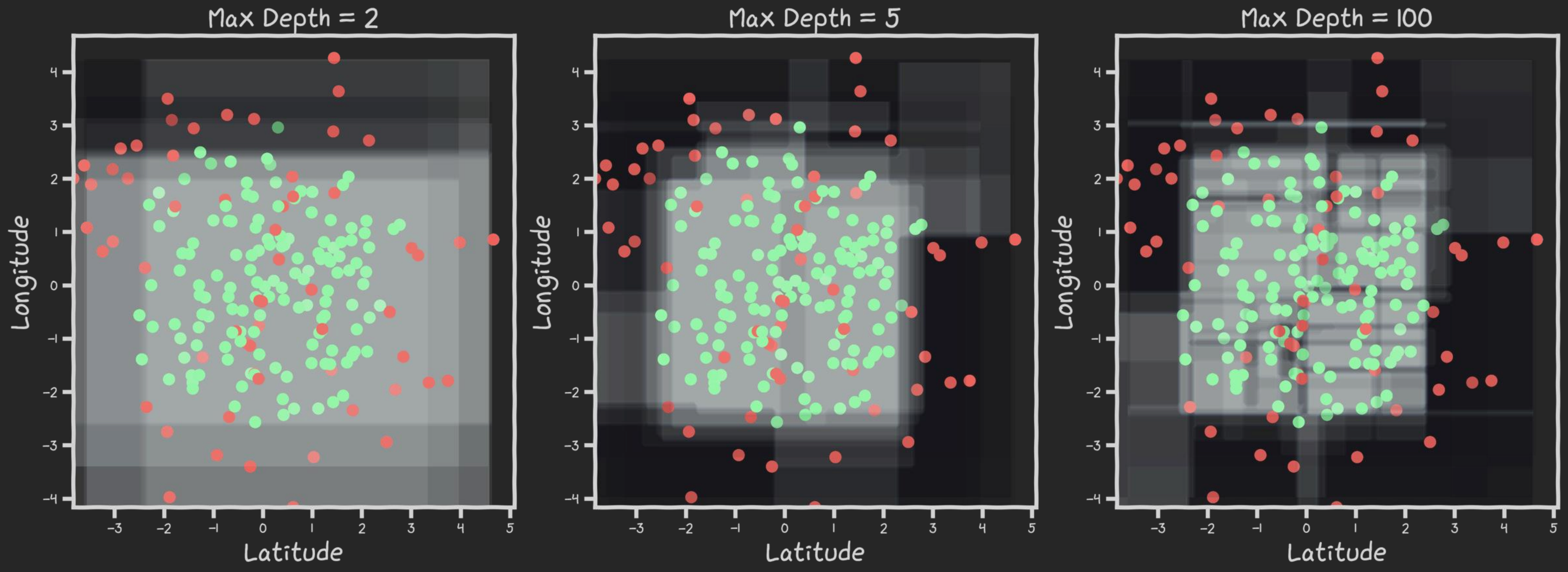
Bagging - Bootstrap + Aggregating

1. **Bootstrap:** we generate multiple samples of training data, via bootstrapping. We train a deep decision tree on each sample of data.

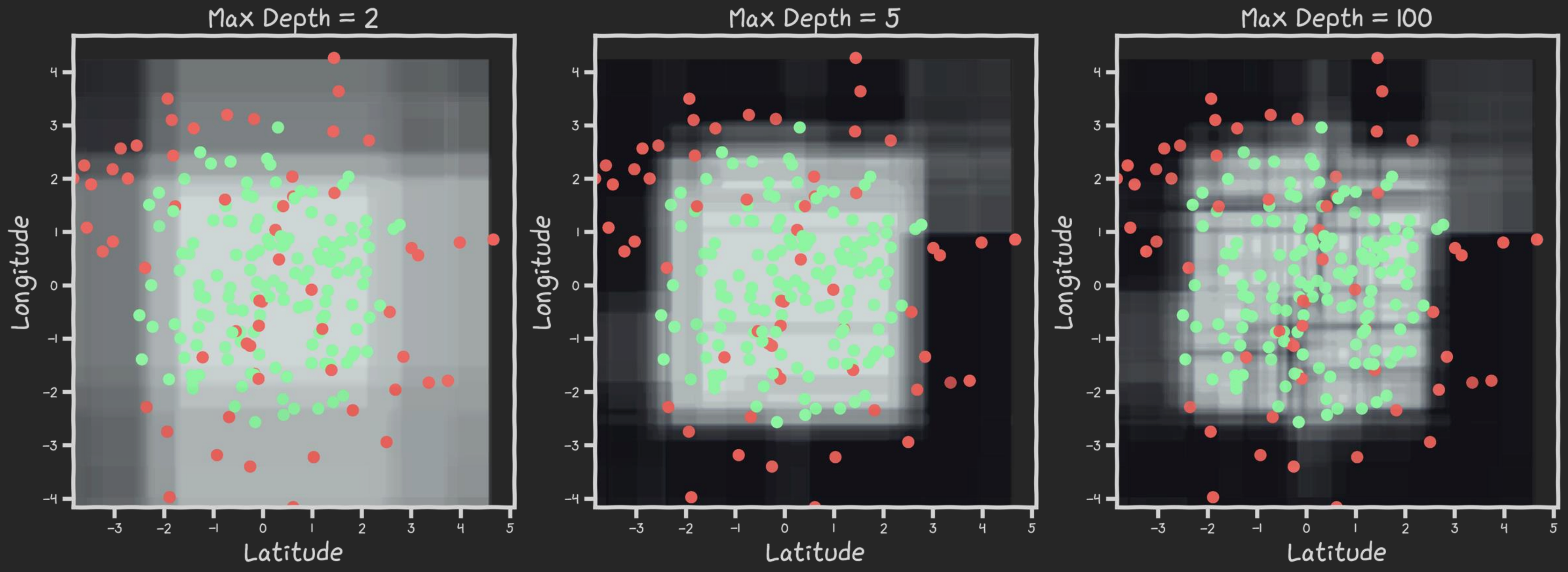
2. **Aggregating:** for a given input, we output the averaged outputs of all the models for that input.



