

Machine Learning crash course

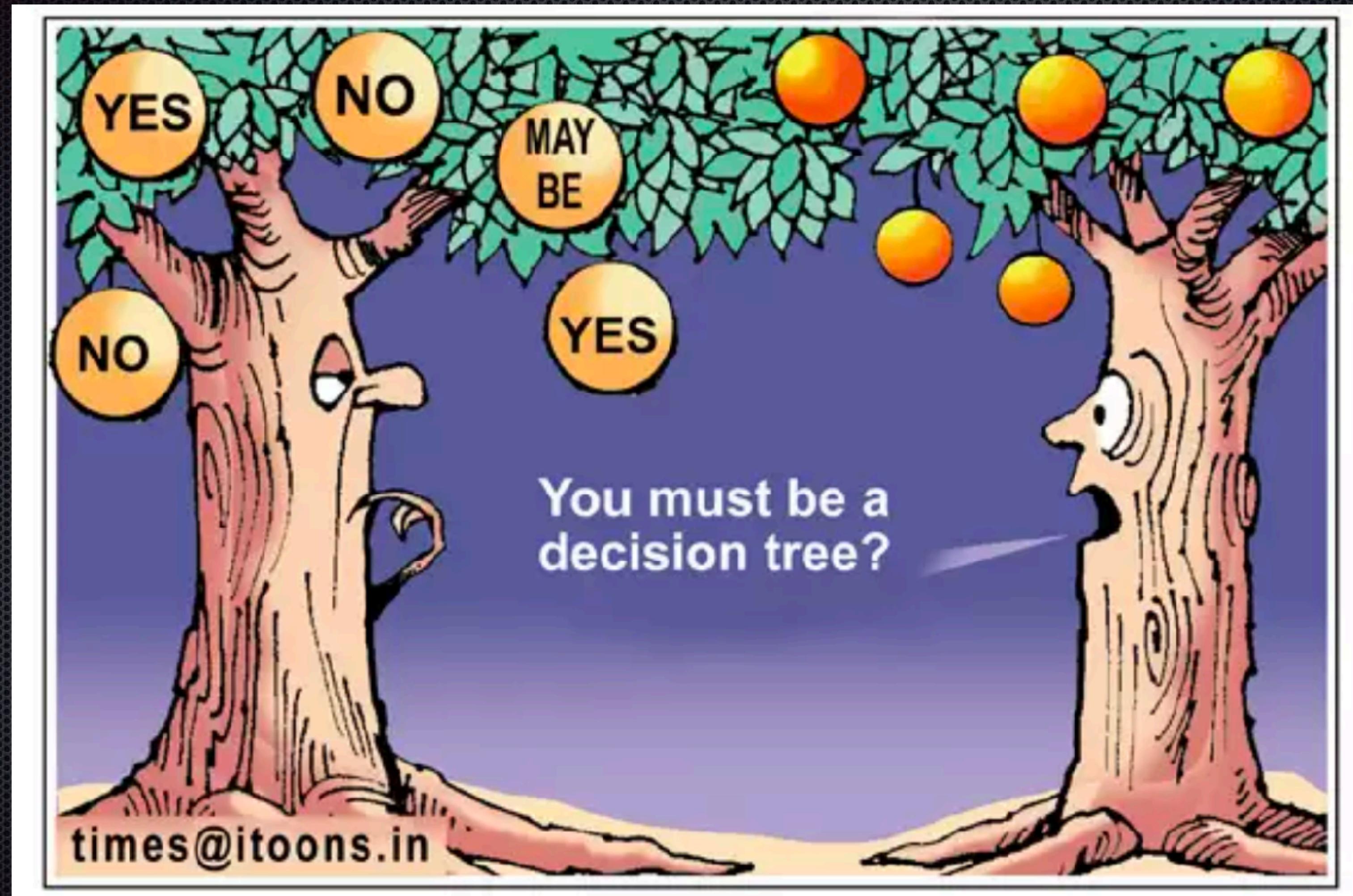
(Part-3)

Ahmed Hammad

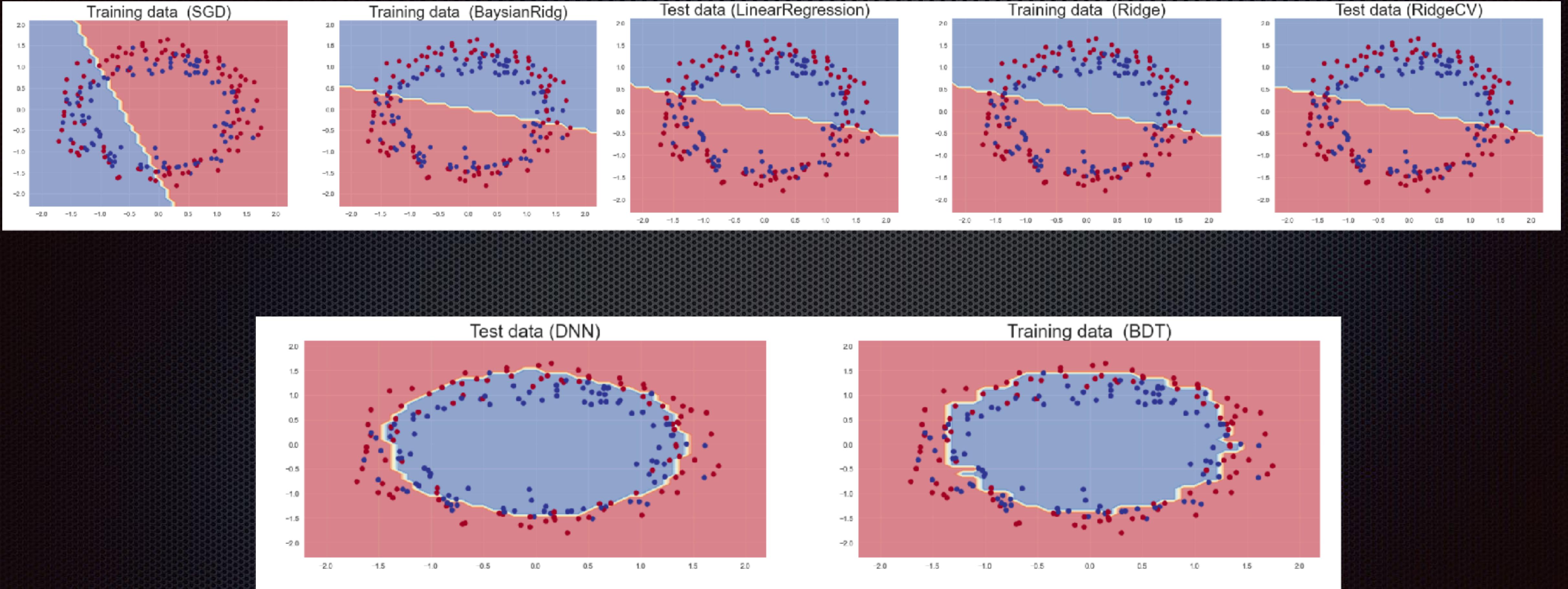
*Ensemble learning, decision trees, random
Forest and boosted decision trees*

Decision Tree

Why do we need the decision tree over the normal one?

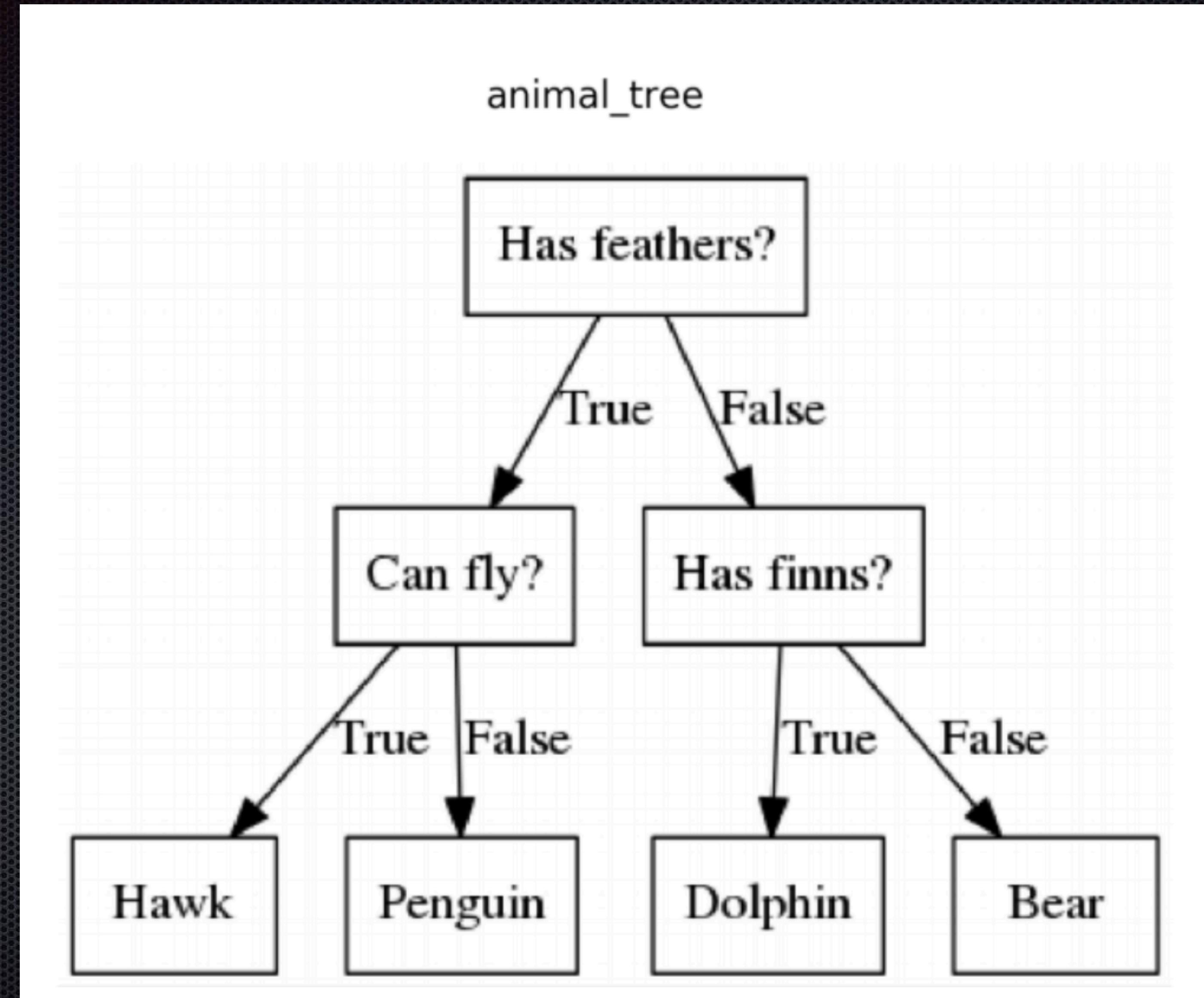


Decision Tree



For non-linear data, more complex machine learning model is needed

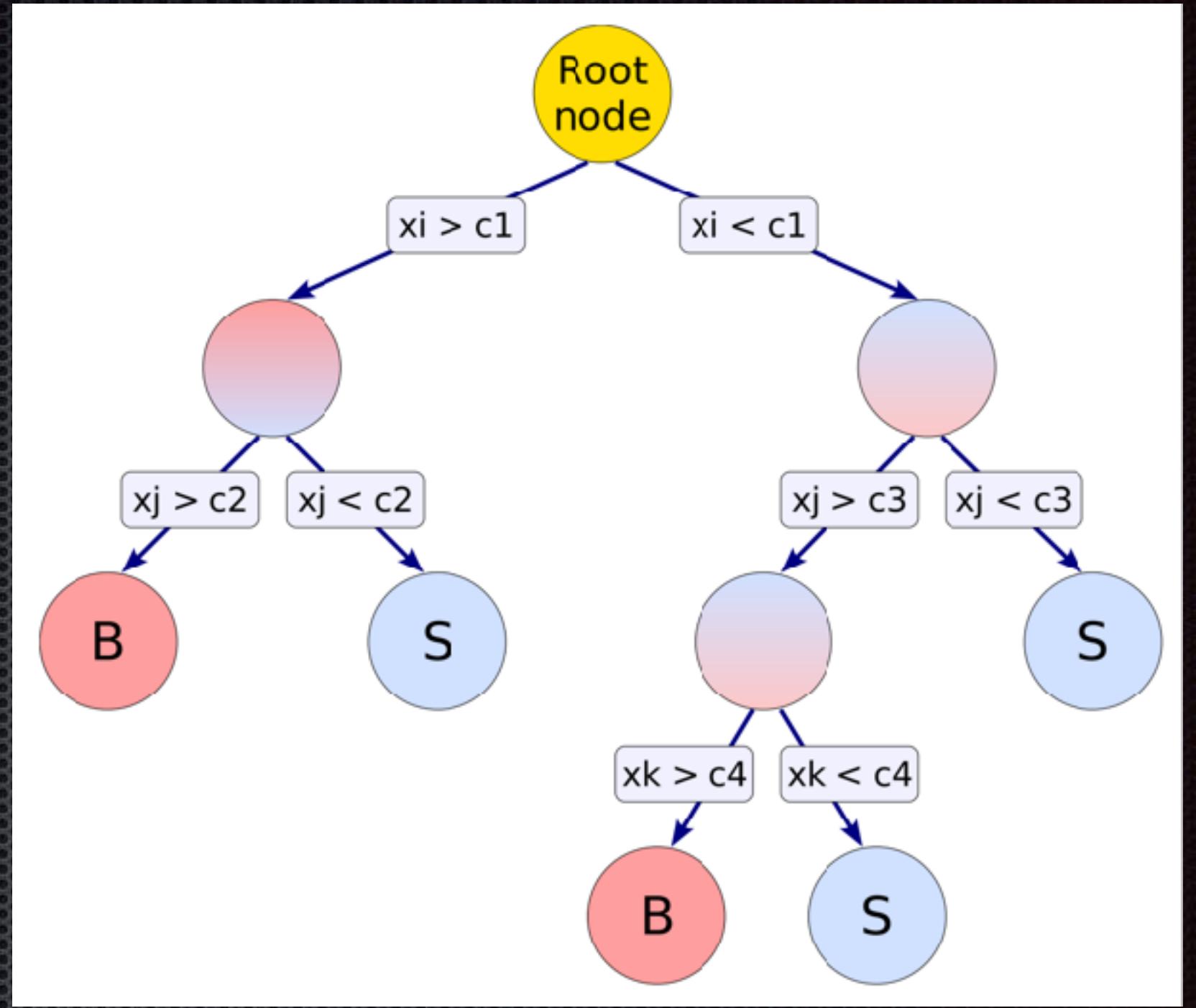
Decision Tree



Decision Tree

Constructing a decision (Regression) Tree as :

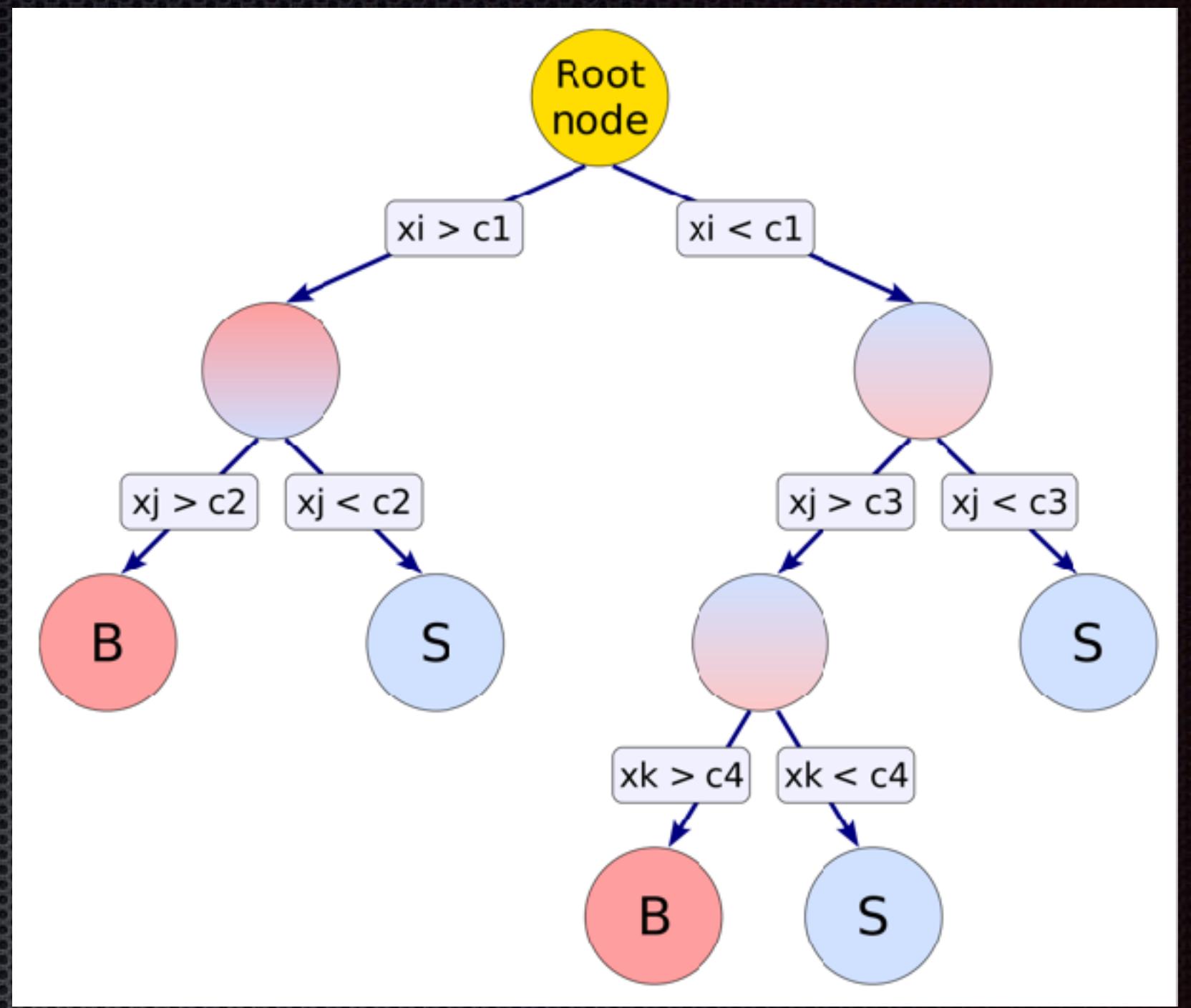
- 1- A root node is created from the sample events
- 2- The algorithm find the optimal threshold that gives the best separation between signal and background by calculating the Loss function
- 3- Once the threshold is found, the node split into two branches
- 4- For the branches that not mostly signal or background (impure) the algorithm consider it as a new node and splits it again.
- 5- Continue splitting the impure branches until we hit a leaf with minimum impurity



