**Gradient Boosted Decision Trees (GBDTs):**

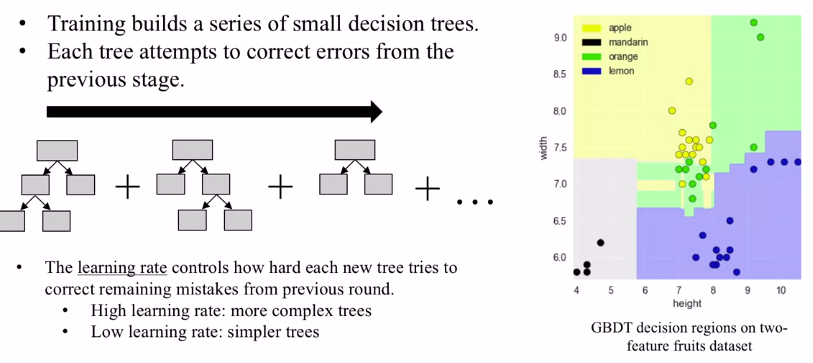
This is another **ensemble** method that has gained wide usage. Like random forests gradient boosted decision trees use an ensemble of decision trees to create a more powerful model for **classification and regression**.

The key difference between random forests and gradient boosted decision trees is that random forests combine randomly different trees in **parallel** while gradient boosted decision trees build trees in **series** where each following trained tree is an **attempt to correct the previous mistakes**.

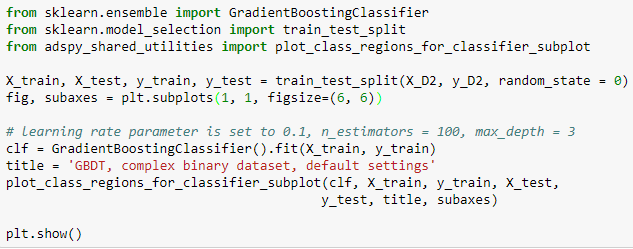
GBDTs use lots of shallow trees known as **weak learner** built in a non-random way to create a model that makes fewer and fewer mistakes as more trees are added.

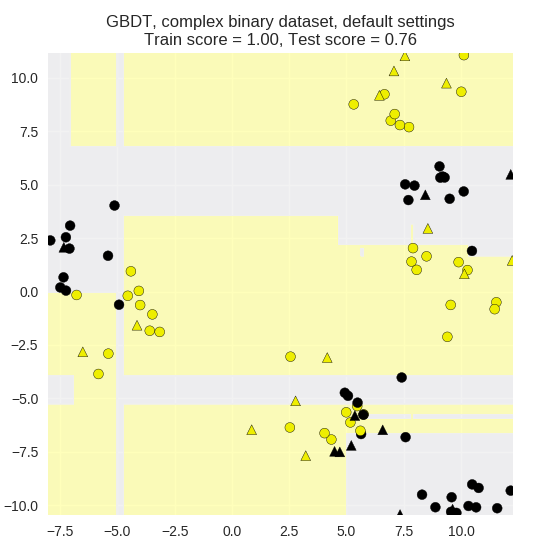
A new parameter we need to use is called the **learning rate** and controls how hard each new tree tires to correct remaining mistakes from the previous round.

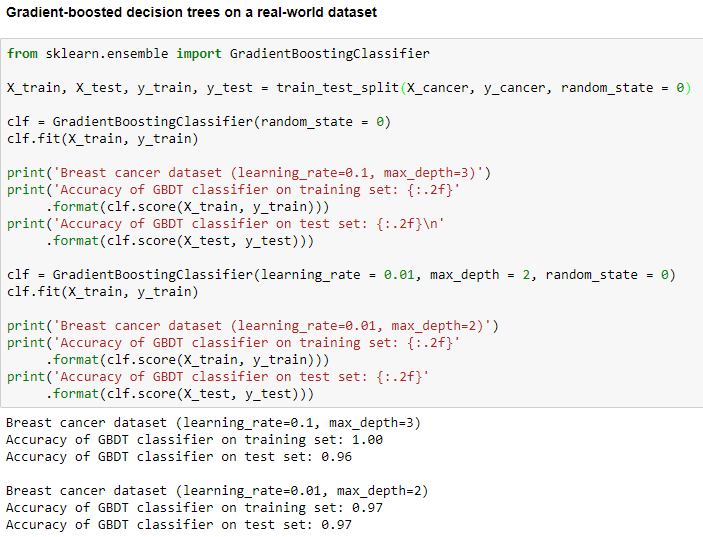
* High learning rate: more complex trees. Potential for overtraining.
* Low learning rate: simpler trees. Potential for undertraining.



**Example:**







Above the first model used with default parameters shows a tendency to overfit the data with a training score of 1. By reducing the learning rate from 0.1 to 0.01 and max depth from 3 to 2 we can see that the model performs better on the test data.

**By reducing the learning rate, we reduce the complexity of the model and allow for a model to generalize better to new data instances. Reducing the max depth also reduces the complexity.**