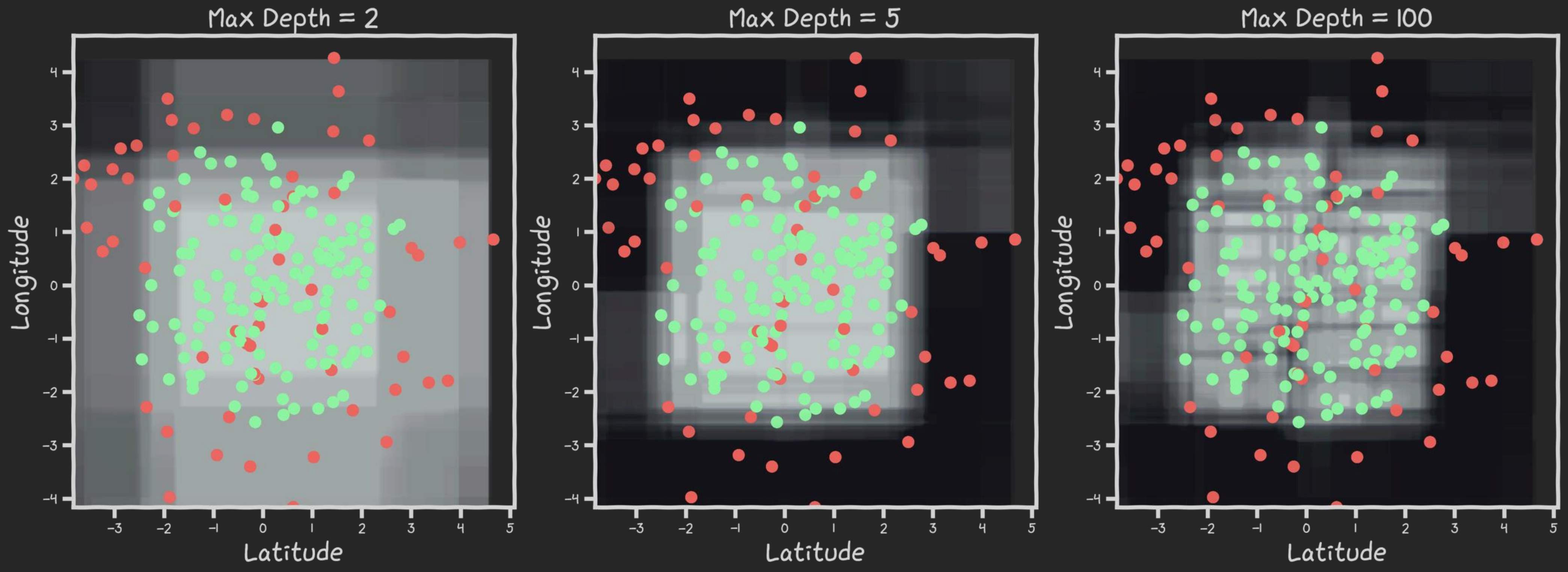
Examples of tree of various depth for 10 bootstrapped samples



Examples of tree of various depth for 100 bootstrapped samples



Examples of tree of various depth for 150 bootstrapped samples



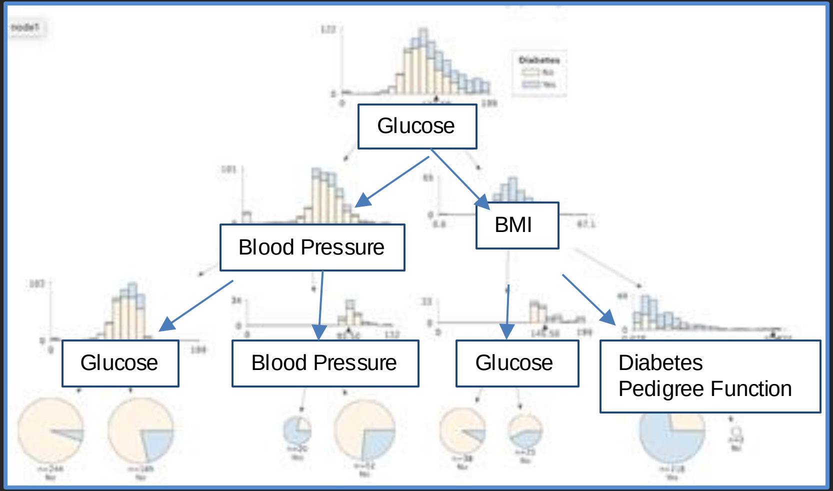
Advantages of Bagging

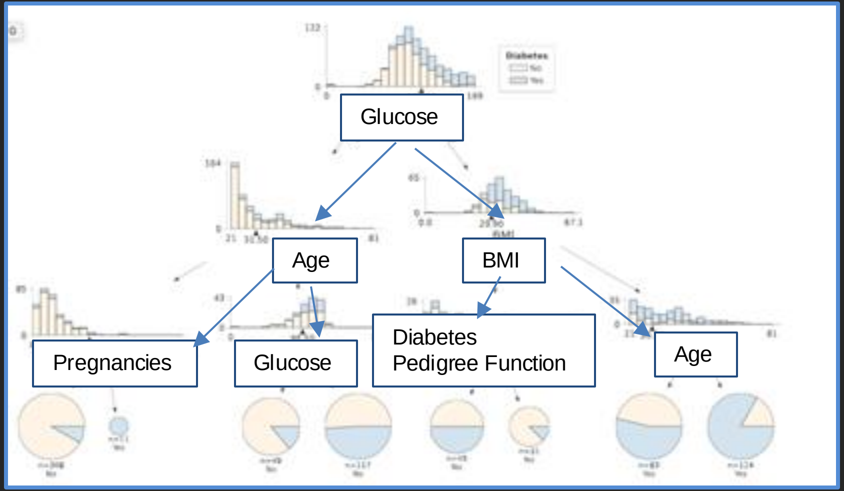
Bagging enjoys the benefits of:

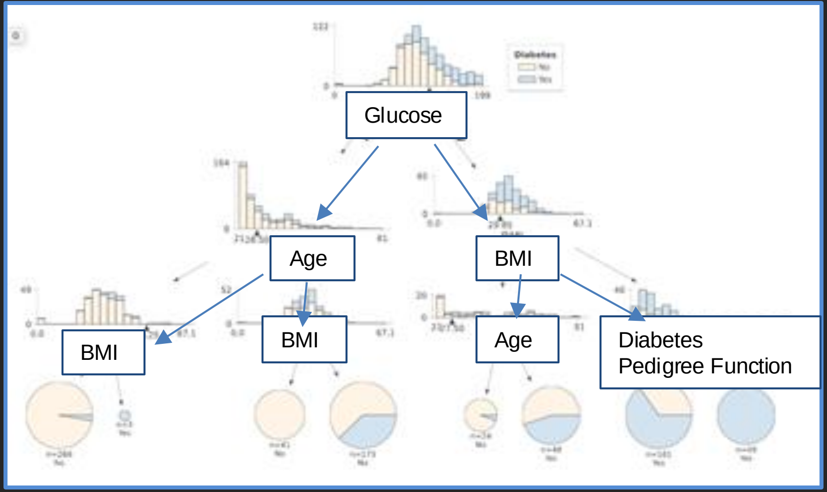
1. High expressiveness - by using deeper trees each model is able to approximate complex functions and decision boundaries.

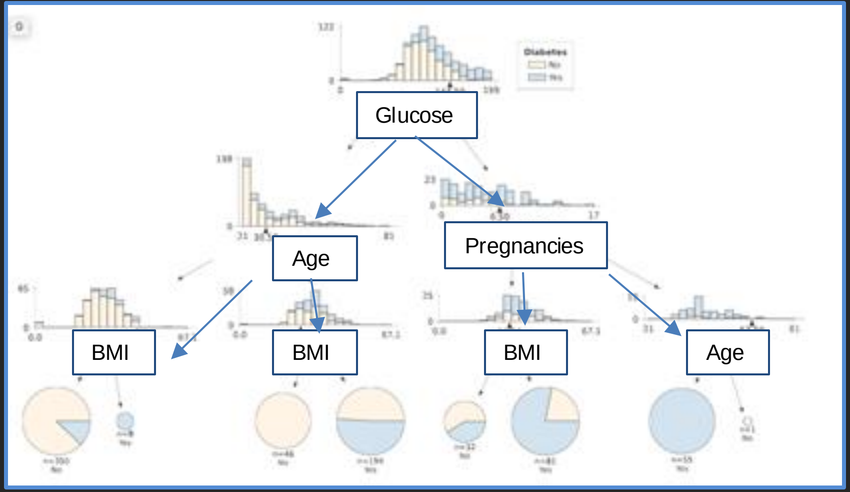
2. Low variance - averaging the prediction of all the models reduces the variance in the final prediction, assuming that we choose a sufficiently large number of trees.