Decision Tree

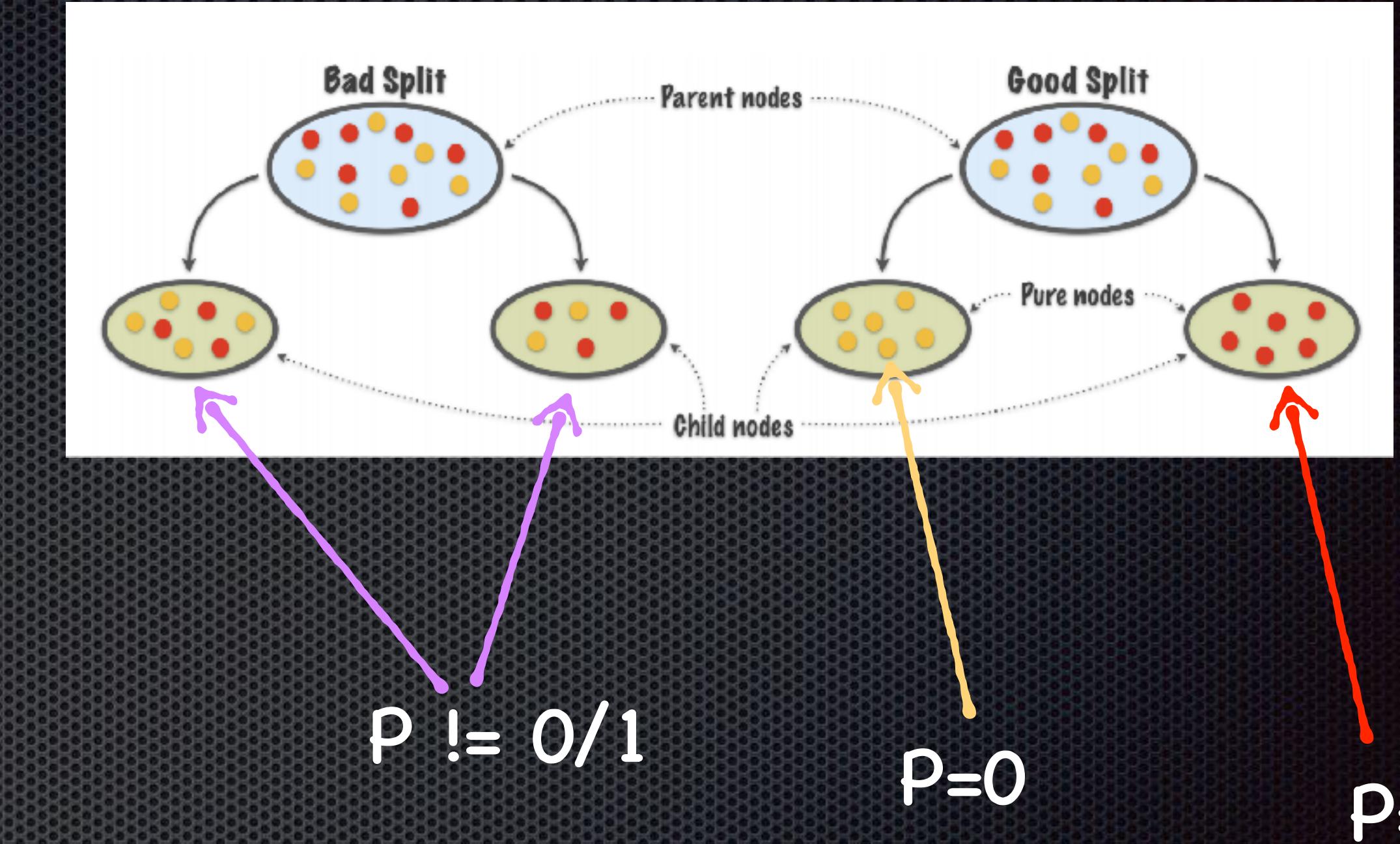
Two questions we need to ask:

- 1- When do we have to split the node
- 2- when do we have to stop the splitting



Splitting criteria

$$\text{Purity (P)} = S / (S+B)$$



Split the tree nodes to maximize the purity in each node
over the splitter node

Splitting criteria

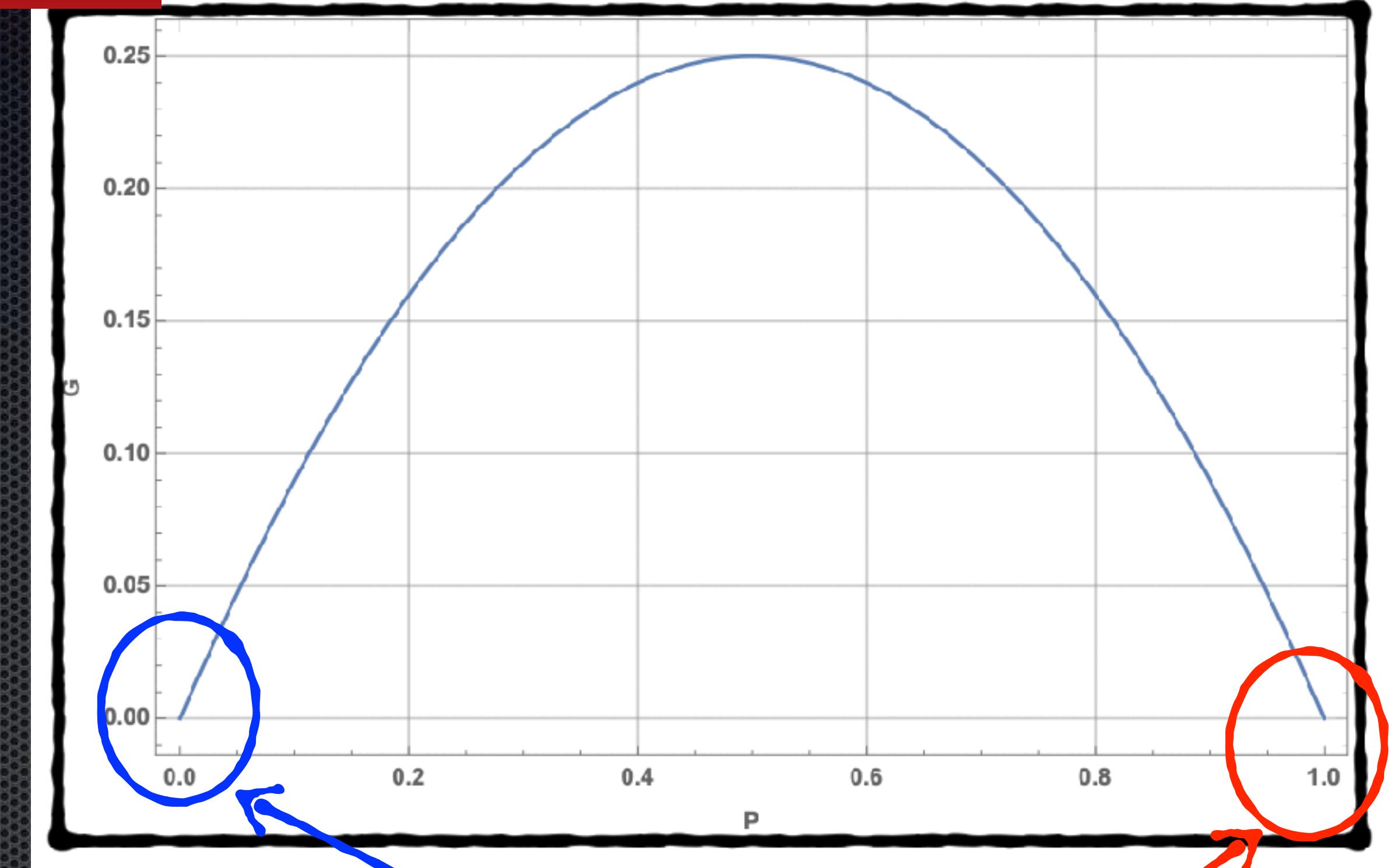
Q1: what is the criteria to split the impure nodes ??



Minimize the Loss function and maximize purity

- *Gini Index* [default], defined by $p \cdot (1 - p)$;
- *Cross entropy*, defined by $-p \cdot \ln(p) - (1 - p) \cdot \ln(1 - p)$;
- *Misclassification error*, defined by $1 - \max(p, 1 - p)$;
- *Statistical significance*, defined by $S/\sqrt{S + B}$;
- *Average squared error*, defined by $1/N \cdot \sum^N (y - \hat{y})^2$ for regression trees where y is the regression target of each event in the node and \hat{y} is its mean value over all events in the node (which would be the estimate of y that is given by the node).

Splitting criteria



Minimize the Gini index with purity > 0.5 or < 0.5
leads two branch with mostly signal like events or background like events

Splitting criteria

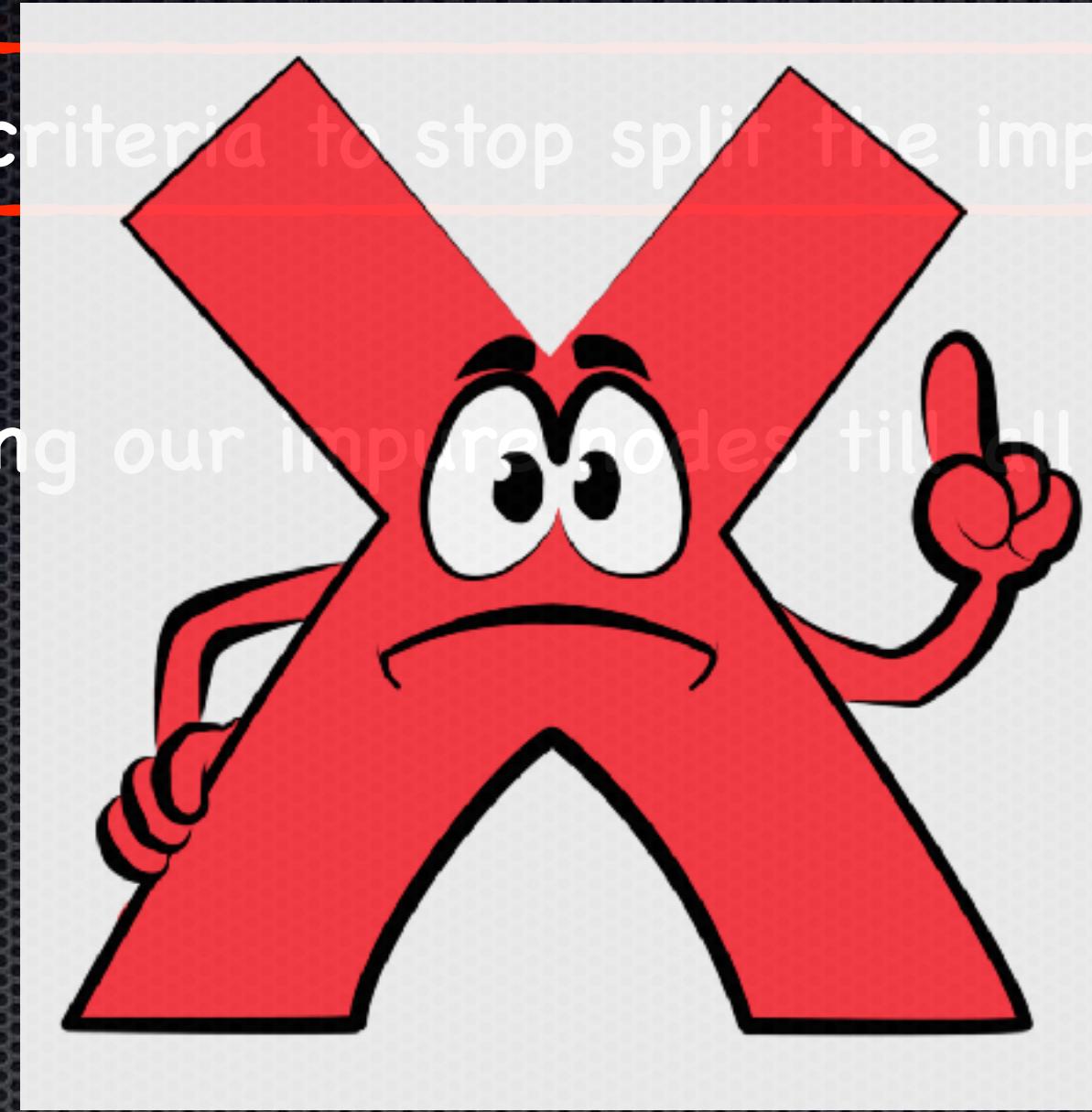
Q2: what is the criteria to stop split the impure nodes ??

Well, in principle we can continue splitting our impure nodes till all events are classified into pure leafs

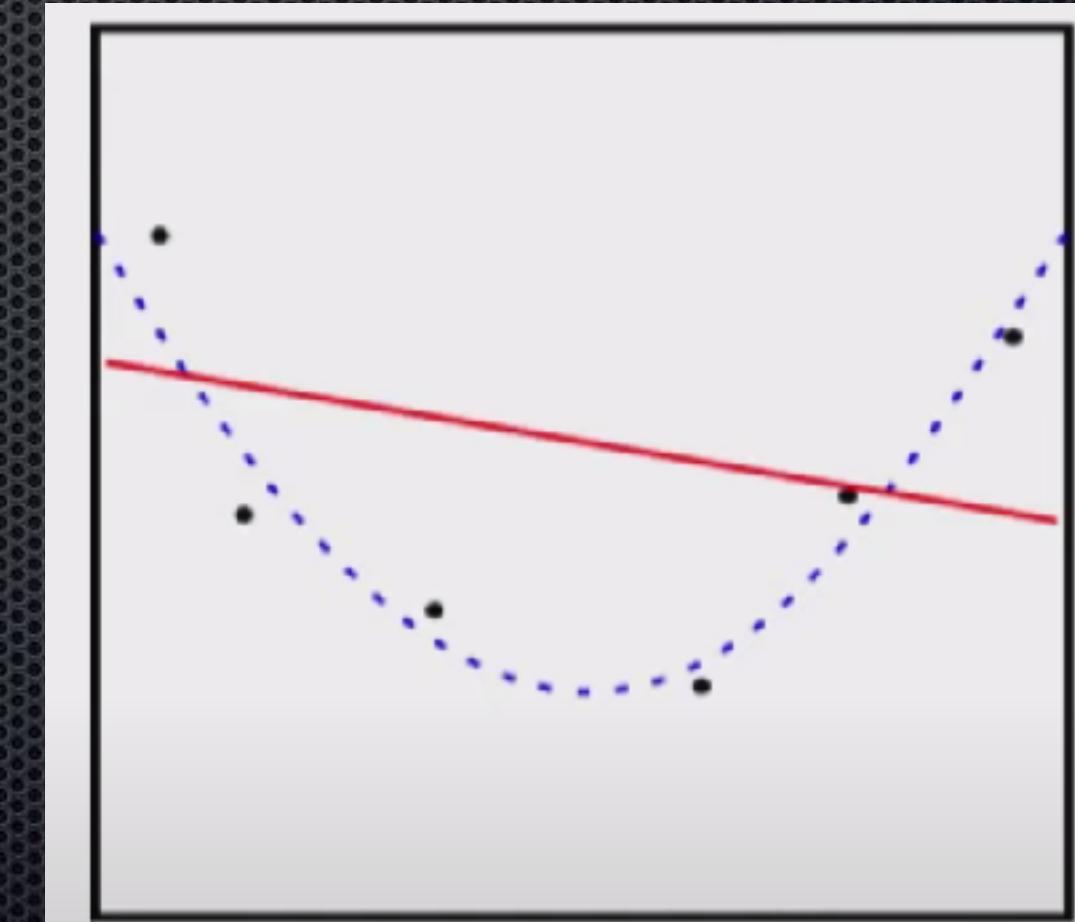
Splitting criteria

Q2: what is the criteria to stop split the impure nodes ??

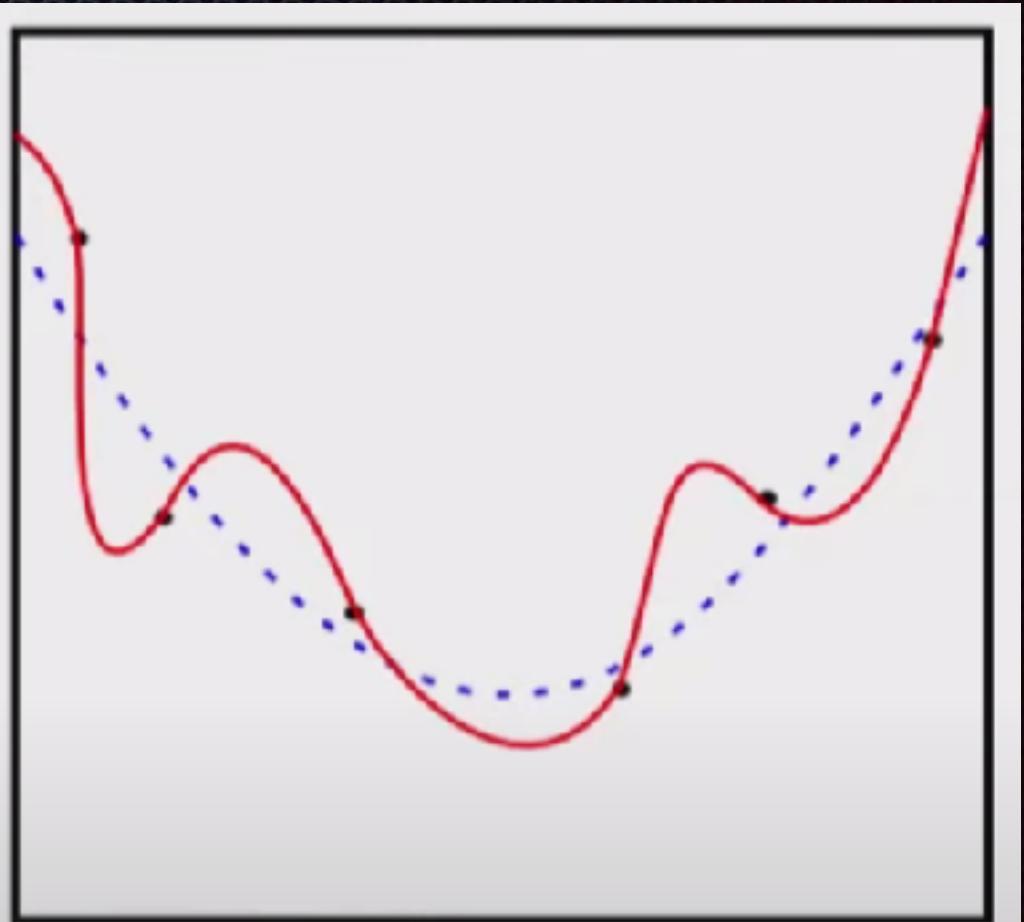
Well, in principle we can continue splitting our impure nodes till all events are classified into pure leafs



