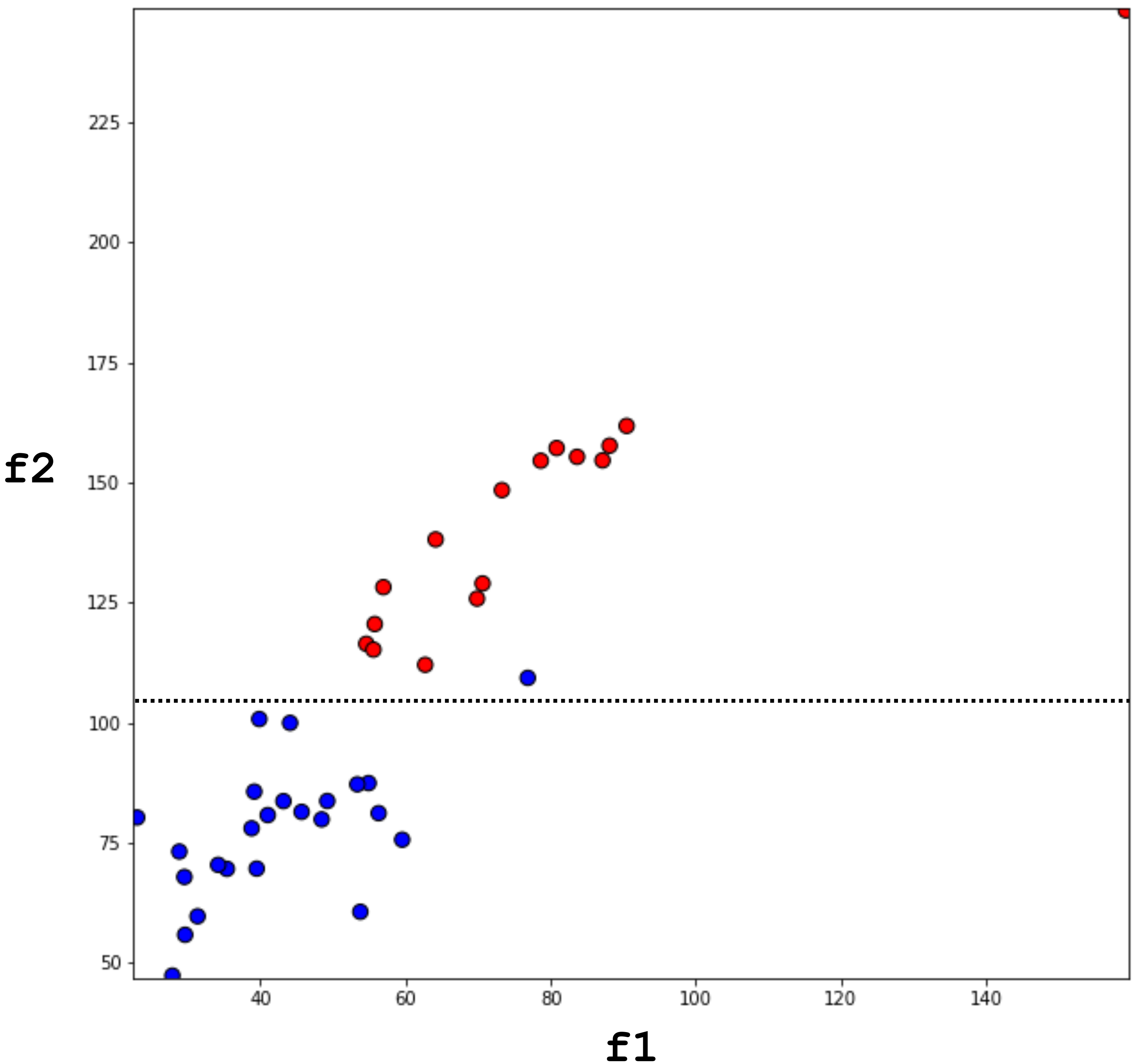
Classification - 2

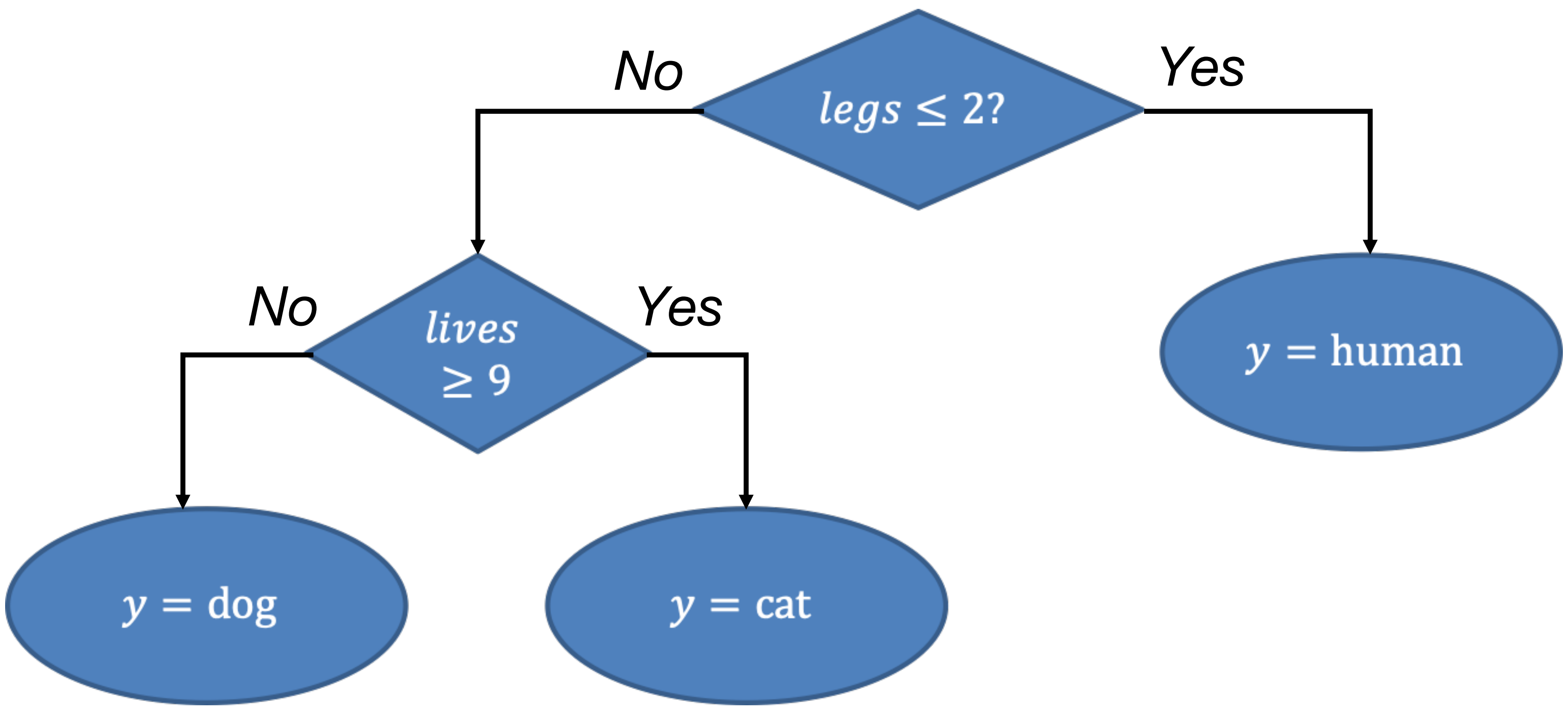
Sai

Decision Trees, Bagging, and Boosting




```
if f2 < 105:
    print 'Blue'
else:
    print 'Red'
```

Decision Trees



Decision Trees

- More explainable
- More interpretable
- Can be quite powerful
- One of the most commonly used algorithms
- Surprisingly useful for many sensing problems

Why do we care?