Classification in Bagging



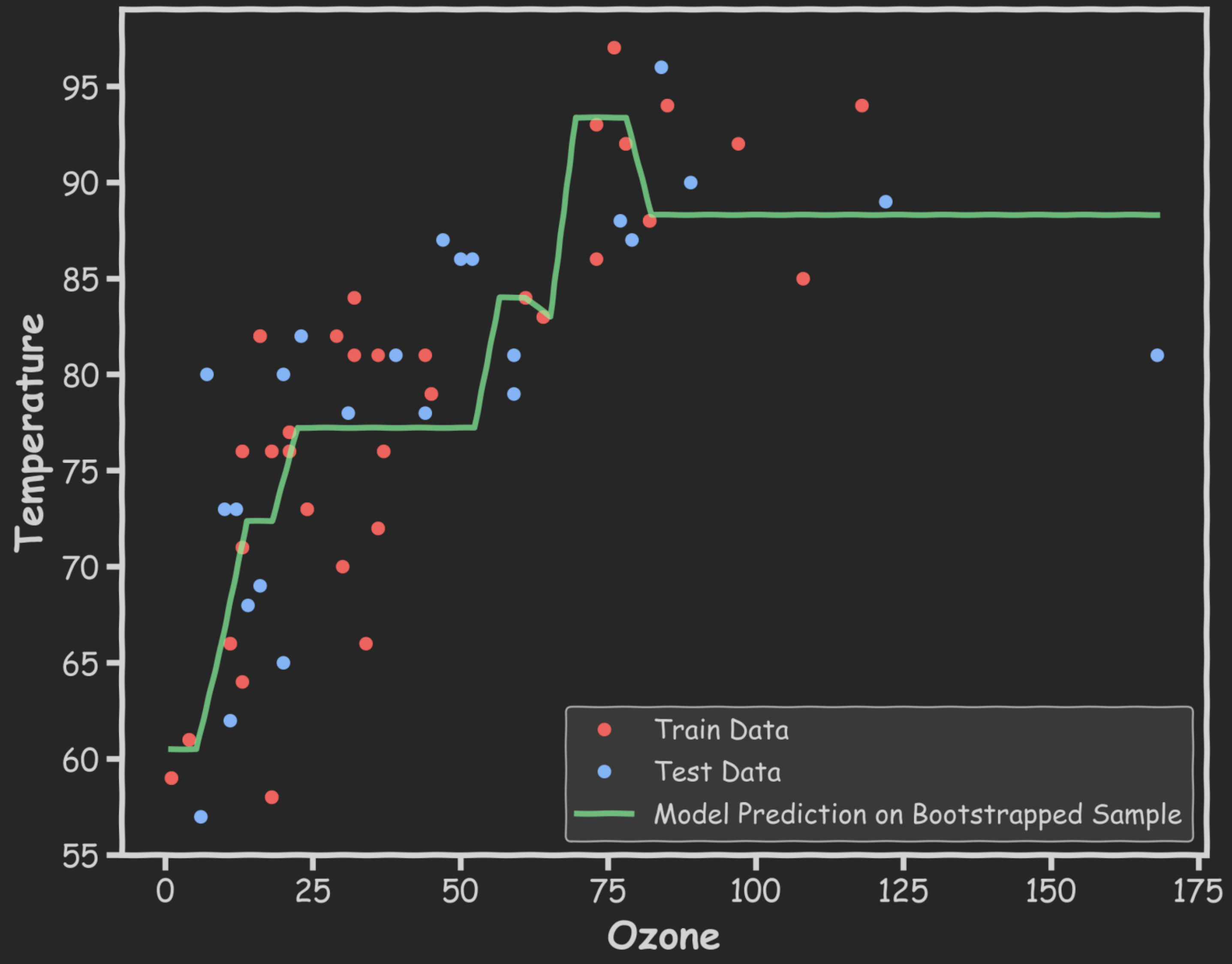






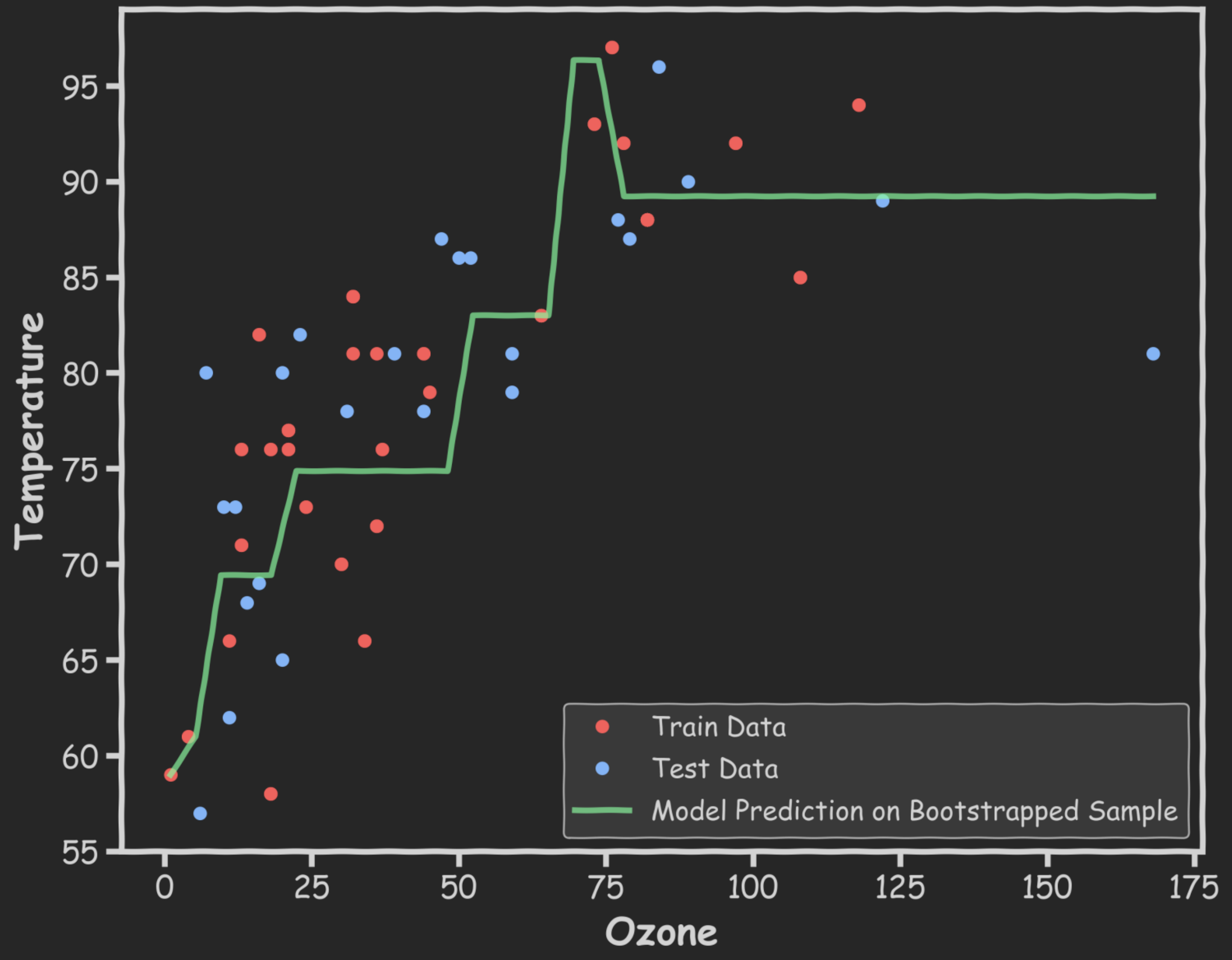
For each bootstrap, we build a decision tree. The results is a combination (majority) of the predictions from all trees.