Over fitting:
When you over train your model it starts to learn about the noise into the data. Accordingly when you apply it test unknown data it gives wrong answers

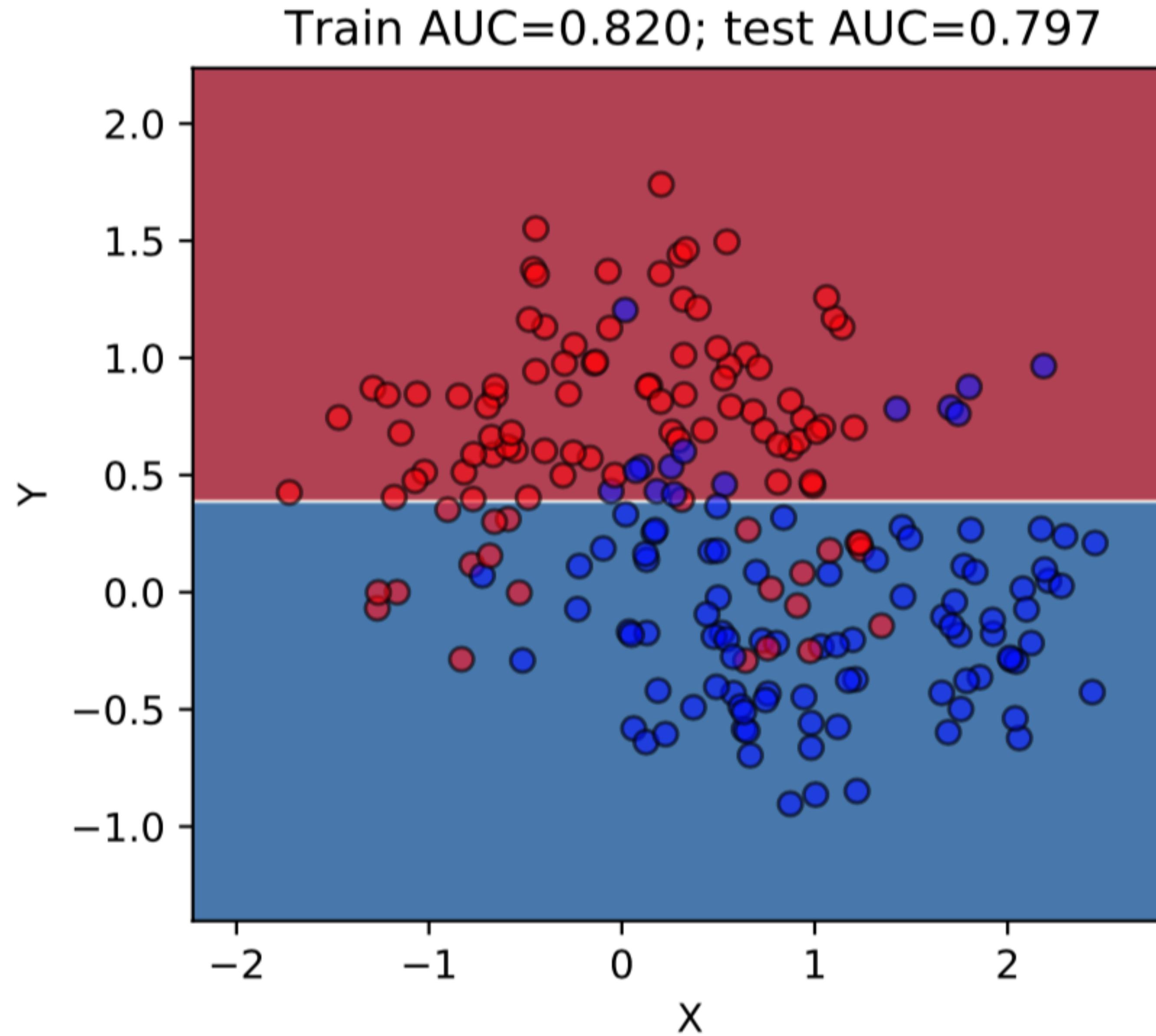
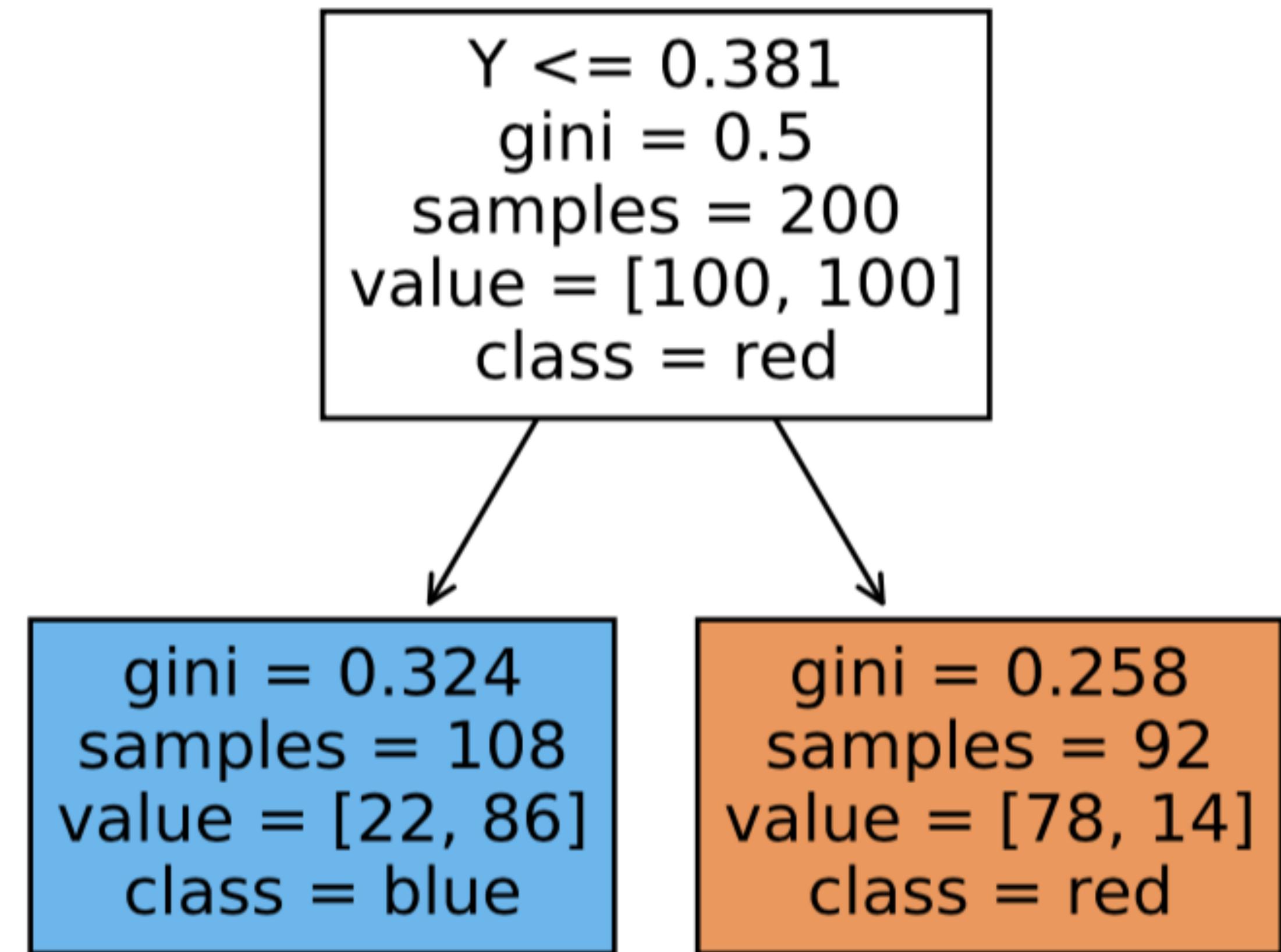