High-Stakes Decisions

- Healthcare: What treatment to recommend to the patient?
- Criminal Justice: Should the defendant be released on bail?

High-Stakes Decisions: Impact on human well-being.

Slide Credit: Hima Lakkaraju, Harvard

What is Interpretability?

- **Defn**: Ability to explain or to present in understandable terms to a human
- No clear answers in psychology to:
  - What constitutes an explanation?
  - What makes some explanations better than the others?
  - When are explanations sought?

Slide Credit: Hima Lakkaraju, Harvard

When and Why Interpretability?

- Not all ML systems require interpretability
    - E.g., ad servers, postal code sorting
    - No human intervention
- No explanation needed because:
    - No consequences for unacceptable results
    - Problem is well studied and validated well in real-world applications → trust system's decision

When do we need explanation then?

Motivation for Interpretability

- ML systems are being deployed in complex high-stakes settings
- Accuracy alone is no longer enough
- Auxiliary criteria are important:
    - Safety
    - Nondiscrimination
    - Right to explanation (Now an EU law!)

Slide Credit: Hima Lakkaraju, Harvard

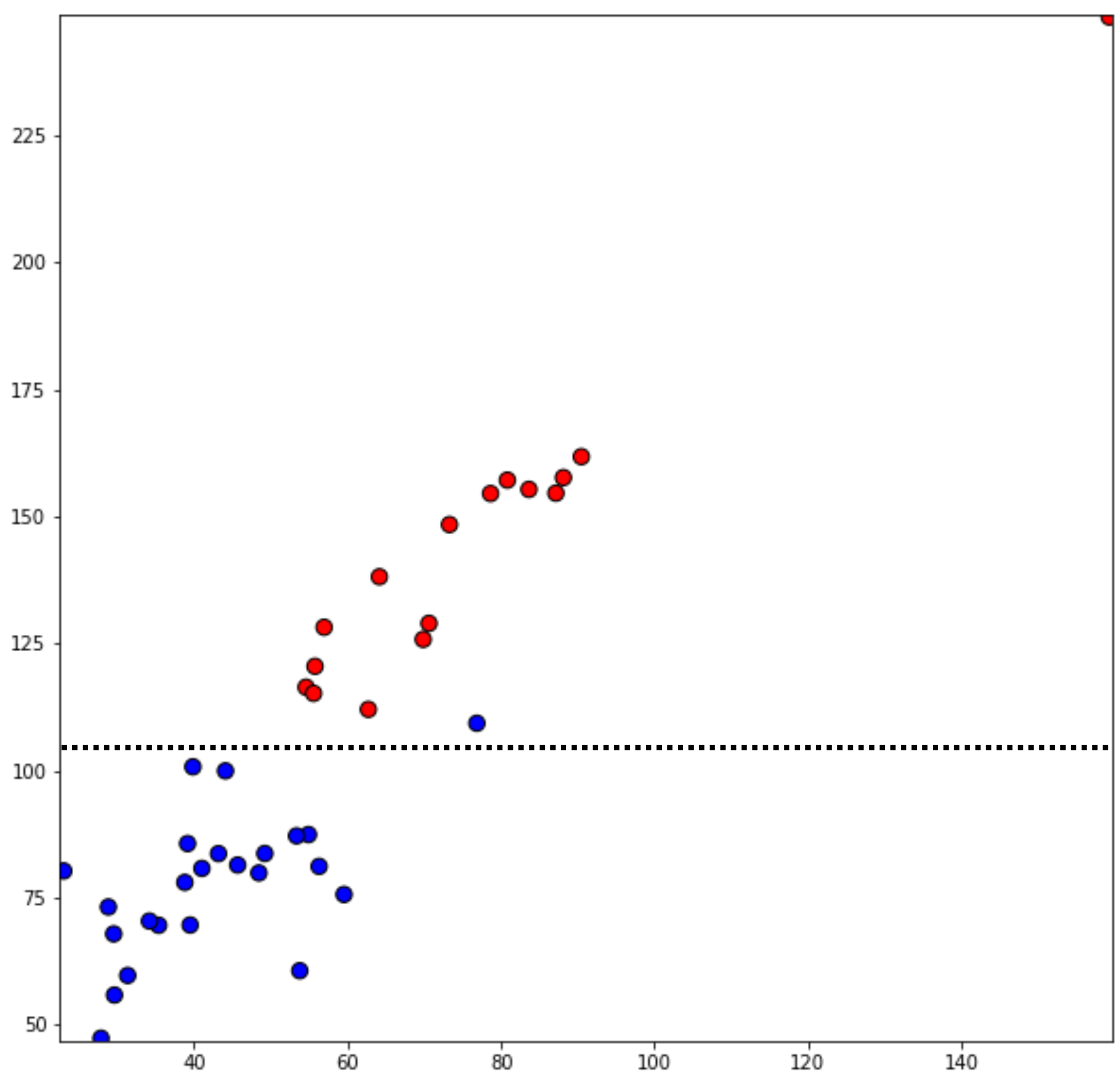
Decision Trees

- More explainable

Why do we care?

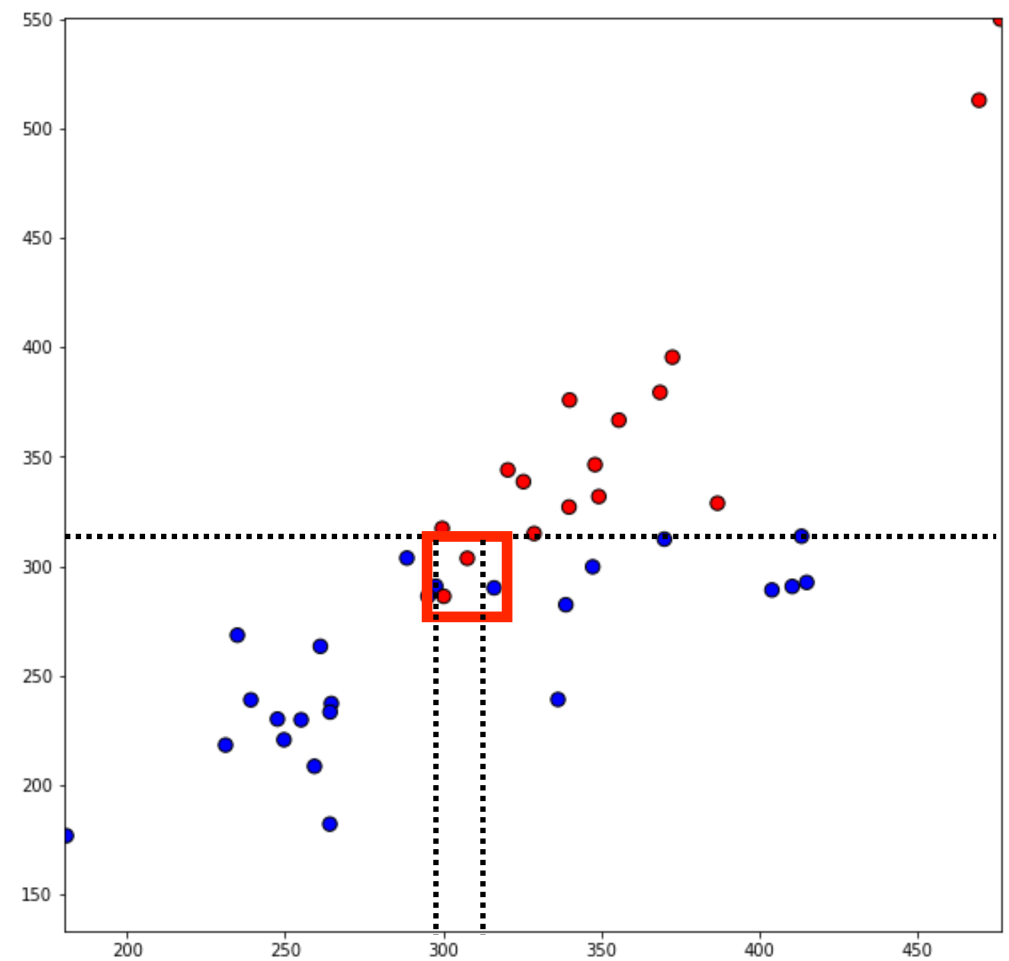
- More interpretable
- Can be quite powerful
- One of the most commonly used algorithms
- Surprisingly useful for many sensing problems