Prediction from Decision Tree #1 with Bootstrapped Sample #1



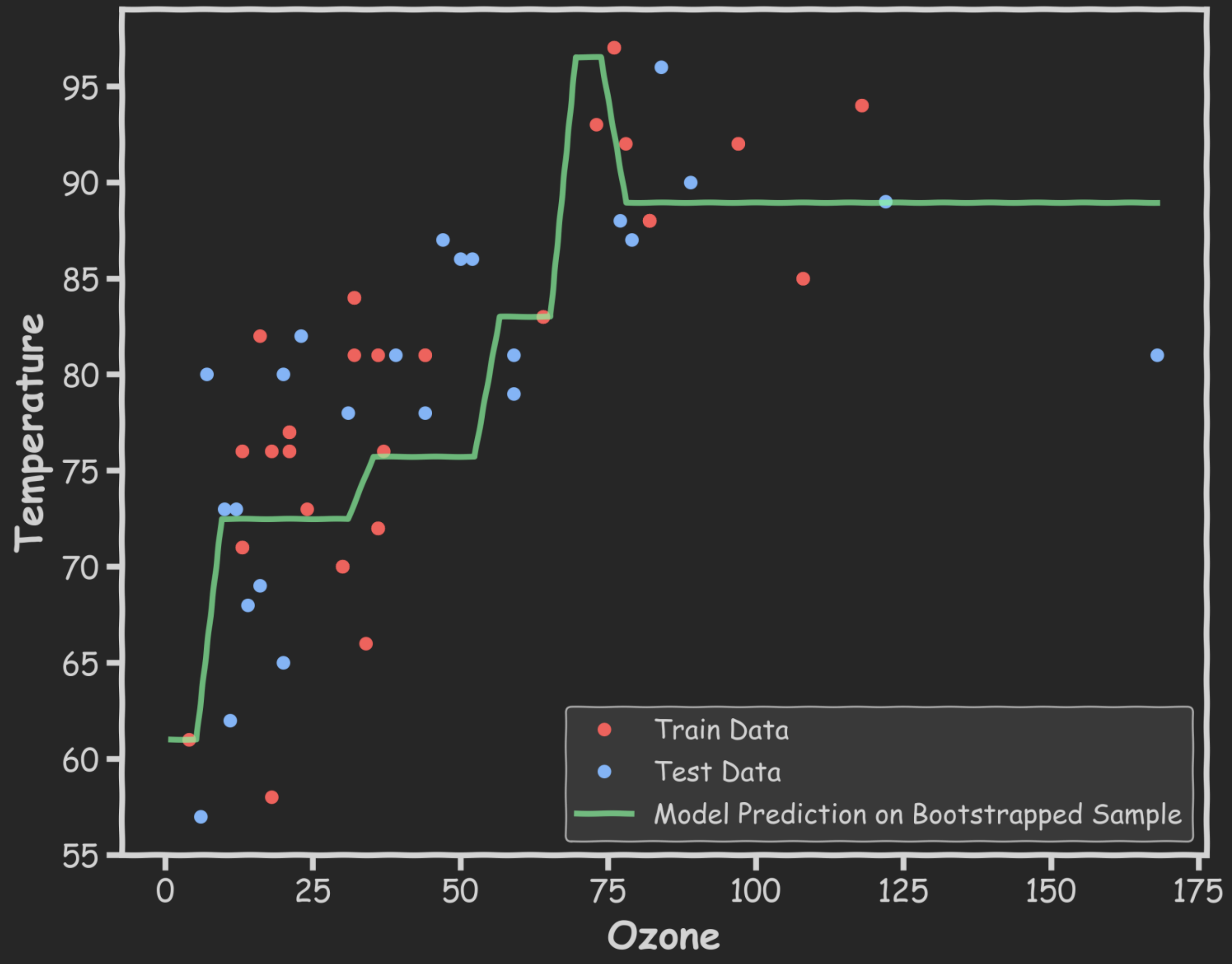
Model prediction on the same test set when fit on one version of the bootstrapped train data