Underfitting

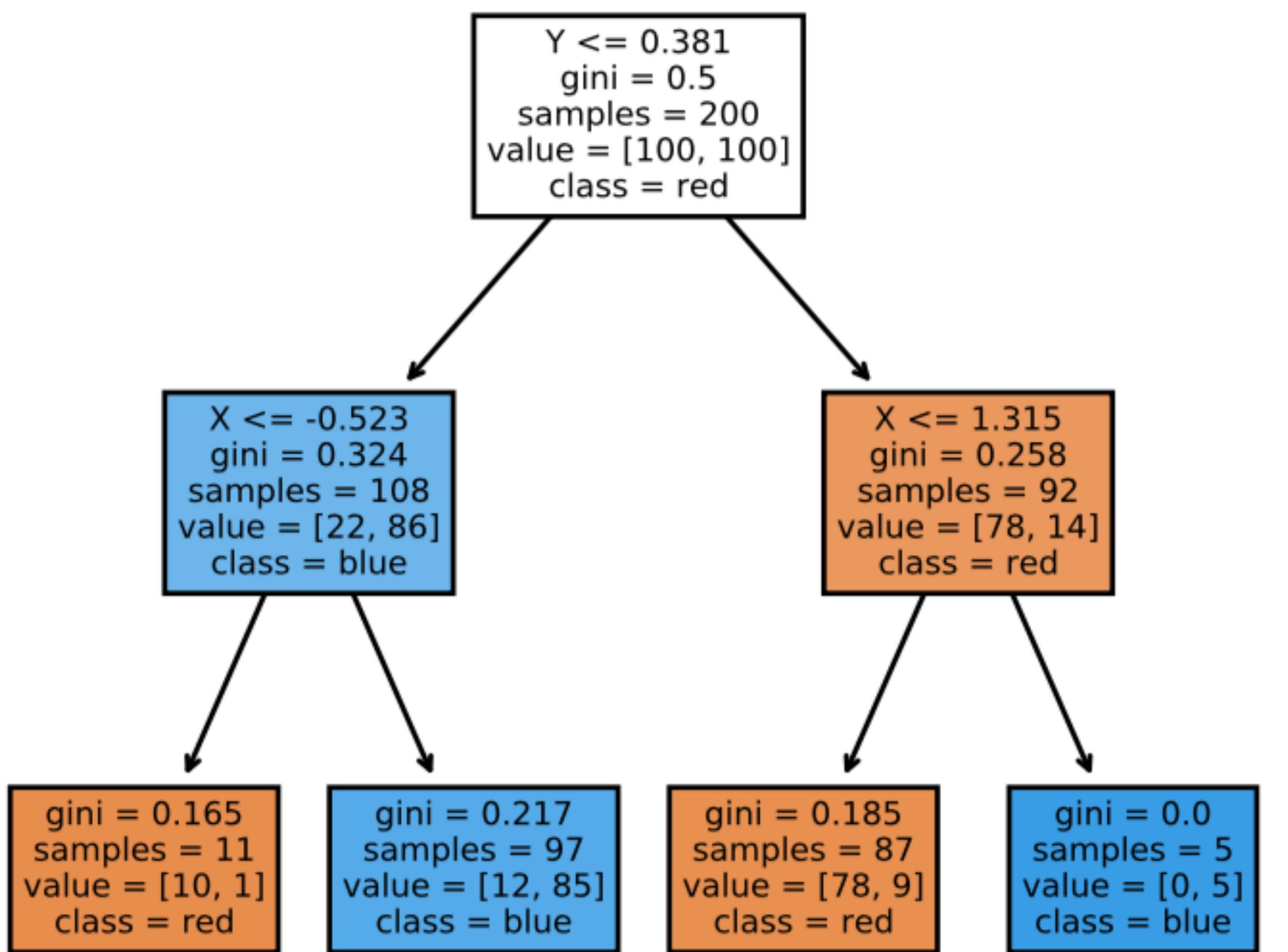


Overfitting

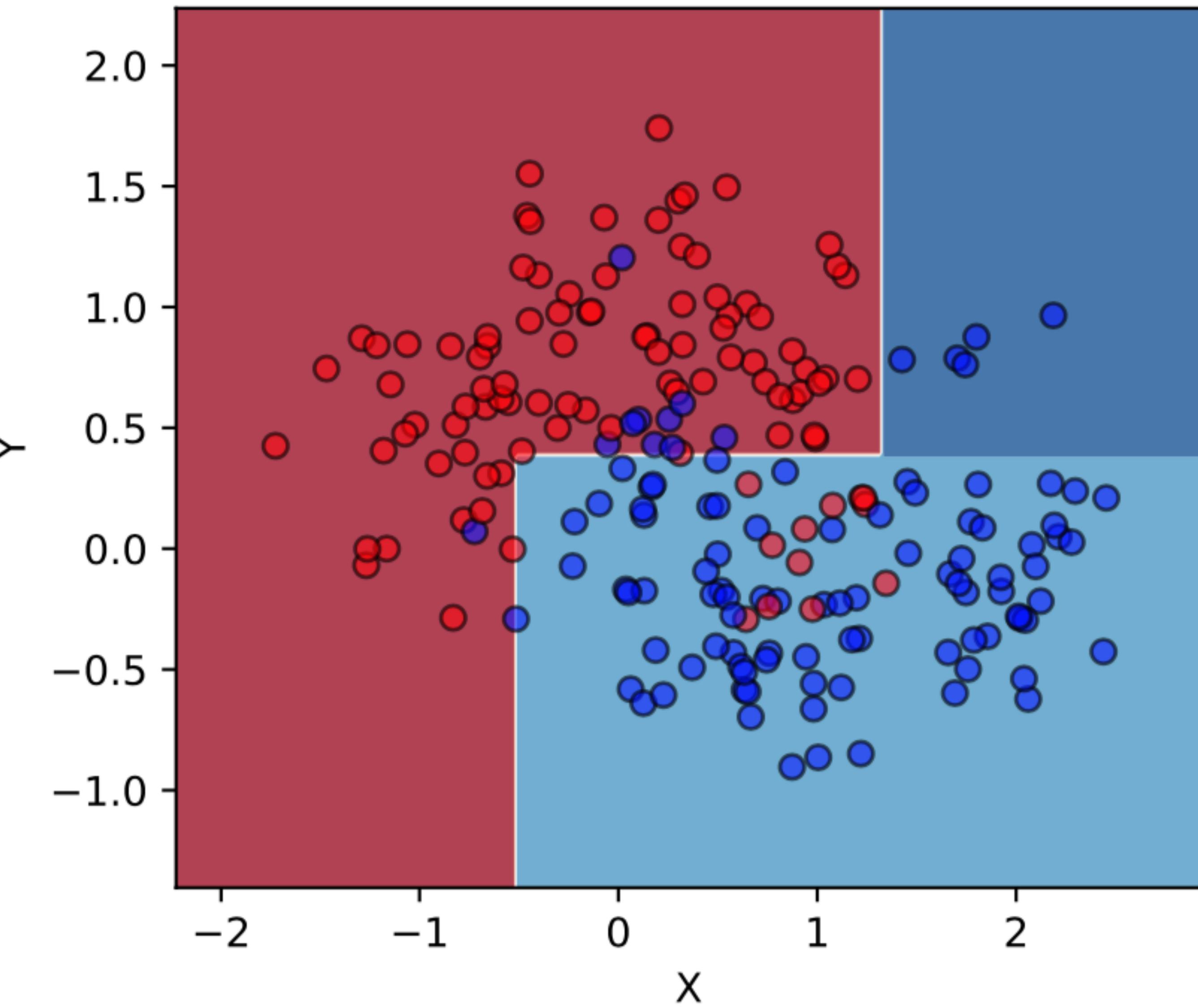
Example



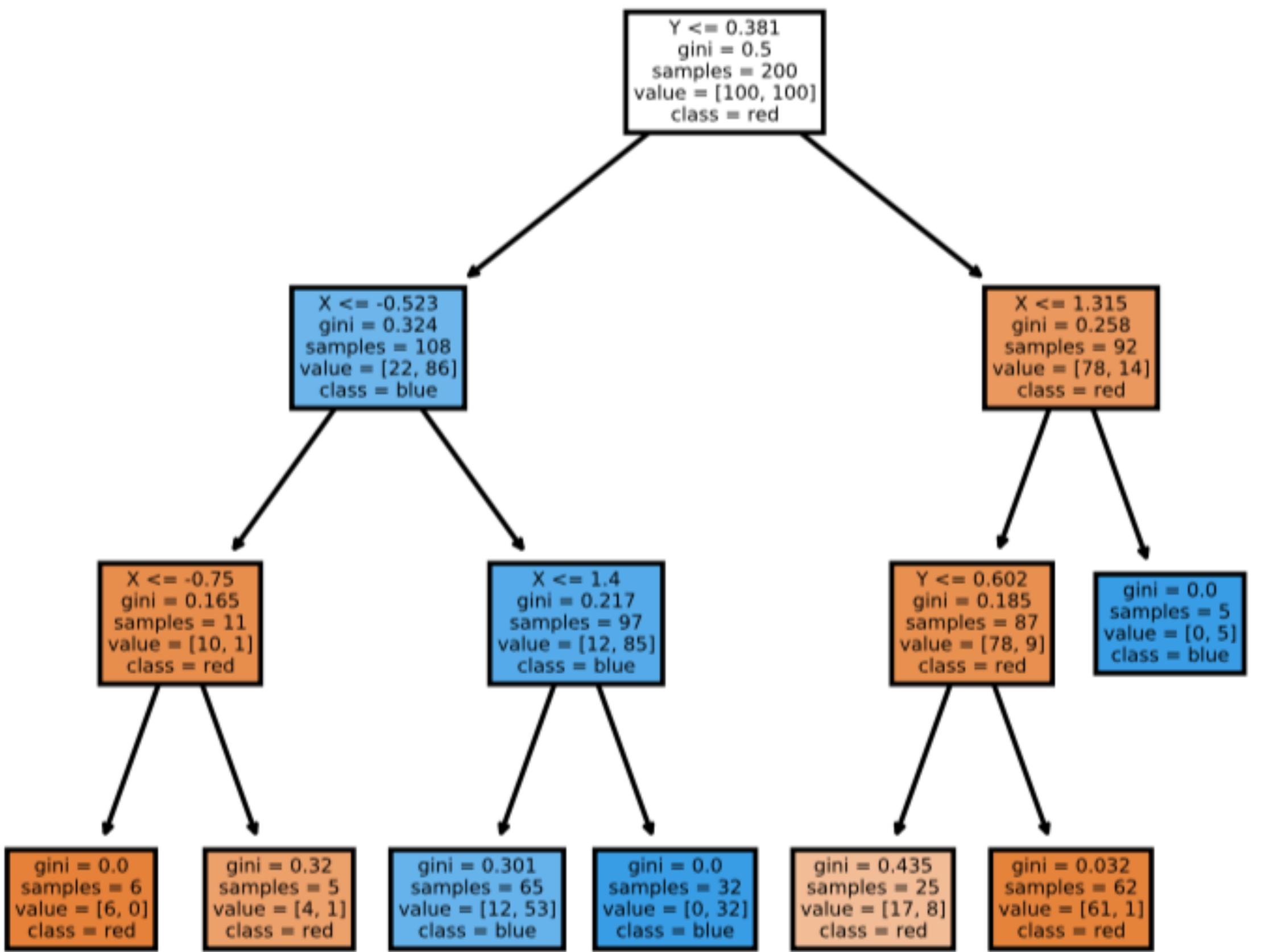
Example



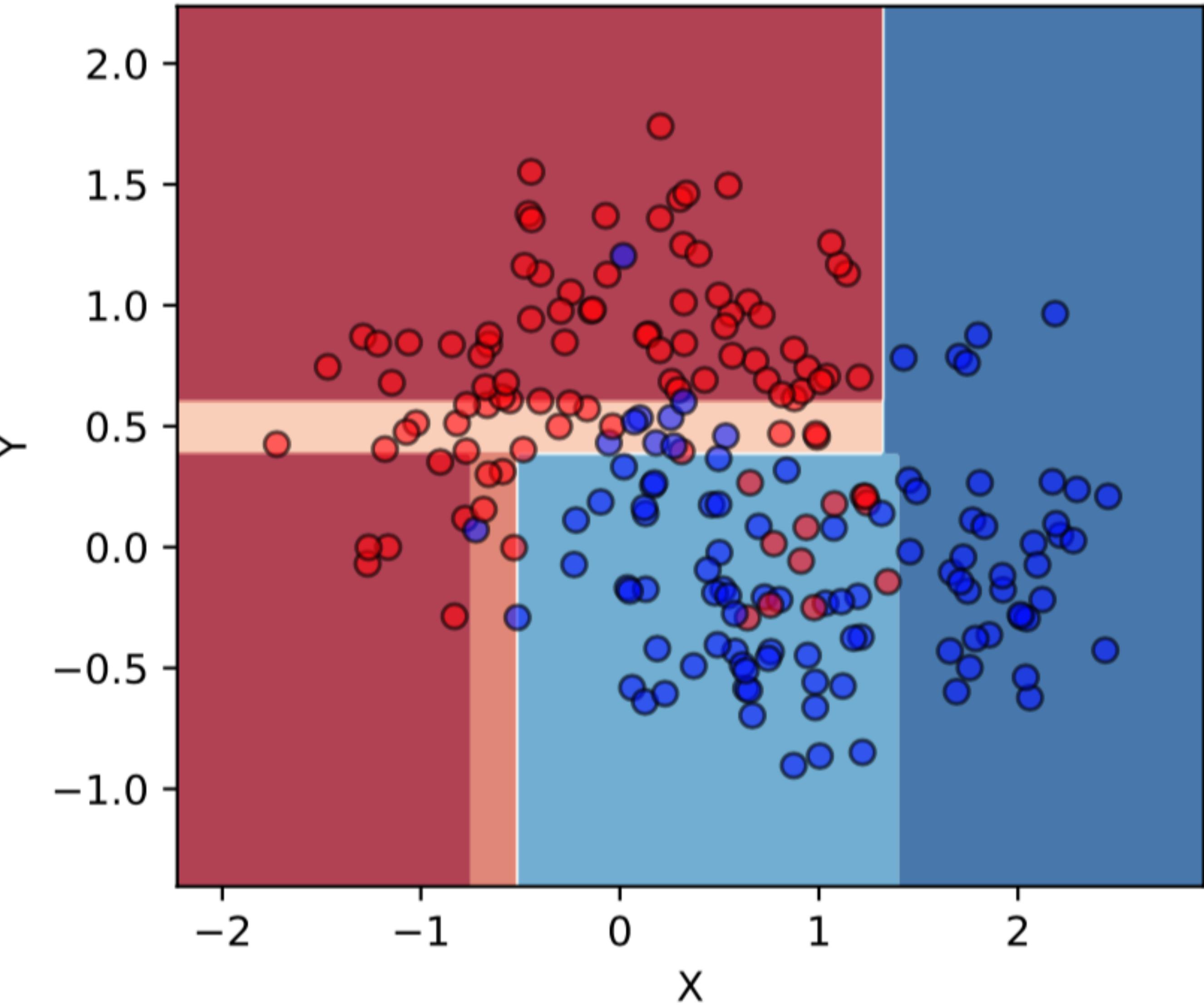
Train AUC=0.894; test AUC=0.885



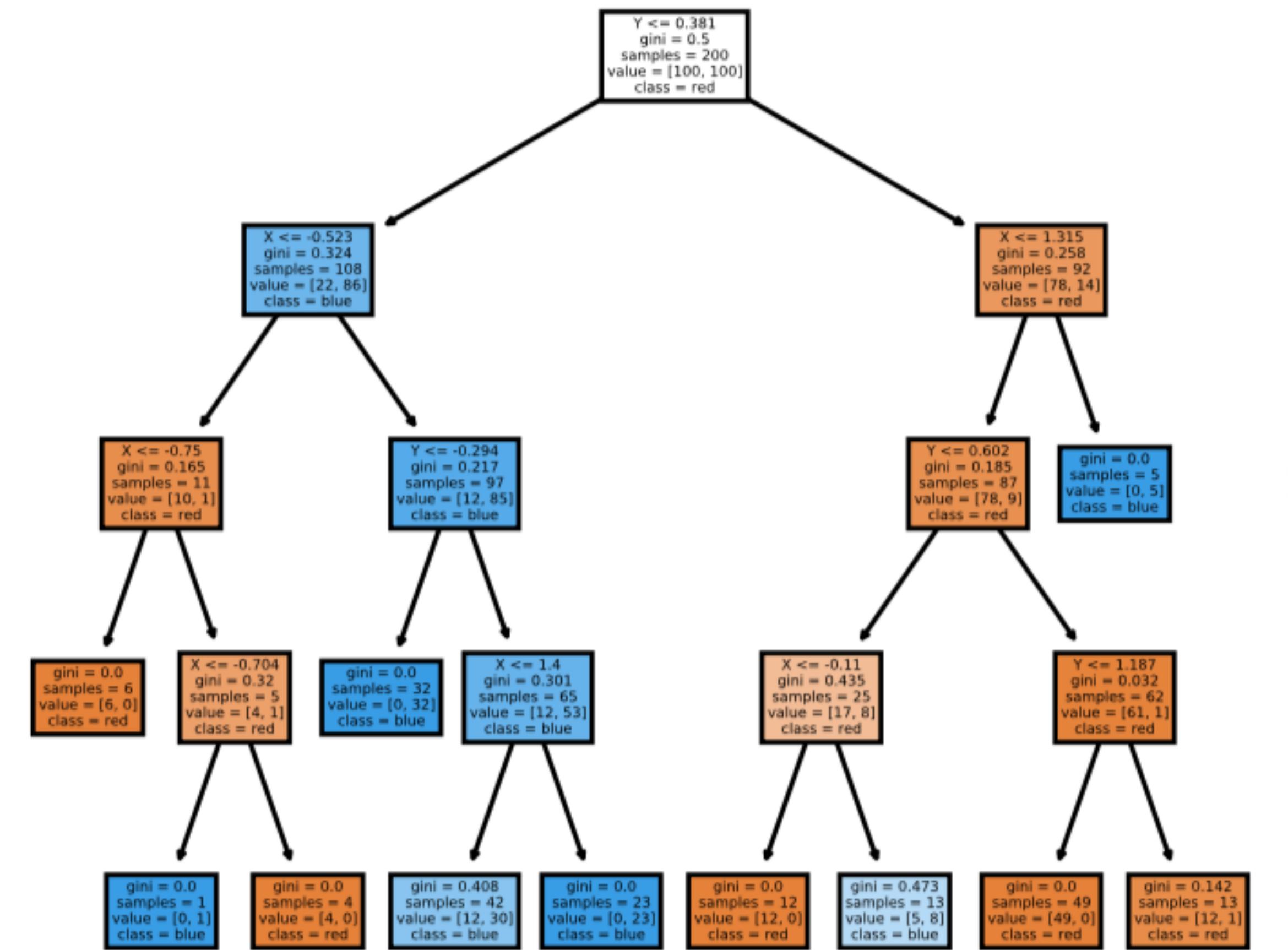
Example



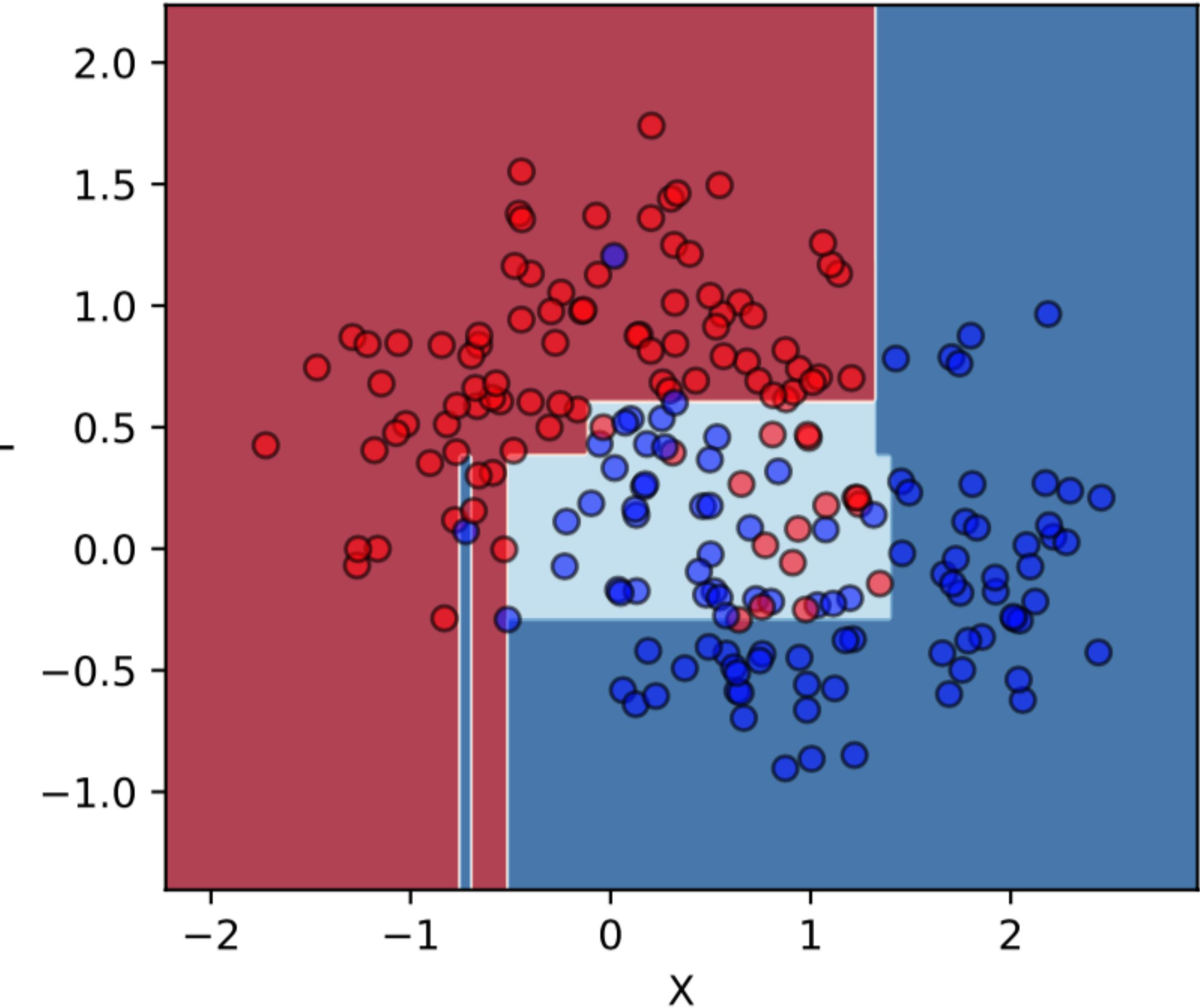
Train AUC=0.942; test AUC=0.916



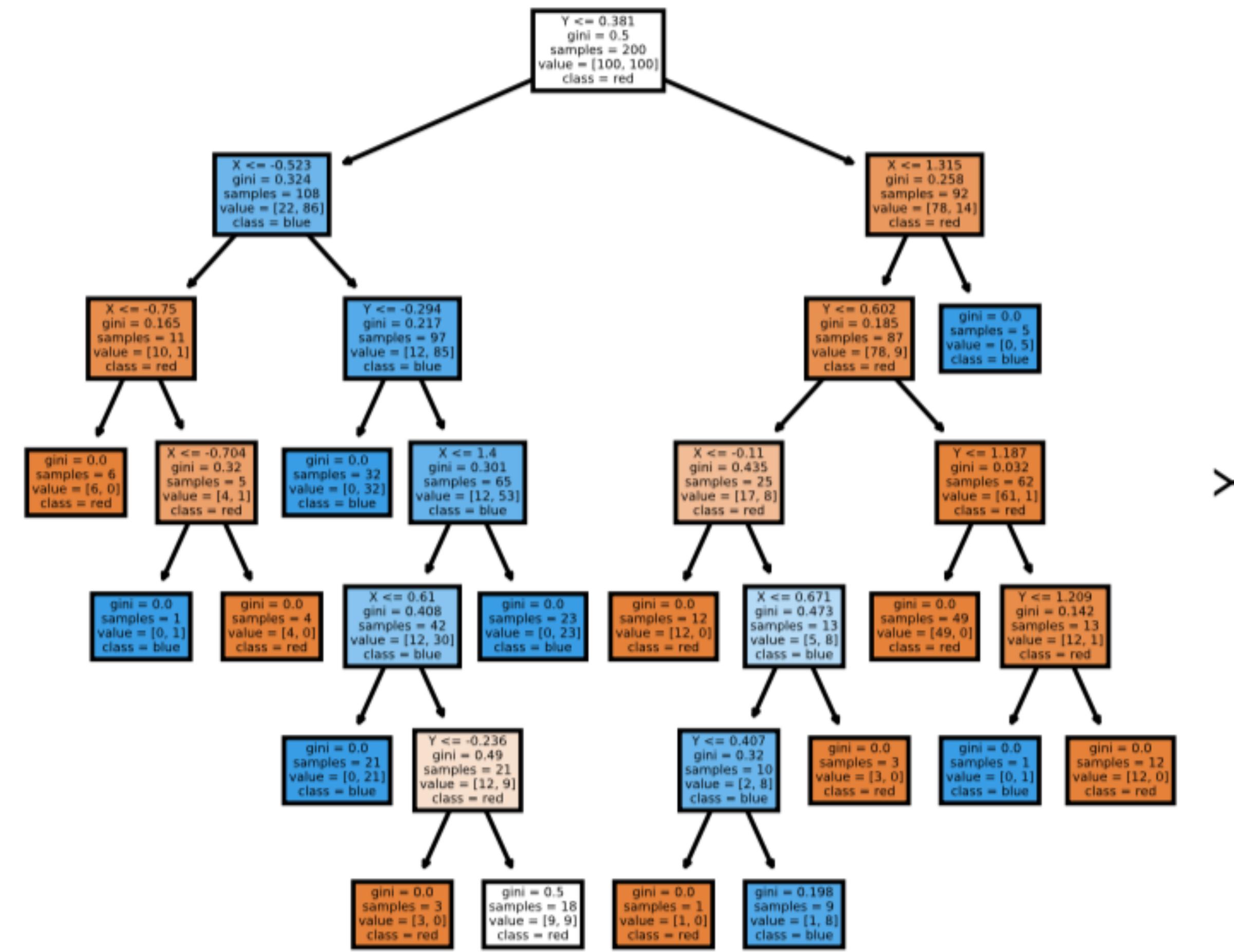
Example



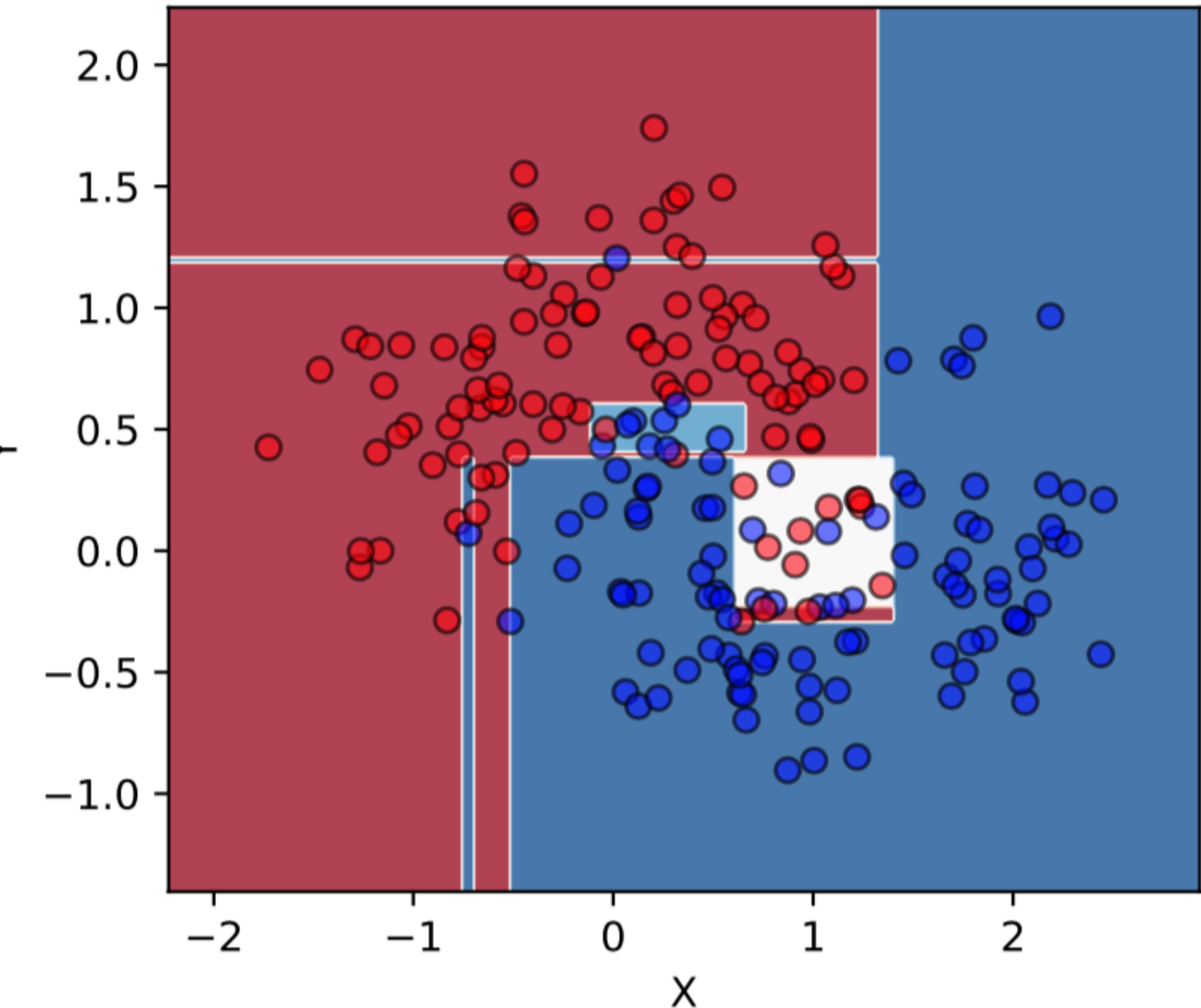