Decision Trees with Sensor Data



- Threshold heuristics can be hard to develop
- DTs figures out the thresholds for you
- Especially in case of other environmental variables

Decision Trees



- The tree can look at the same variable multiple times
- Come up with a long sequence of decision points to develop the complete tree

Building a Decision Tree

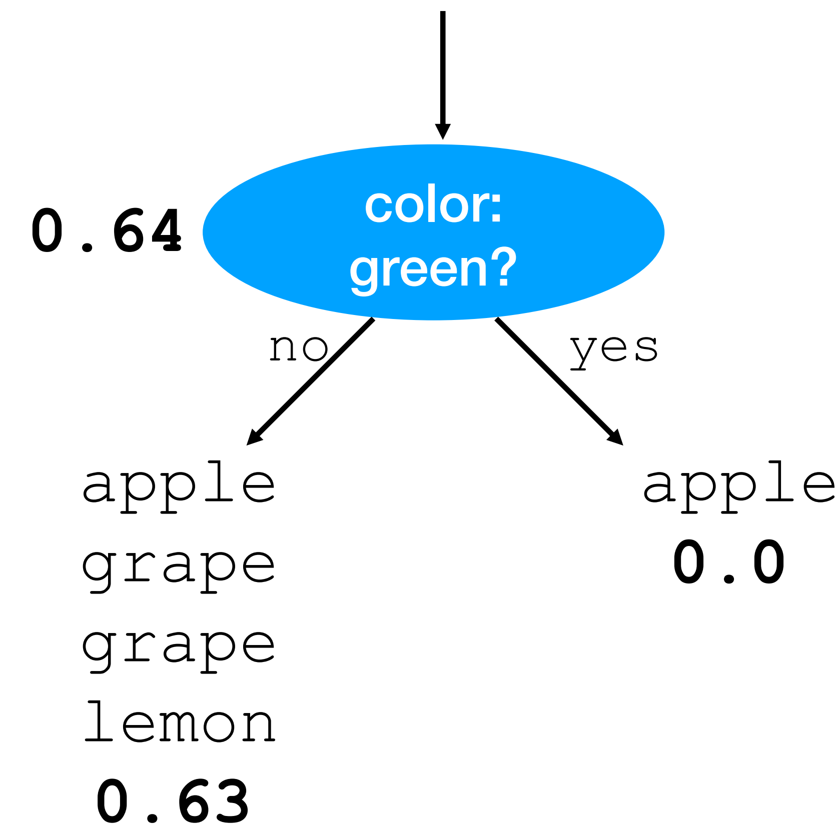
Figuring out when to ask what question!

Gini Impurity: What is the probability of me randomly picking up class A and being correct?

Color	Measurement	Fruit
Green	15	Apple
Yellow	15	Apple
Red	5	Grape
Red	5	Grape
Yellow	10	Lemon