Prediction from Decision Tree #2 with Bootstrapped Sample #2



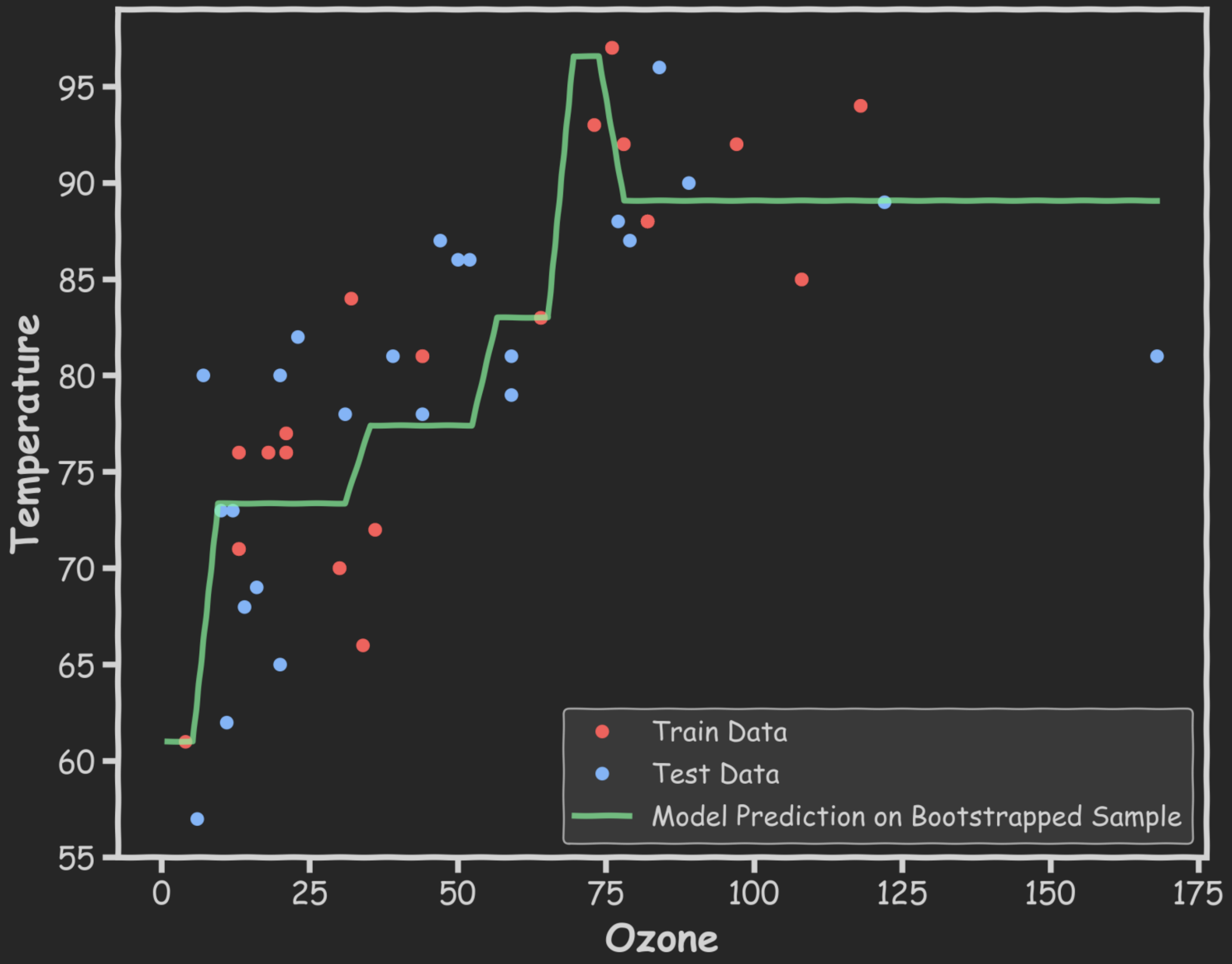
Model prediction on the same test set when fit on one version of the bootstrapped train data

Prediction from Decision Tree #3 with Bootstrapped Sample #3



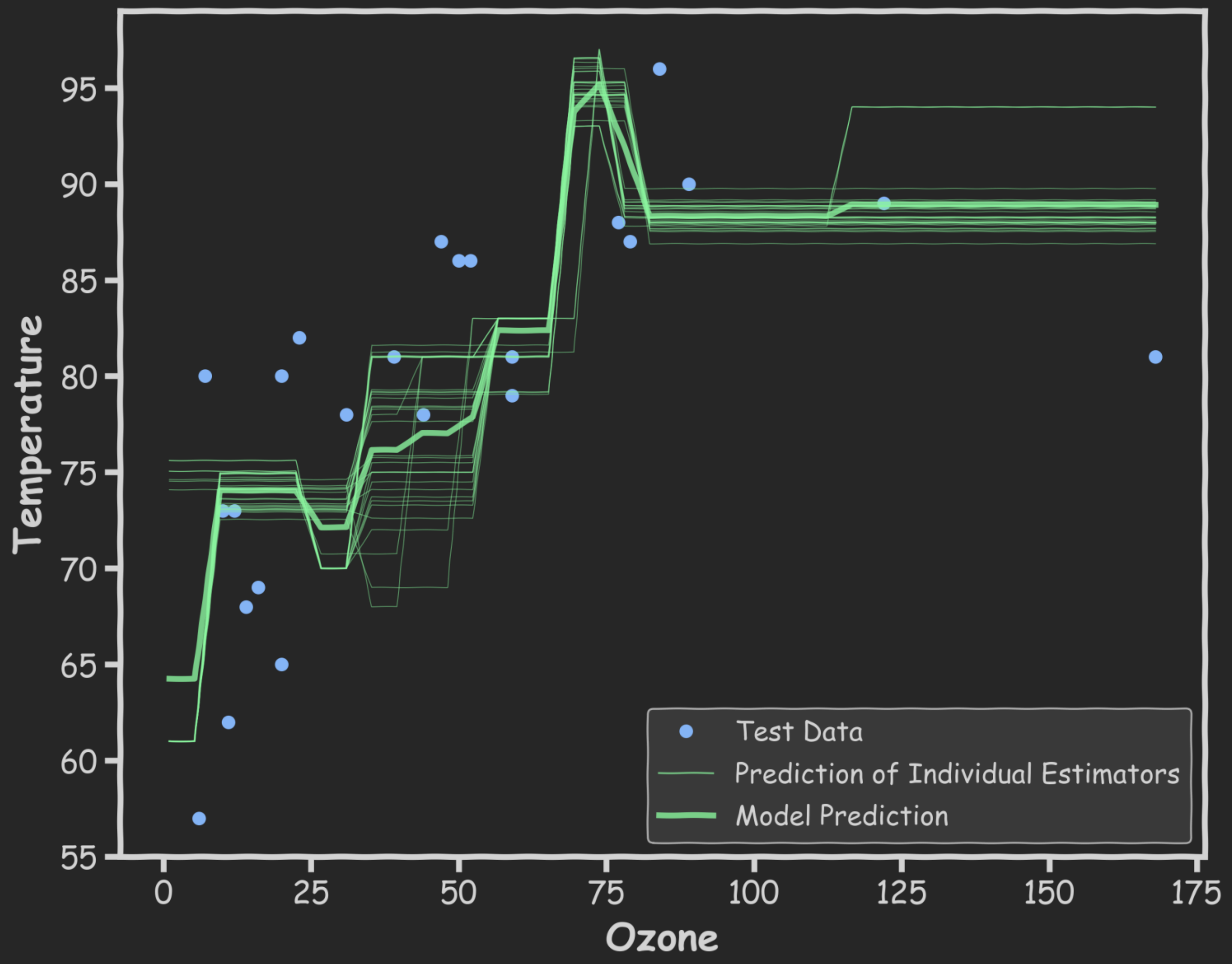
Model prediction on the same test set when fit on one version of the bootstrapped train data

Prediction from Decision Tree #4 with Bootstrapped Sample #4



Model prediction on the same test set when fit on one version of the bootstrapped train data

Prediction from all Decision Trees



The prediction by the bagging regressor model is the average of all the individual predictions of the trees.