Train AUC=0.968; test AUC=0.925



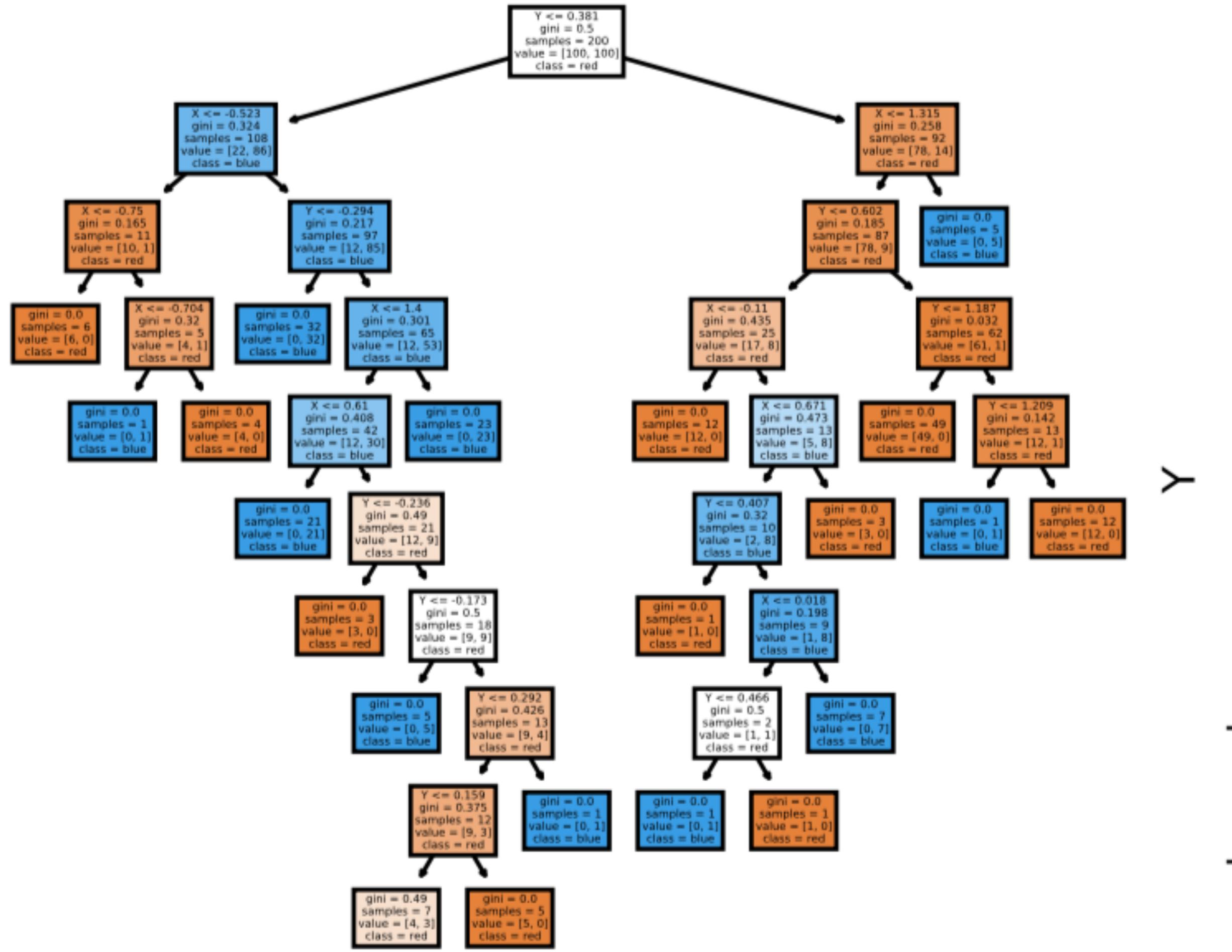
Example



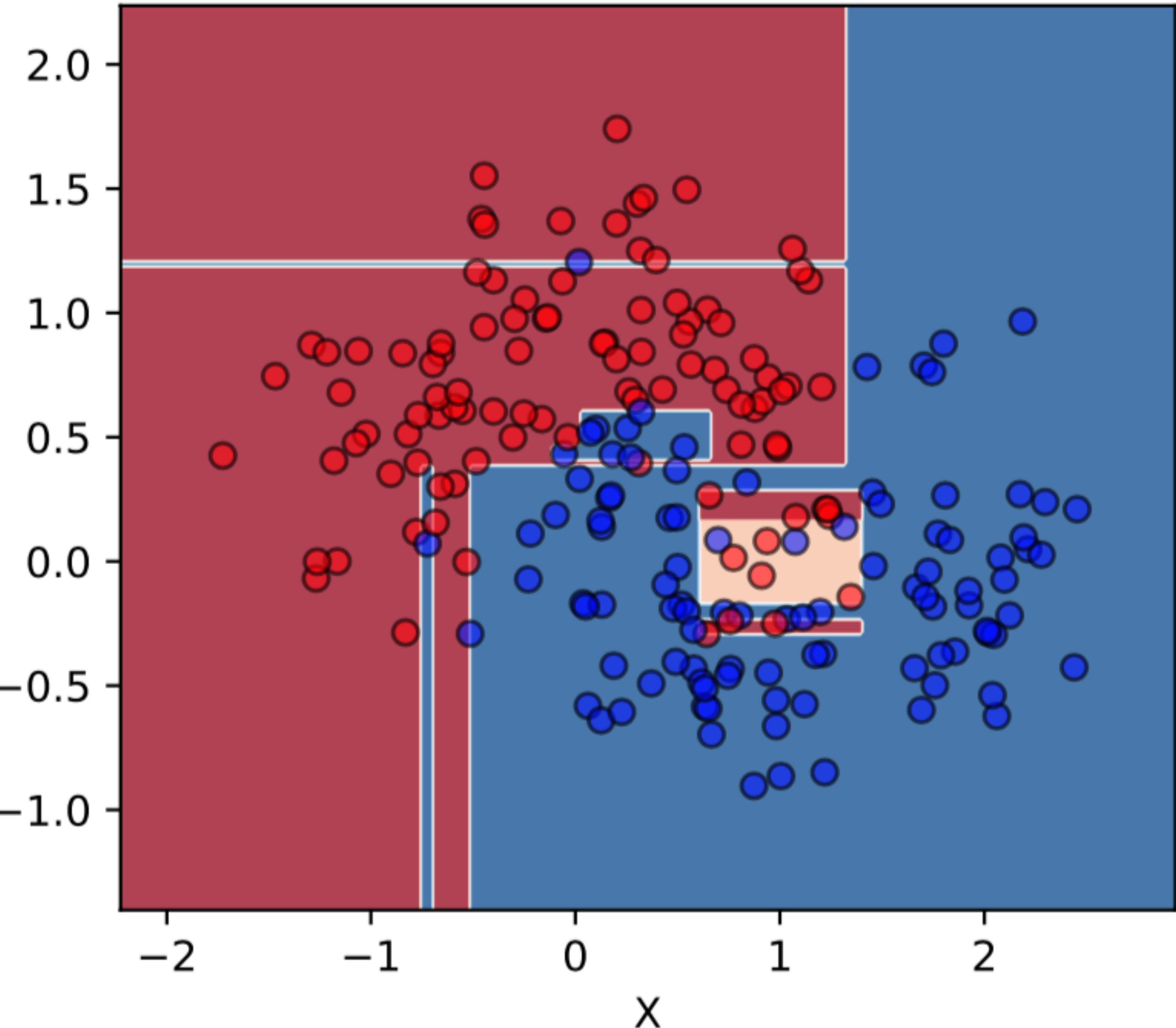
Train AUC=0.995; test AUC=0.912



Example



Train AUC=0.999; test AUC=0.885

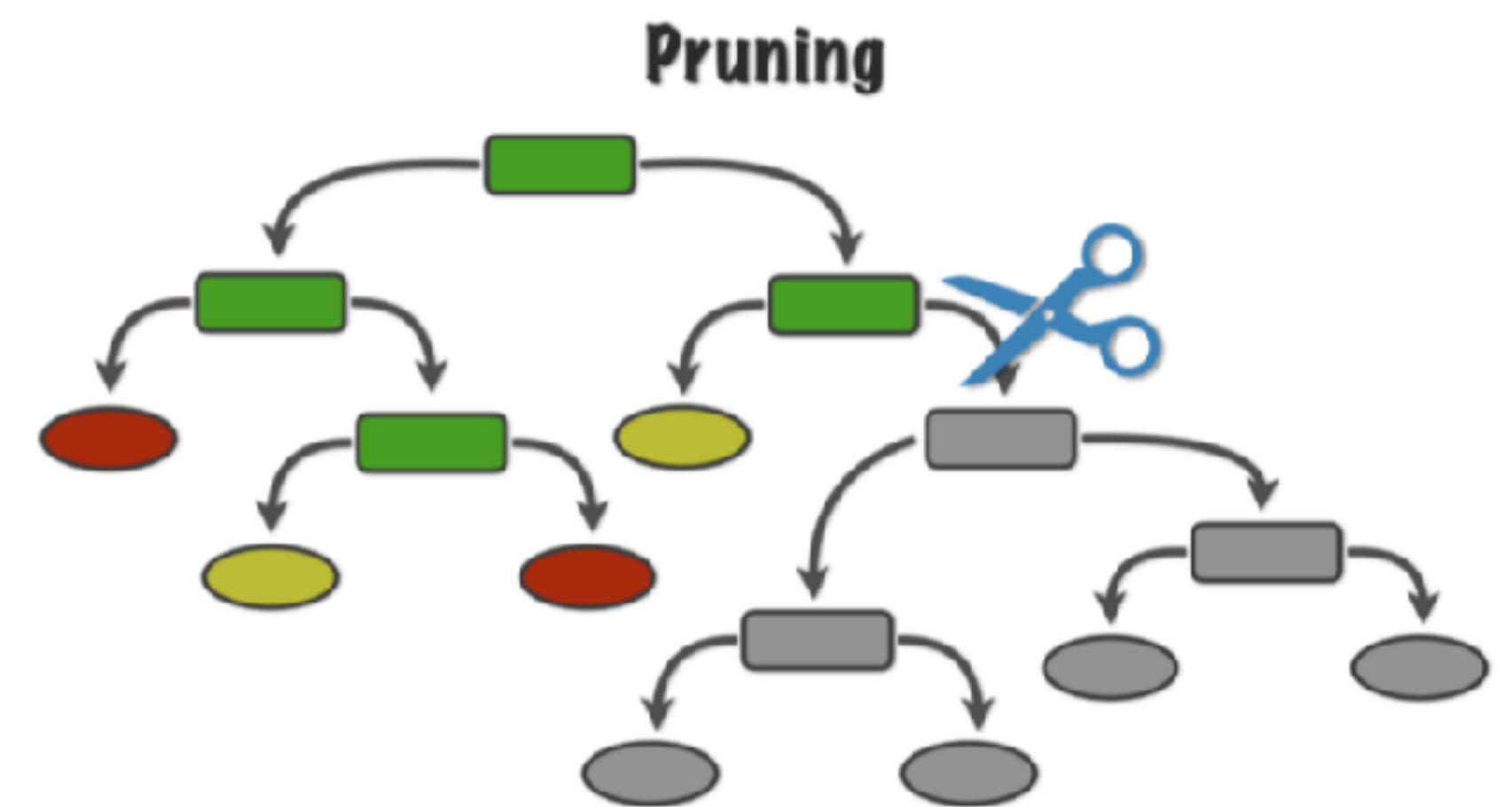


Decision Tree

Two ways to control the Decision tree depth :

1- Pruning the statistically insignificant nodes

2- Optimize the tree depth as a hyper-parameter



Decision tree can always reach a 100% accuracy, but with two main problems:

1- Infinite time for very large data

2- Overfitting for high uncertain data

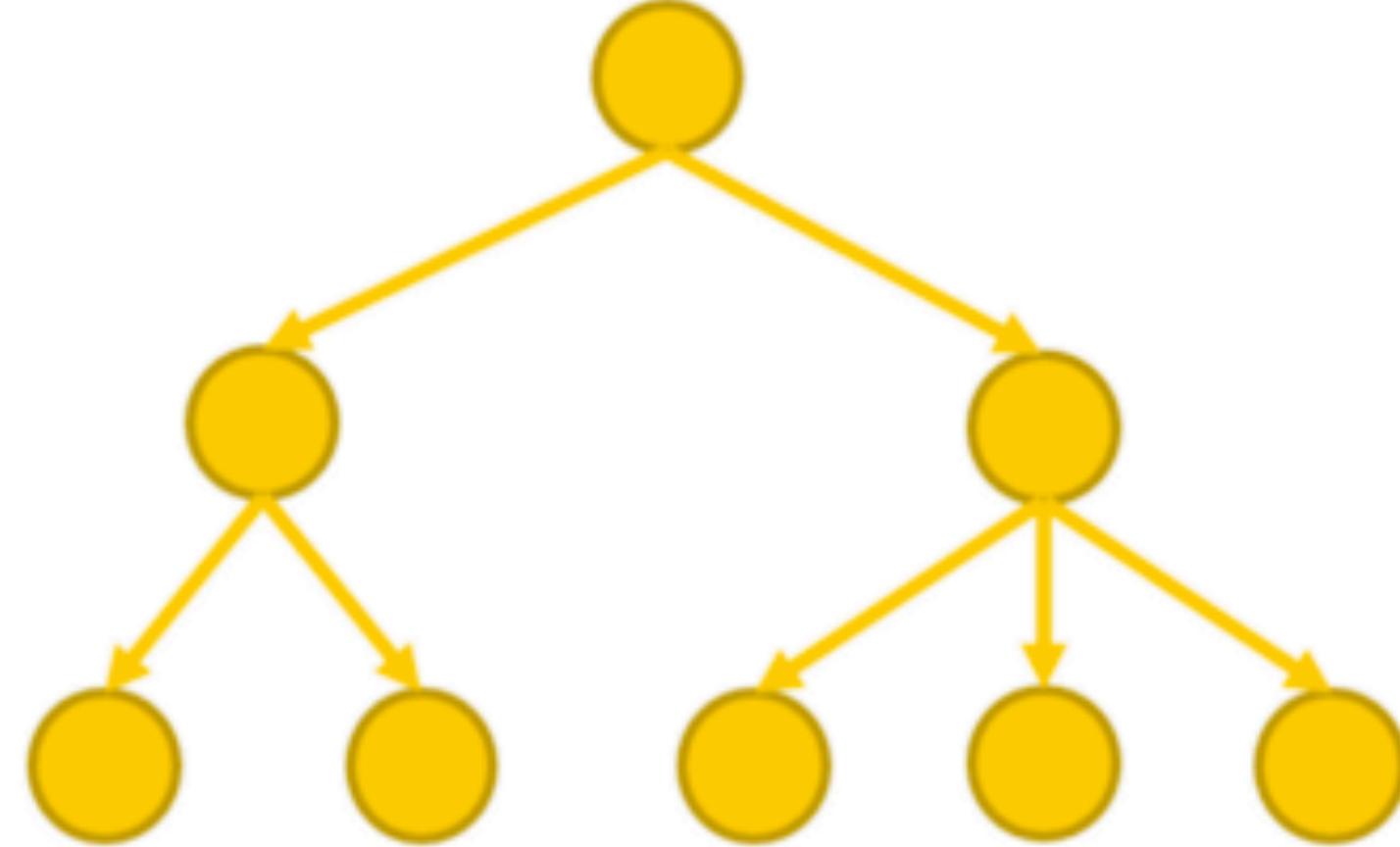
Ensemble Learning

Boosted decision Trees & Random forest

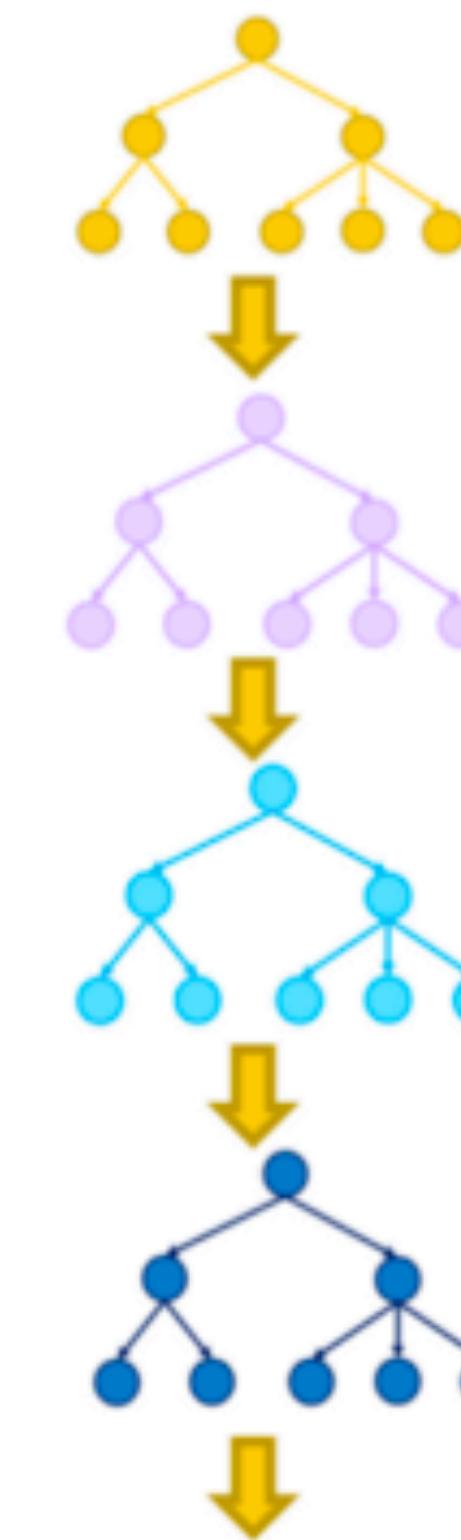


Ensemble learning

Single Decision Tree



Boosted Trees

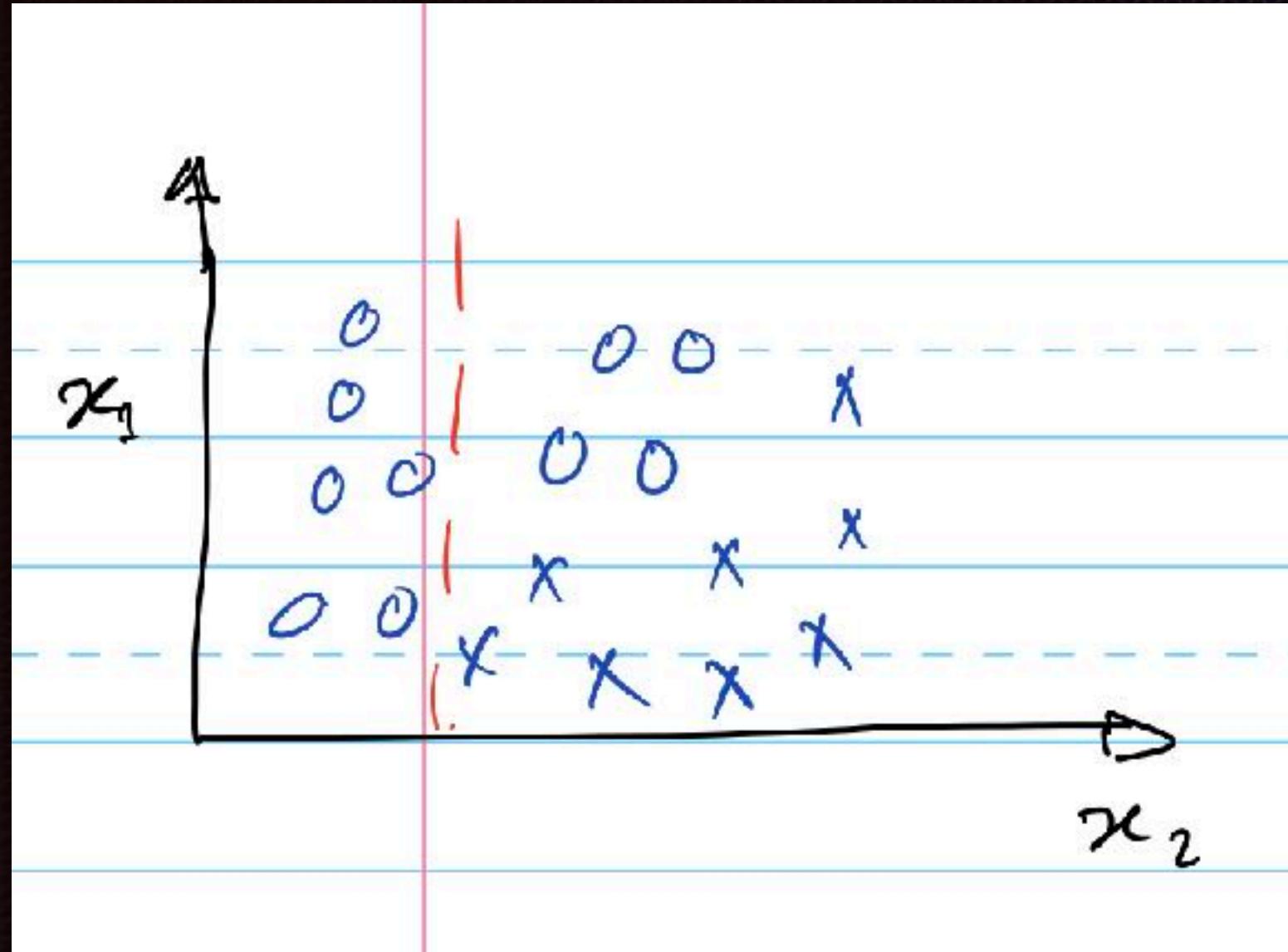


Random Forest



Boosted decision trees

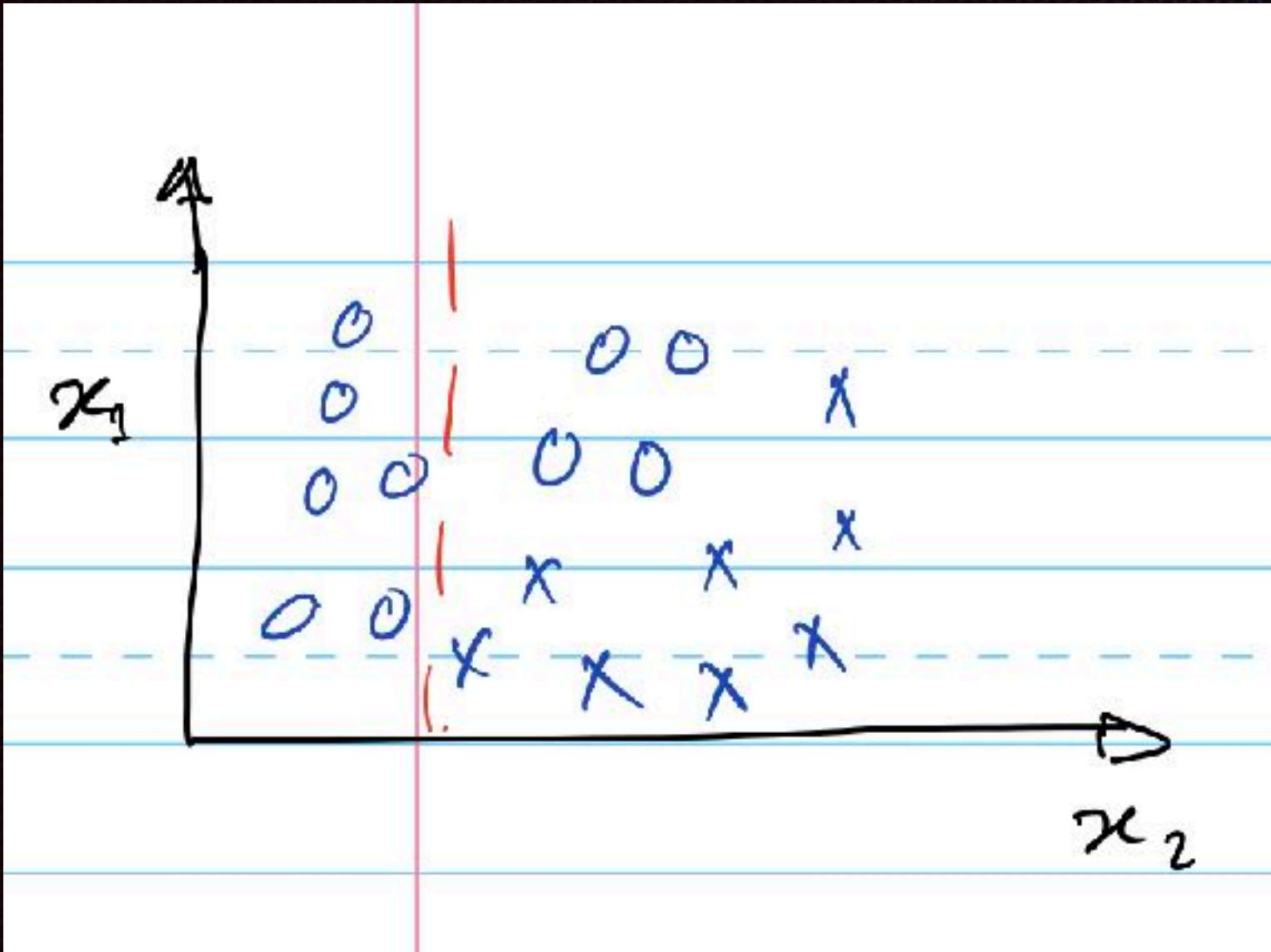
Boosting is a method of combining many weak learners (trees) into a strong classifier.



Start with a decision Tree

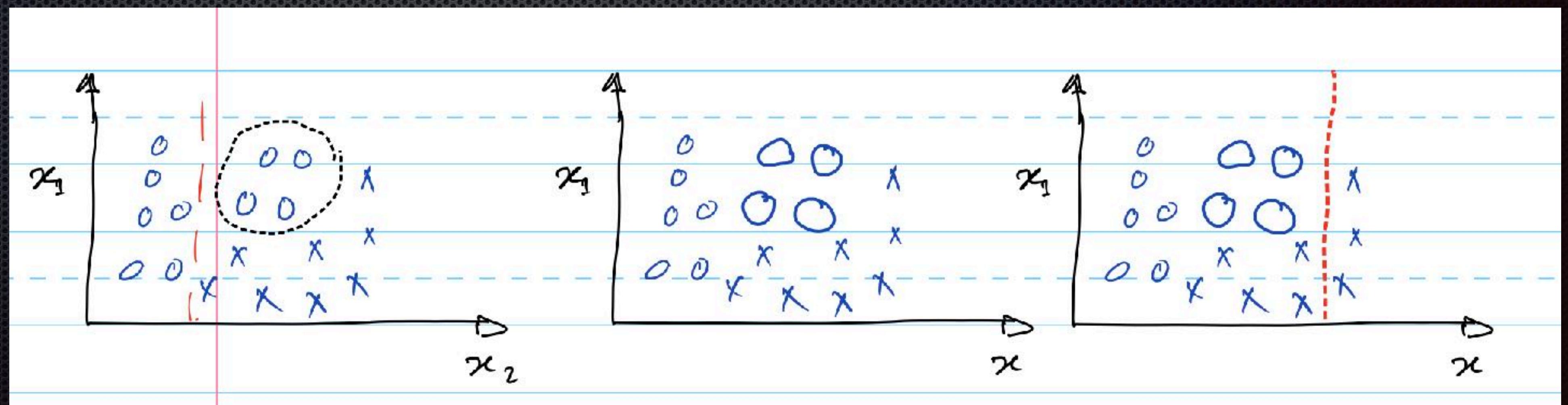
Boosted decision trees

Boosting is a method of combining many weak learners (trees) into a strong classifier.



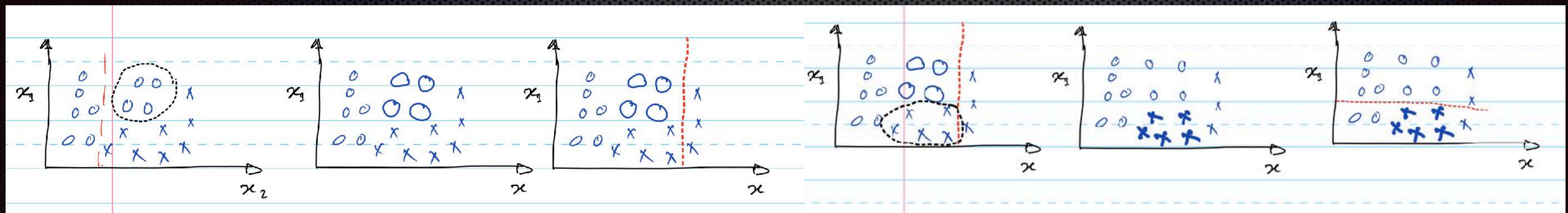
Start with a decision Tree

Boost the classification boundaries for first tree



Boosted decision trees

Boost the purity in
trees nodes sequentially

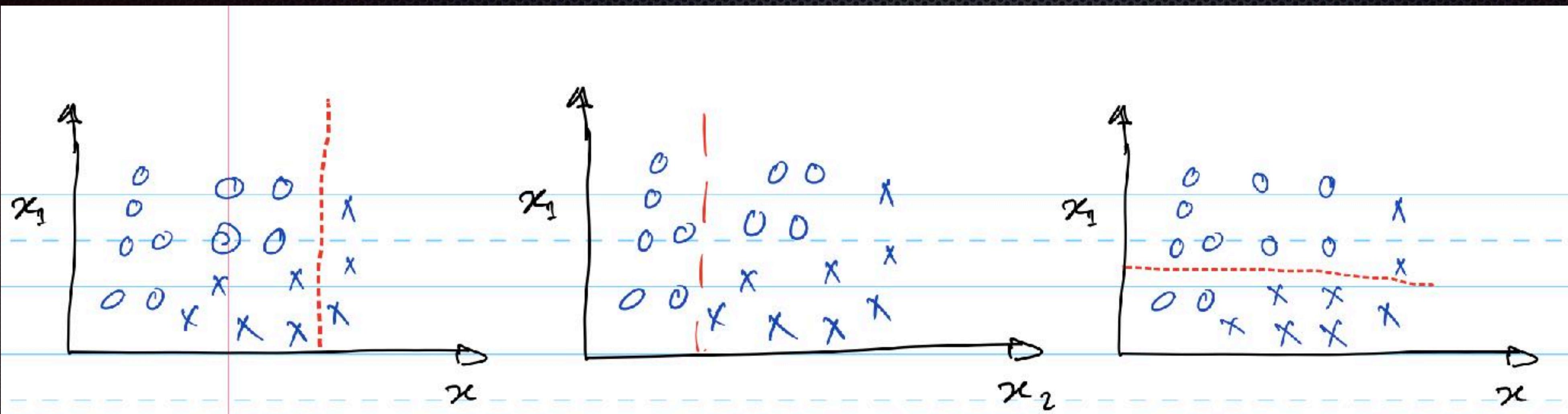


$$F_T(x) = \sum_{t=1}^T f_t(x)$$

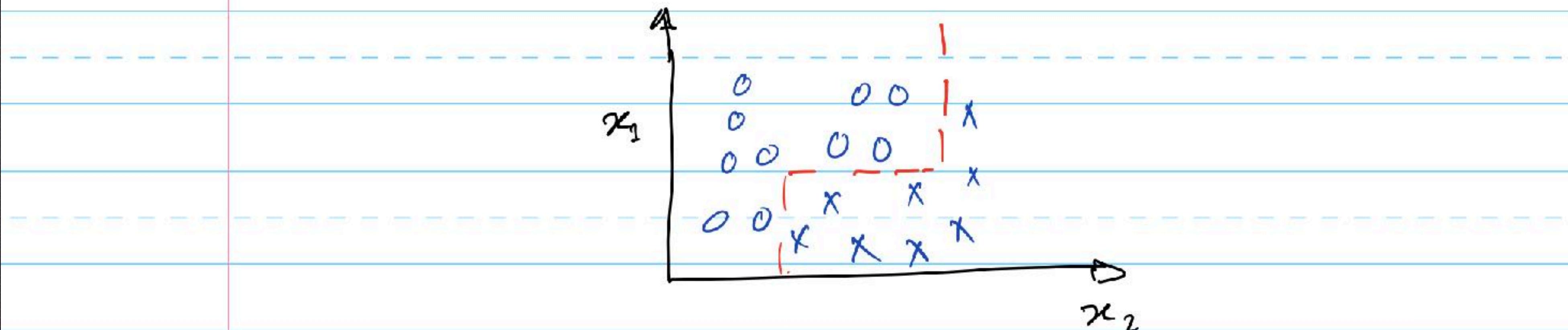
Final classifier is obtained by summing weak learners