Courtesy: Josh Gordon, Google

Color	Measurement	Fruit
Green	15	Apple
Yellow	15	Apple
Red	5	Grape
Red	5	Grape
Yellow	10	Lemon



$$
\begin{aligned}
G.I. = {} & p(\text{apple}) \cdot (1 - p(\text{apple})) \\
& + p(\text{grape}) \cdot (1 - p(\text{grape})) \\
& + p(\text{lemon}) \cdot (1 - p(\text{lemon}))
\end{aligned}
$$

$$G.I. = 0.64$$

Average Impurity

= 4/5 * 0.63 + 1/5 * 0.0

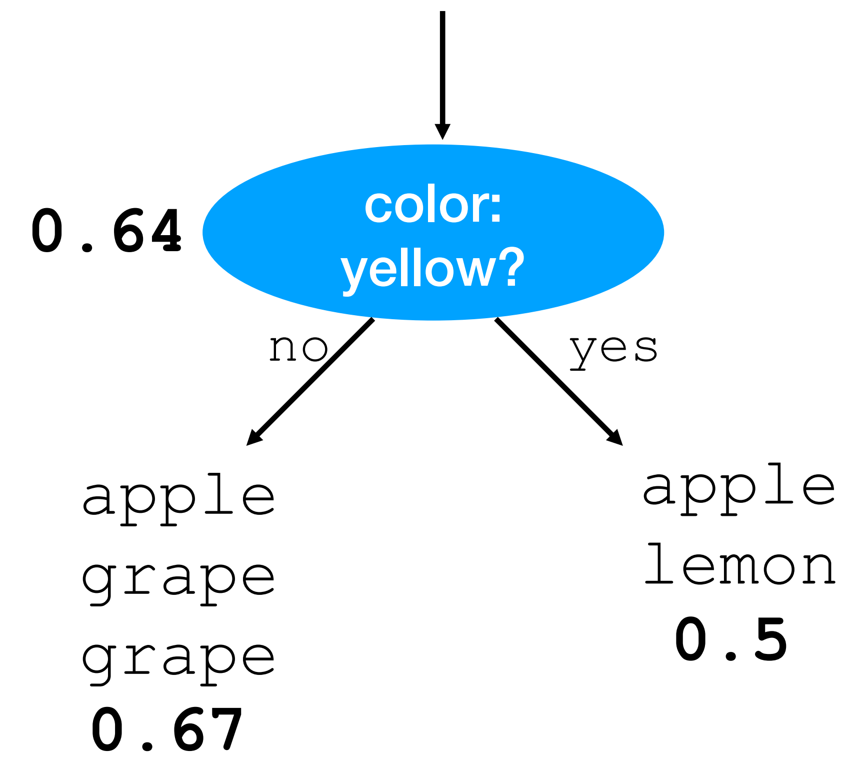
= ~0.5

Information Gain

= 0.64 - 0.5

= 0.14

Color	Measurement	Fruit
Green	15	Apple
Yellow	15	Apple
Red	5	Grape
Red	5	Grape
Yellow	10	Lemon



G.I. = p (apple) . (1 − p (apple))
+ p (grape) . (1 − p (grape))
+ p (lemon) . (1 − p (lemon))