Drawbacks of Bagging

Interpretability -

A **major drawback** of bagging (and other ***ensemble methods*** that we will study) is that the averaged model is no longer easily interpretable - i.e. one can no longer trace the 'logic' of an output through a series of decisions based on predictor values!

Underfitting and overfitting -

Drawbacks of Bagging

Interpretability -

A **major drawback** of bagging (and other ***ensemble methods*** that we will study) is that the averaged model is no longer easily interpretable - i.e. one can no longer trace the 'logic' of an output through a series of decisions based on predictor values!

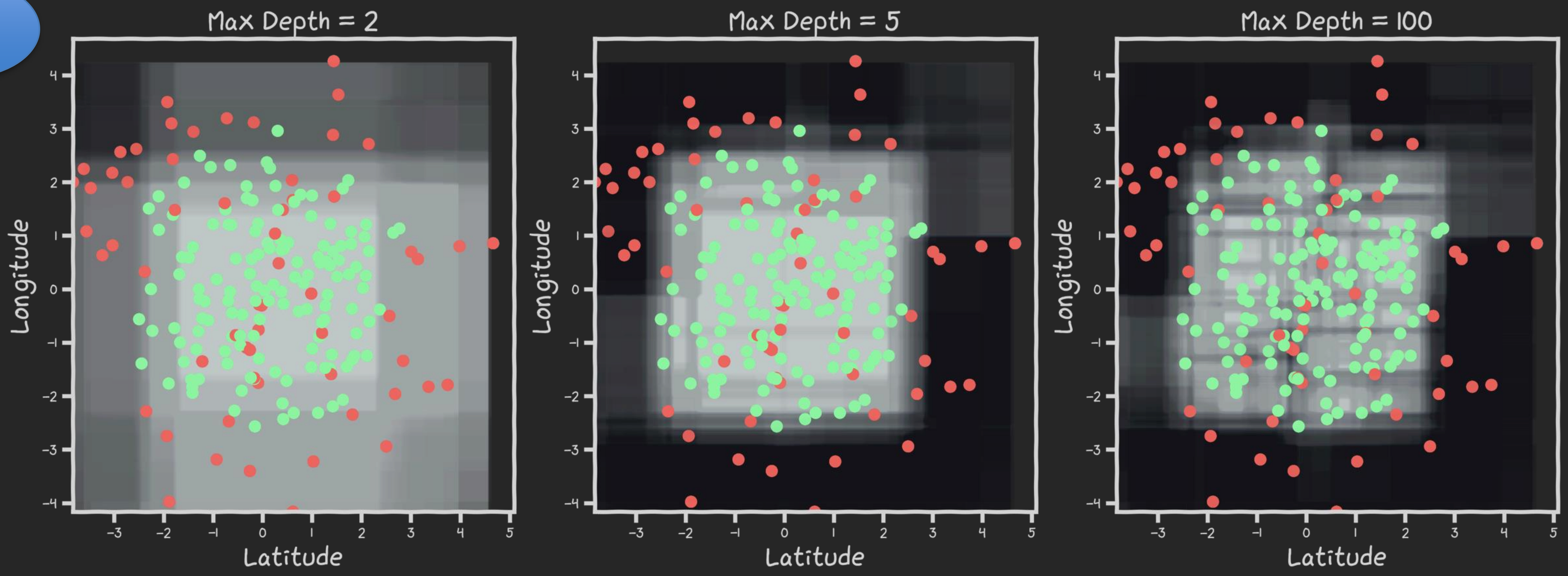
This interpretability challenge will be addressed in our upcoming lecture on random forests, where we will introduce methods like MDI (Mean Decrease Impurity) and permutation importance to provide insights into model decisions.

Underfitting & Overfitting

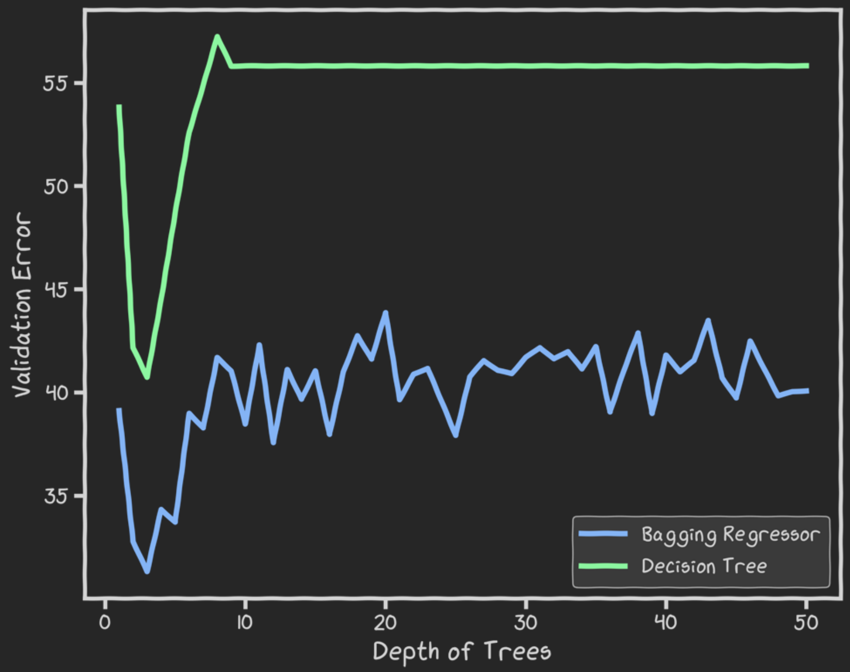
Here we fit 100 trees using bootstrapped samples. Even with multiple estimators, the shallow tree will not be able to capture the real pattern.