$f_t(x)$ is the weak learner that takes x and reutrns the correspondong class

Boosted decision trees

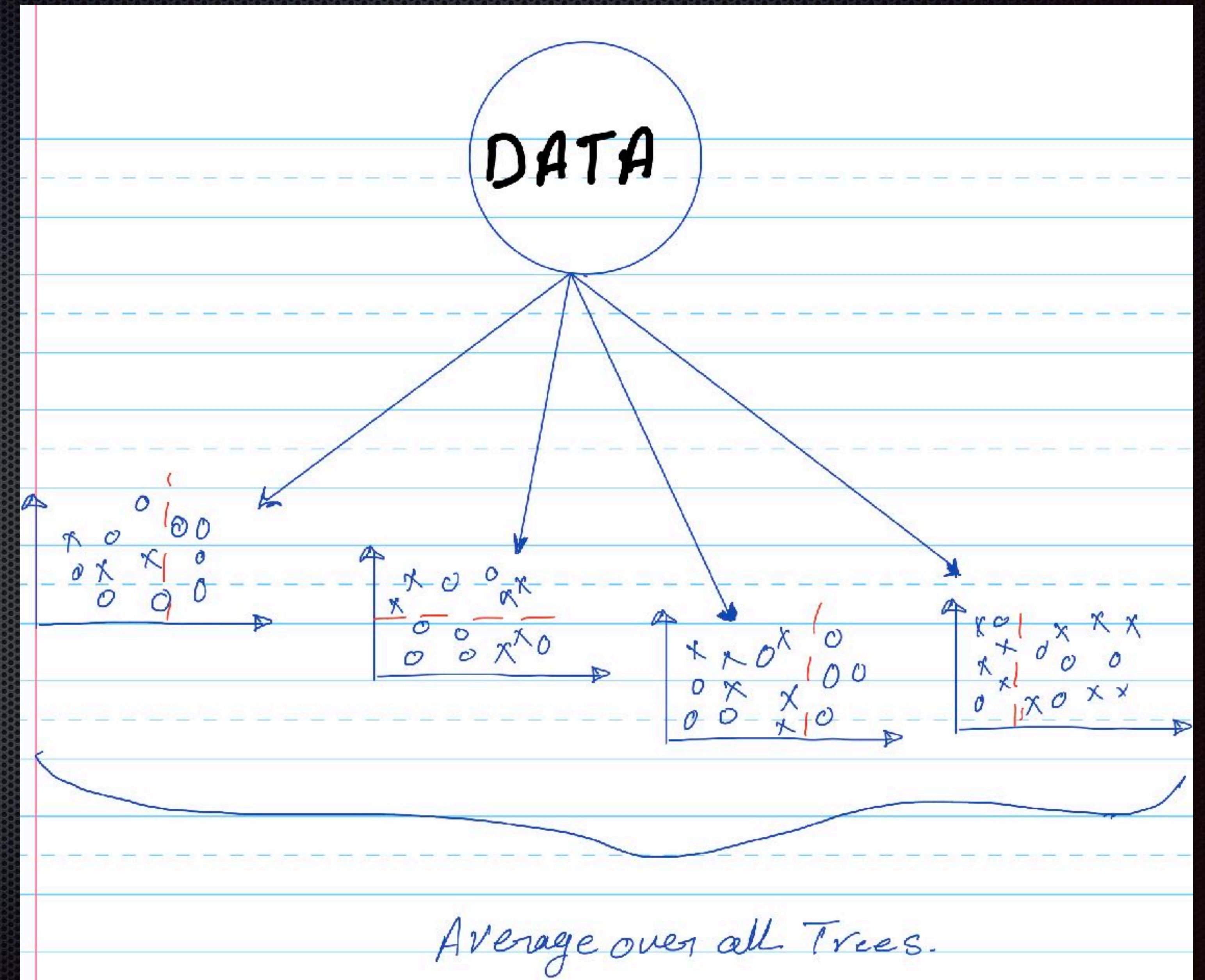


2 weak classifiers \Rightarrow strong classifier



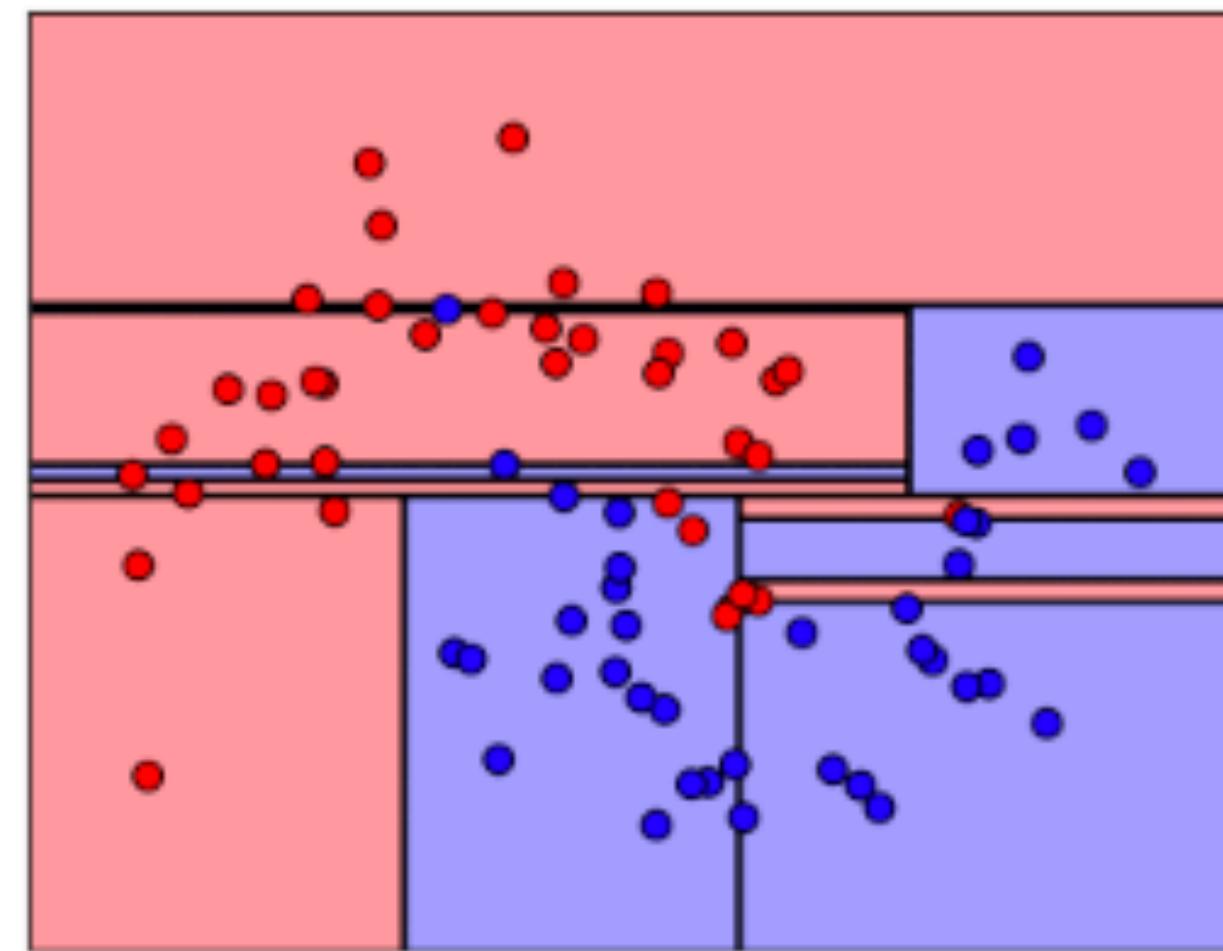
Random forest

Bagging: Bootstrap+aggregation

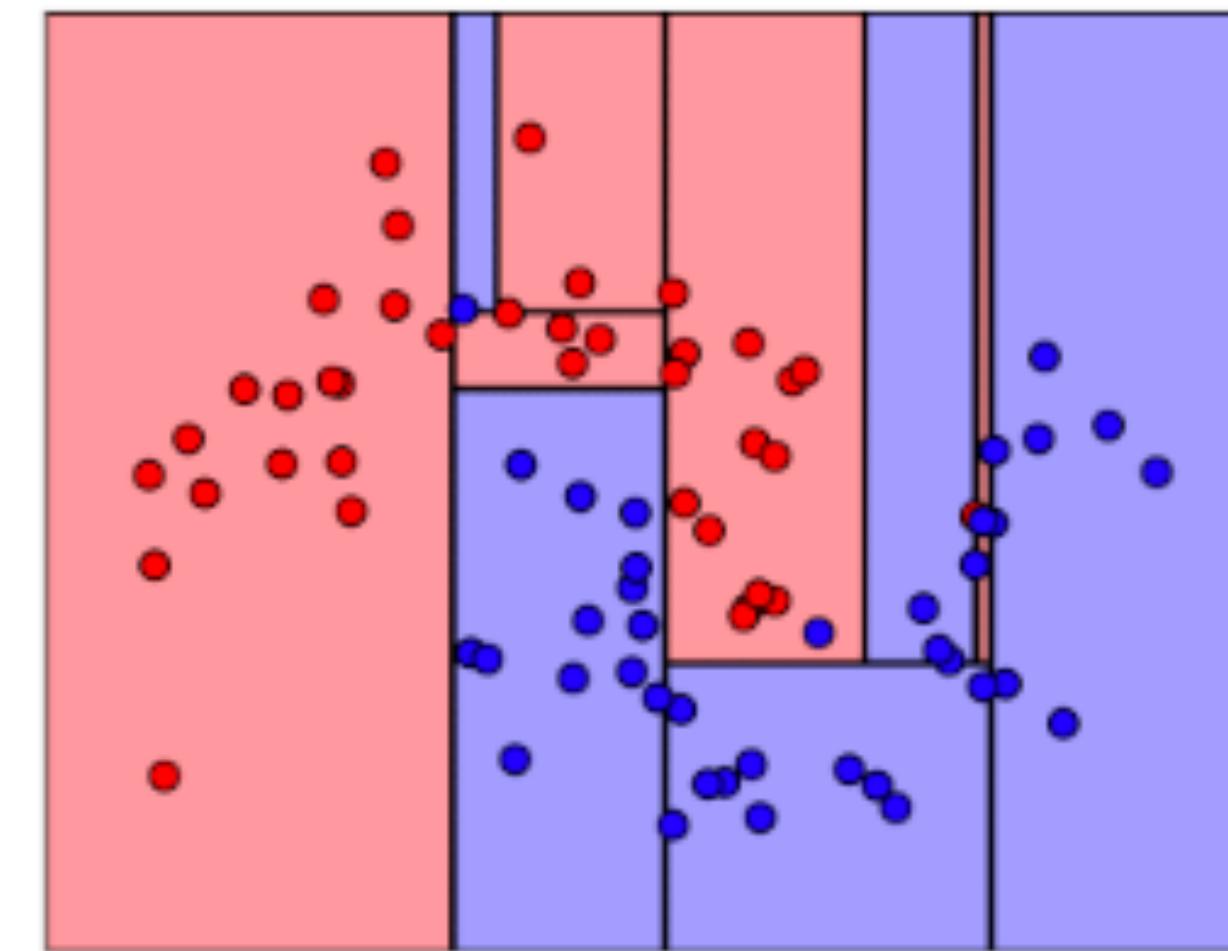


Random forest

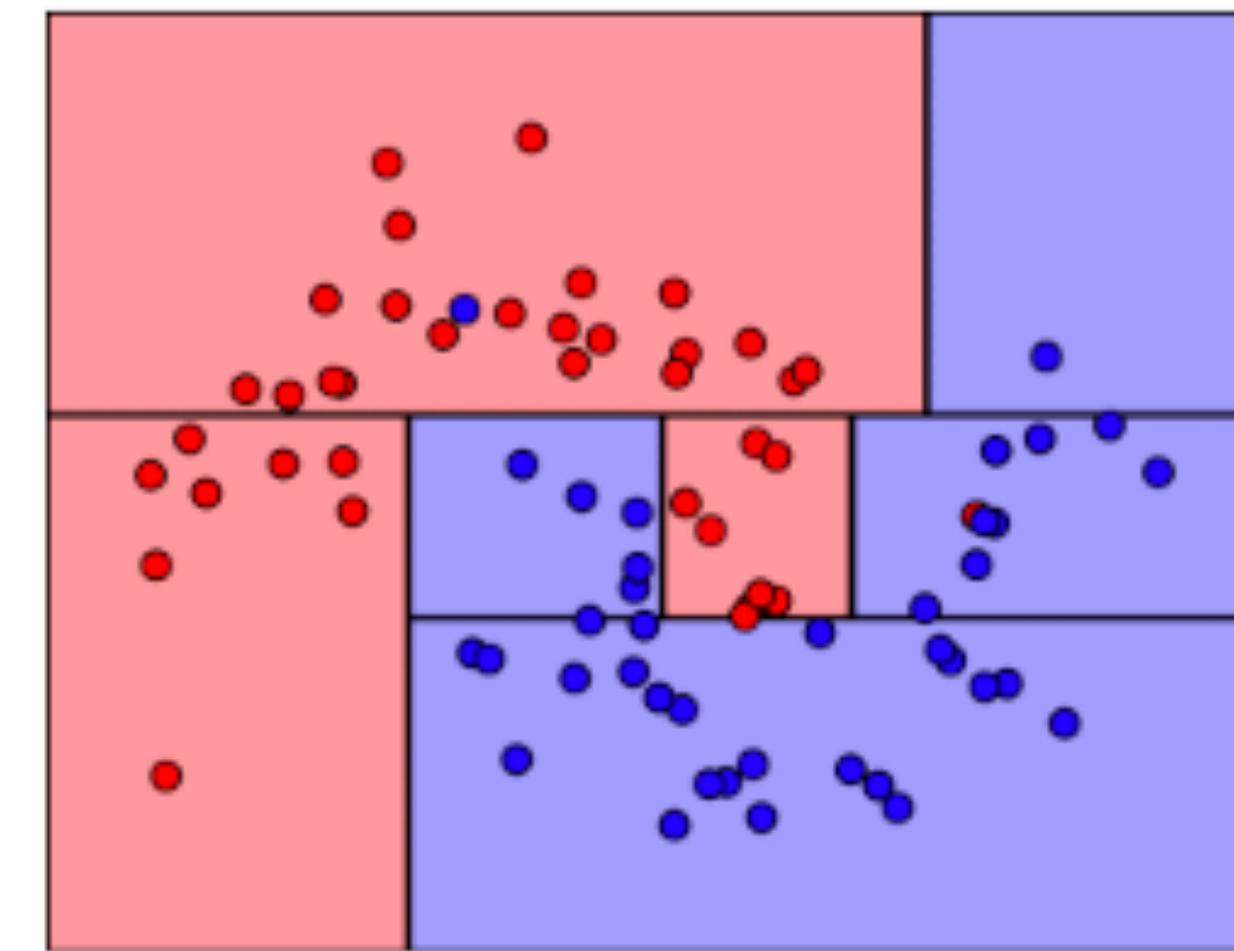
tree 0



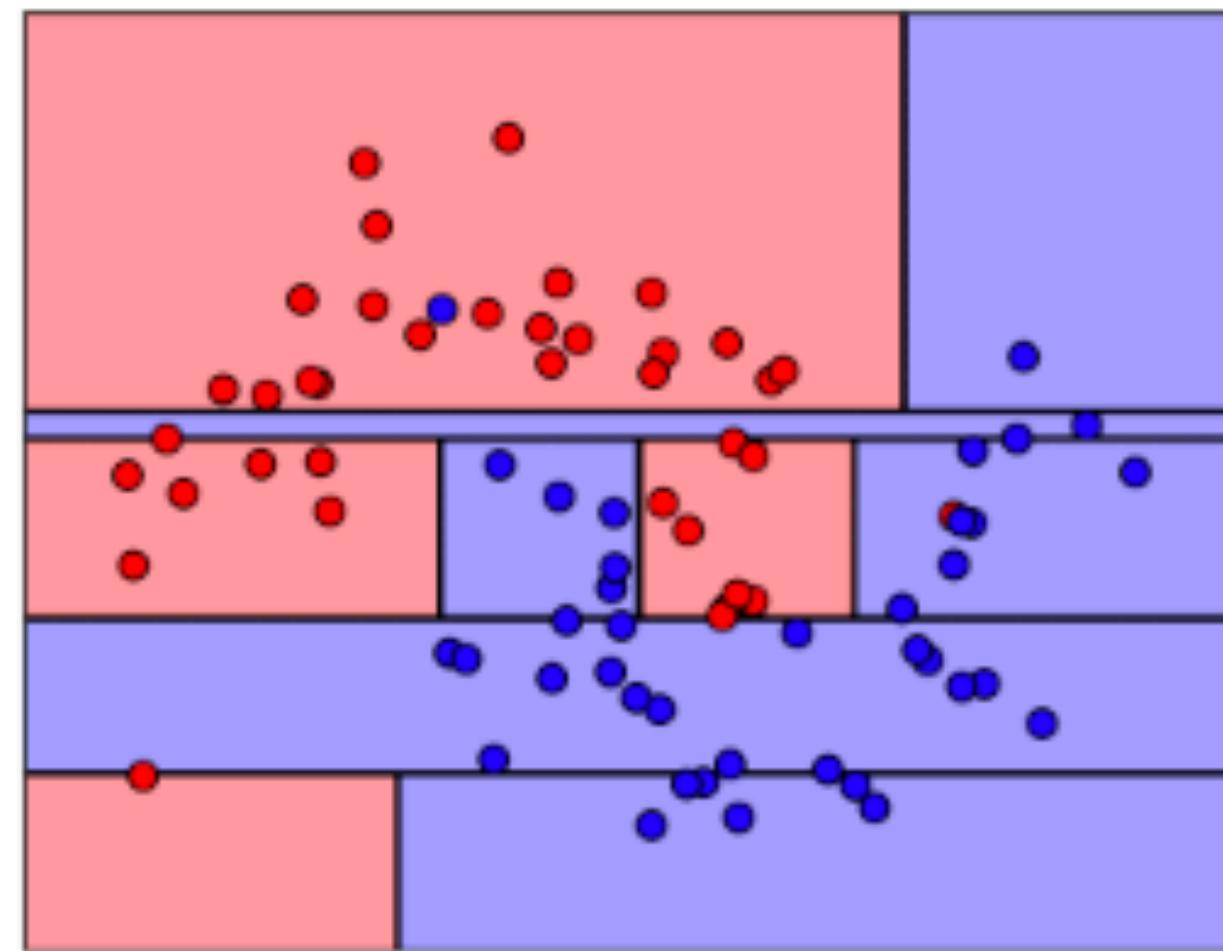
tree 1



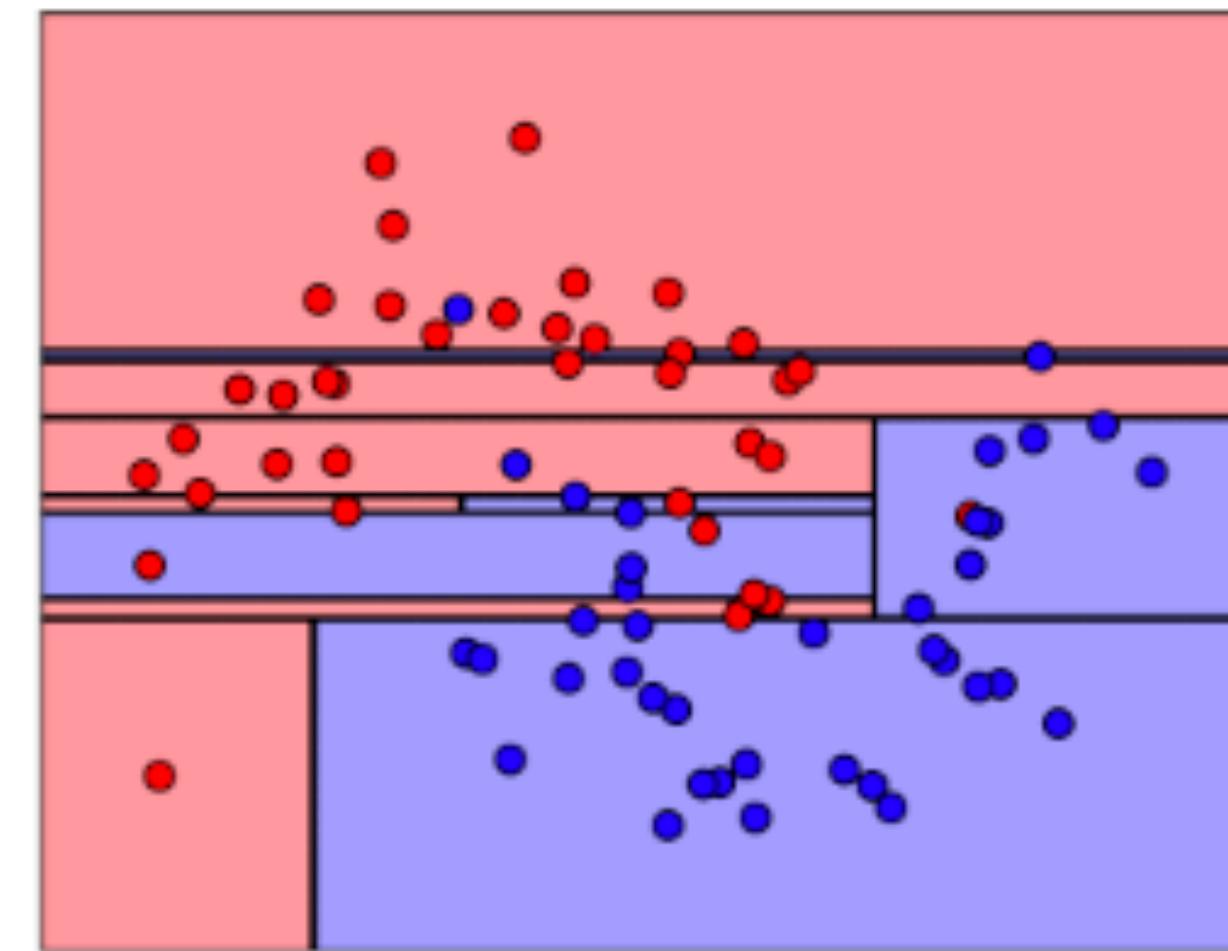
tree 2



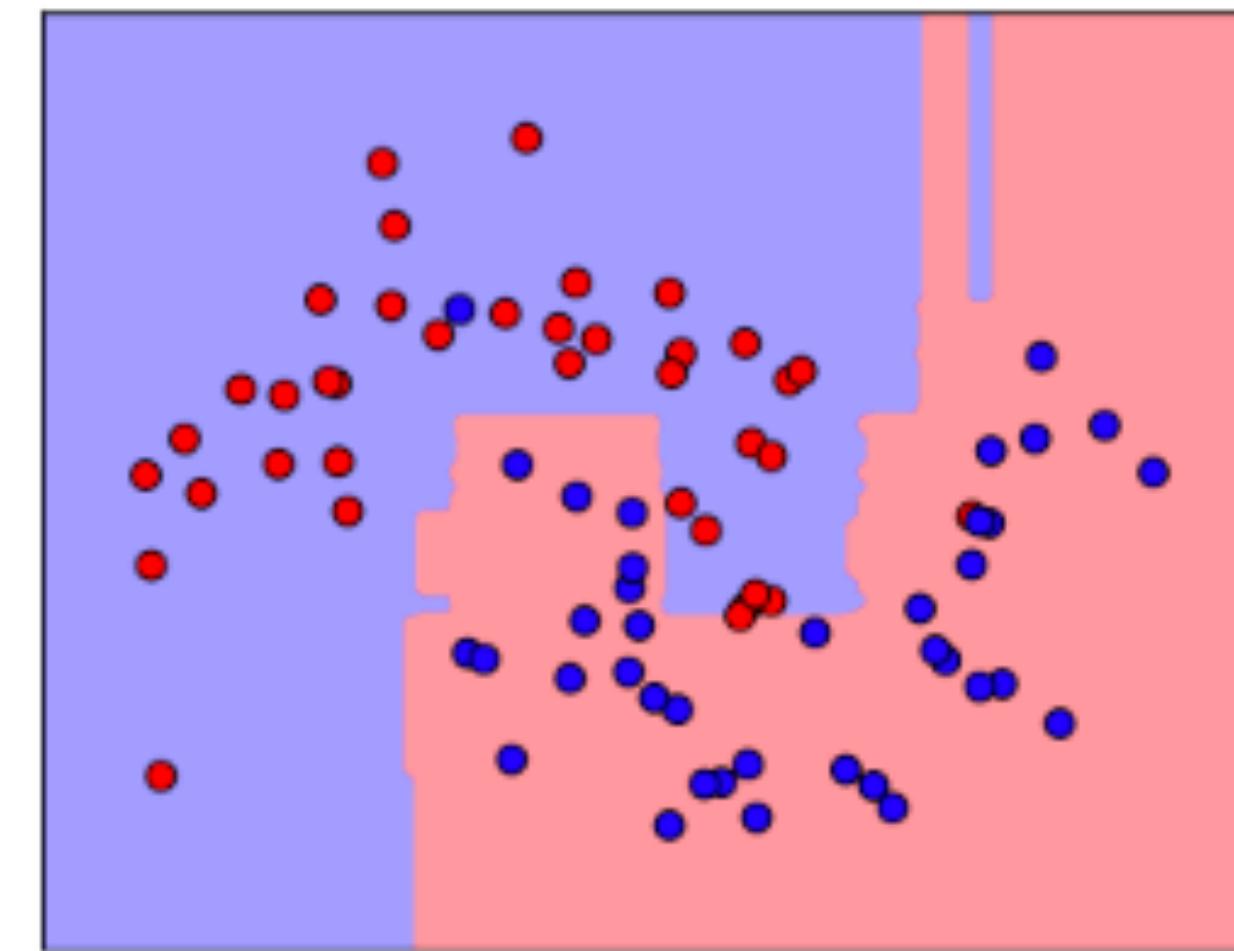
tree 3



tree 4

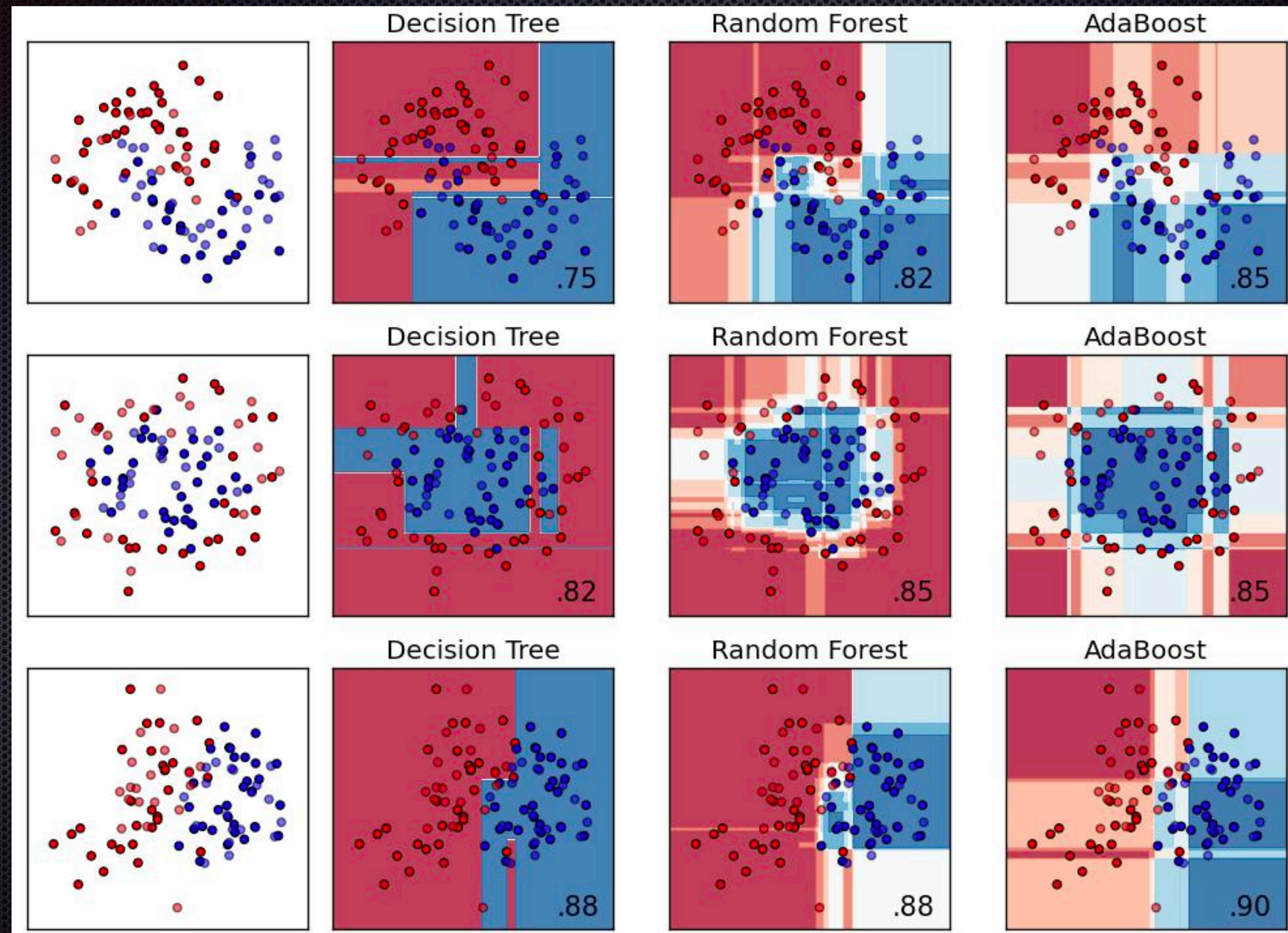


random forest



Ensemble learning

http://jimeladu9.leofile.ru.net/jele_463403_gradient-boosted-decision-trees-tutorial.php

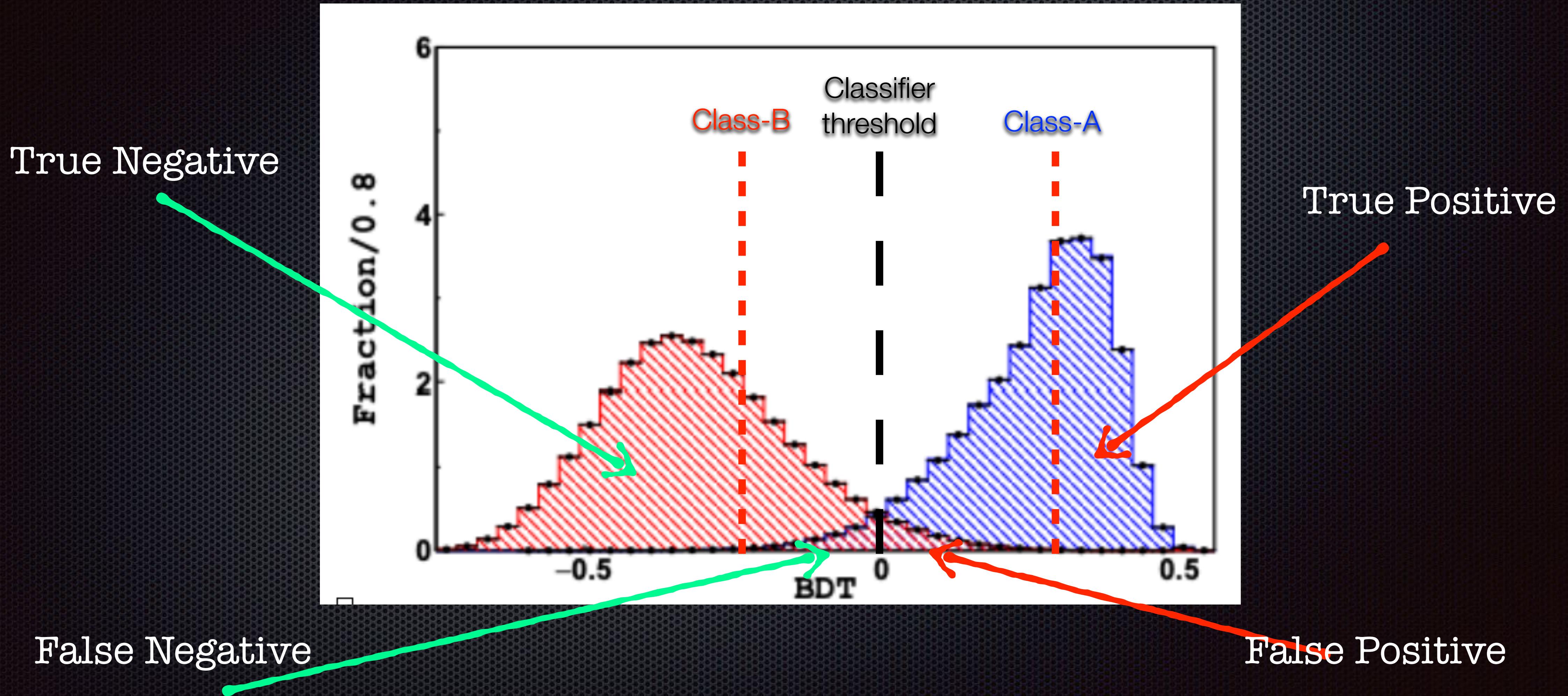


Ensemble learning

	Boosted decision Trees	Random forest
High classification accuracy	Yes	No
Statistical stability	No	Yes
Easy overfitting	Yes	No
Time consuming	Yes	No

Model evaluation

Model output



Model evaluation



Model evaluation

A Hi GPT

 Hello! How can I assist you today?

A I am a doctor can AI take my job ? answer precisely

 No, AI is unlikely to completely take the job of a doctor. While AI can assist doctors in various tasks such as diagnosis, treatment recommendations, and administrative tasks, it cannot replace the unique combination of clinical expertise, judgment, empathy, and