G.I. = 0.64

Average Impurity

= 3/5 * 0.67 + 2/5 * 0.5

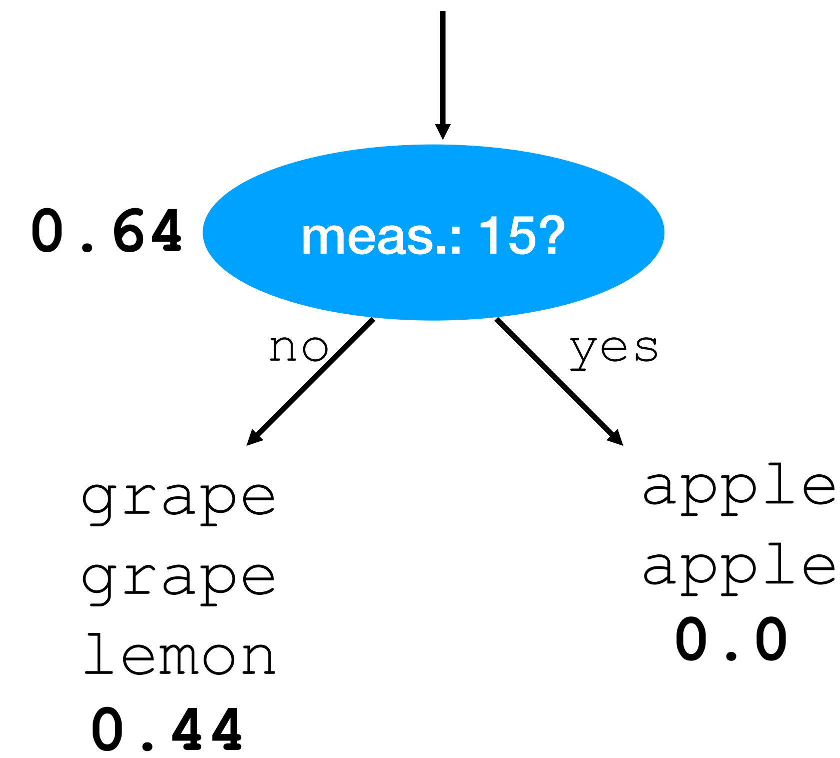
= ~0.6

Information Gain

= 0.64 - 0.6

= 0.04

Color	Measurement	Fruit
Green	15	Apple
Yellow	15	Apple
Red	5	Grape
Red	5	Grape
Yellow	10	Lemon



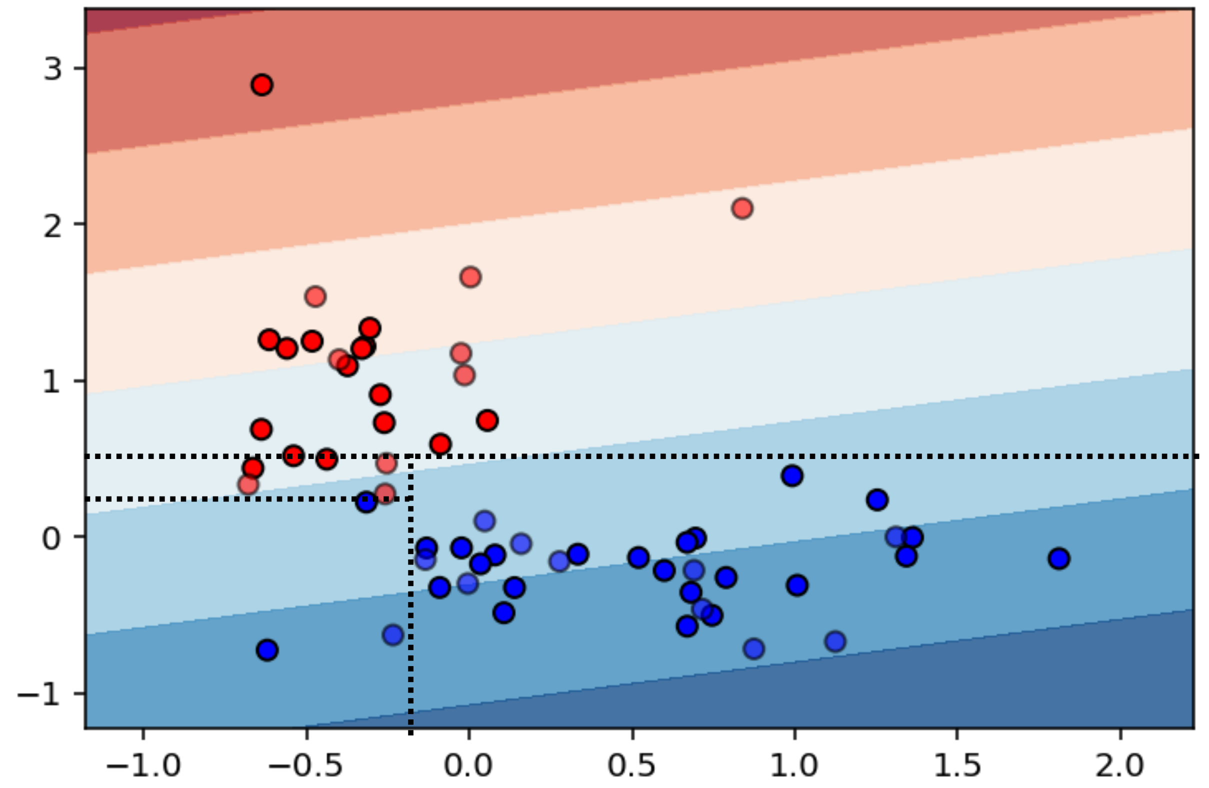
G.I. = p (apple) . (1 - p (apple))
+ p (grape) . (1 - p (grape))
+ p (lemon) . (1 - p (lemon))