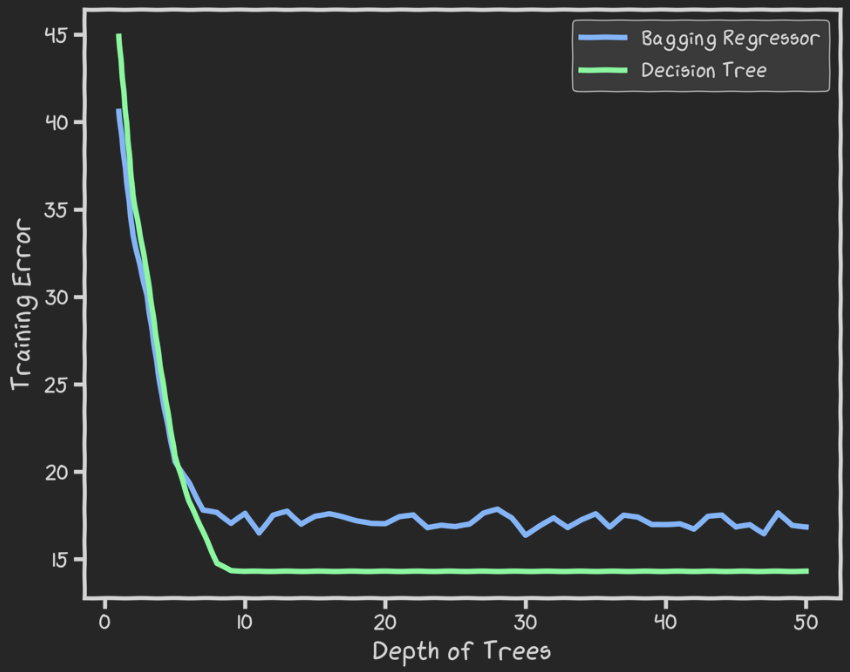
Trees

Number of Bootstraps = 100



Underfitting and Overfitting in Decision Tree vs. Bagging





- The graphs shows that the **more depth we add to Decision Trees, the faster the model overfits**

- However, for Bagging, there is still some form of **better model performance maintained after increase of depth of trees** although eventually it does **lead to no improvement of performance after a certain point as well**

Cases of underfitting and overfitting in Bagging

Cross validation is cool but computationally expensive and requires a larger dataset.

Before we conclude the bagging session, we will present another method of measuring the performance of ensemble methods.

Out-of-bag error

Summary

What are the limitations of single decision tree models, and how does bagging address these limitations?
Single decision tree models, particularly deep ones, are prone to overfitting and may have high variance. Bagging addresses these limitations by creating an ensemble of multiple decision trees trained on different bootstrap samples of the data. By averaging the predictions of these trees, bagging reduces variance and mitigates overfitting.

Explain the concept of bootstrapping and its significance in bagging.
Bootstrapping is a resampling technique where multiple datasets are created by randomly sampling with replacement from the original data. In bagging, bootstrapping is used to generate diverse training sets for each decision tree in the ensemble.

How does bagging leverage ensemble learning to improve prediction accuracy?
Bagging utilizes ensemble learning by combining the predictions of multiple decision trees trained on different bootstrap samples. This approach helps to mitigate individual tree errors and biases, as the final prediction is based on the consensus or average of multiple models. This aggregation process improves overall prediction accuracy.

Summary

Describe the aggregation process in bagging for both classification and regression tasks.

For classification tasks, the aggregation process in bagging involves majority voting. Each tree in the ensemble makes a prediction, and the class with the most votes across all trees is selected as the final prediction. In regression tasks, the aggregation process typically involves averaging the predictions of all trees to obtain the final prediction.

What are the advantages of bagging over single decision tree models?

Bagging offers several advantages over single decision tree models: increased prediction accuracy by reducing variance and overfitting, improved generalization ability, and robustness to noisy data.

Explain the main drawback of bagging in terms of model interpretability.

A major drawback of bagging is the loss of interpretability compared to single decision trees. While decision trees are inherently interpretable, the aggregation process in bagging creates a complex model where the reasoning behind predictions is not easily traceable. This makes it challenging to understand the underlying relationships between variables and predictions.

Summary

Can bagging lead to underfitting? If so, explain why and provide an example.
Yes, bagging can lead to underfitting if the individual decision trees are too shallow. If the trees are not complex enough to capture the underlying patterns in the data, the bagged model may not be able to adequately represent the true relationship between variables, resulting in underfitting. For example, if we are trying to model a complex non-linear relationship with very shallow decision trees, even with many trees, bagging might still underfit the data.

Describe the relationship between the depth of individual trees in a bagged model and the model's tendency to overfit or underfit.
The depth of individual trees in a bagged model plays a significant role in its tendency to overfit or underfit. Deeper trees have higher complexity and can capture more intricate patterns in the data, but they are also more susceptible to overfitting. Shallow trees, on the other hand, have lower complexity and are less prone to overfitting, but they may underfit the data if the true relationship is complex. The optimal tree depth for a bagged model depends on the specific dataset and the balance between bias and variance.