interpersonal skills that doctors possess. The practice of medicine involves complex decision-making and requires a deep understanding of individual patients, which AI alone cannot replicate. The future is expected to involve a collaboration between doctors and AI, with AI augmenting and enhancing the work of doctors rather than replacing them.



ChatGPT confirms this will not be the case in the future

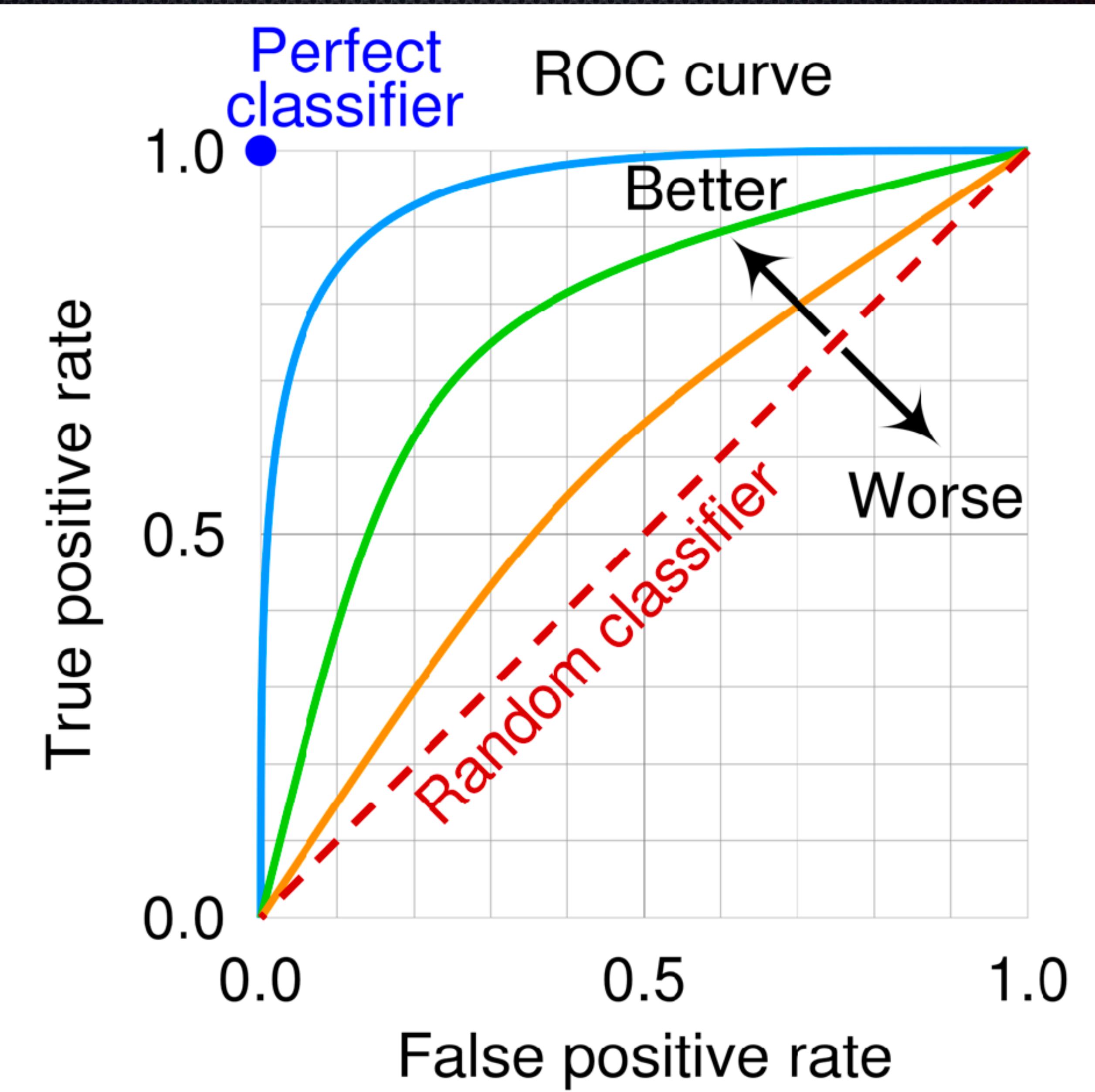
Recursive Operating classifier curve

https://en.wikipedia.org/wiki/Receiver_operating_characteristic#/media/File:Roc_curve.svg

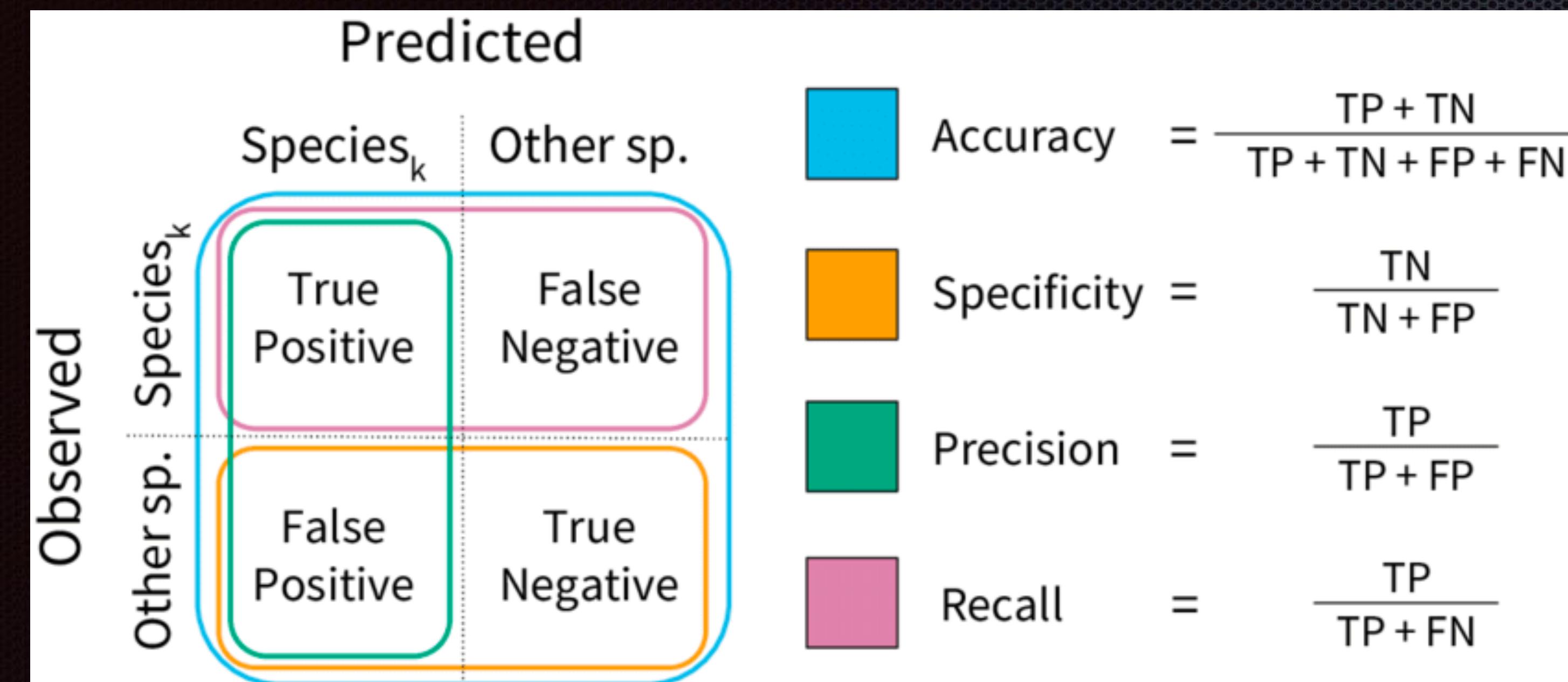
The area under the ROC curve (AUC) quantifies the model prediction performance



The larger AUC the larger model classification accuracy



Evaluation classification metrics



Accuracy:

Quantify the model ability Of making true prediction from both classes

Precision:

Quantify how good the model is at predicting a specific category

Recall:

Quantify how good the model is at detecting a specific category

Question

Assuming you are a data scientist who is responsible for airport flights analytics.

The airport authority told you that they suspect **a terrorist** will be on a flight

with 1000 other passengers. You asked to train a ML model to predict the terrorist.

What classification metric you are going to consider ?

To be continued...