G.I. = 0.64

Average Impurity
= 3/5 * 0.44 + 2/5 * 0.0
= 0.26

Information Gain
= 0.64 - 0.26
= 0.38

Decision Trees



- This is the output from an SVM with a linear kernel
- DTs cannot build such angled lines
- No “smooth” separations
- Too dependent on the axes
- **No need to scale though!**

Decision Trees ++

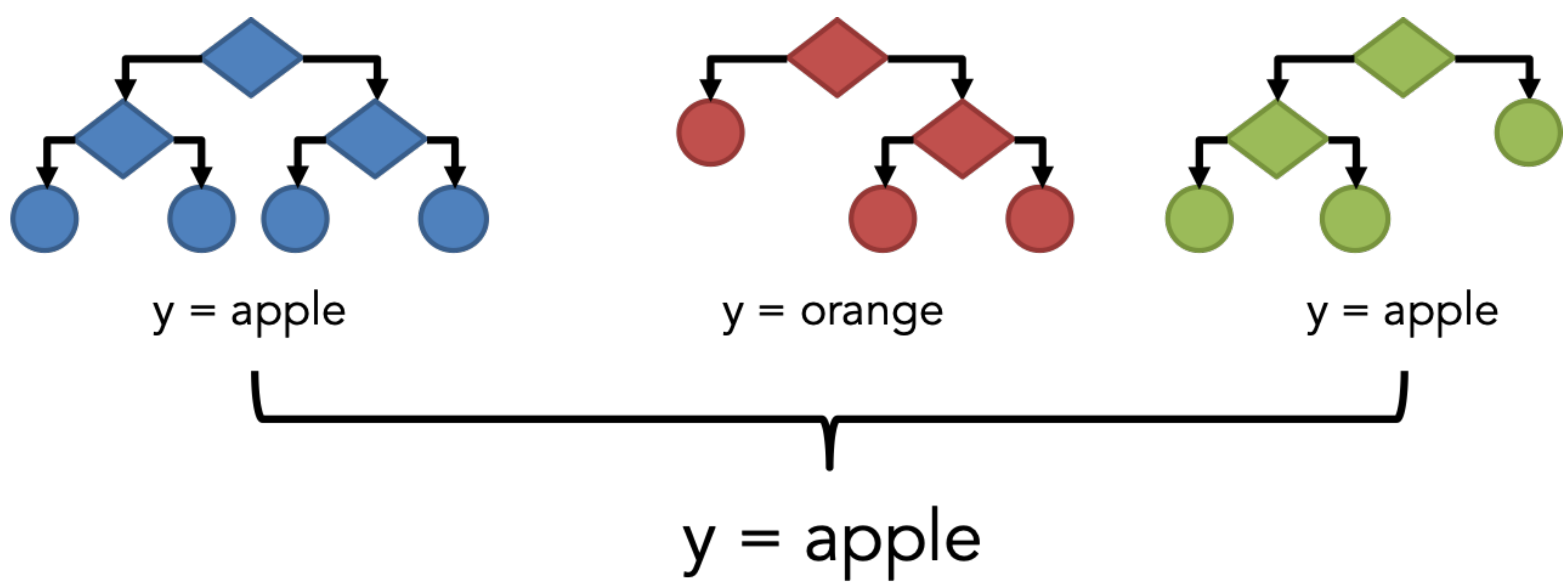
Ensemble of Trees

Build multiple trees and combine their output

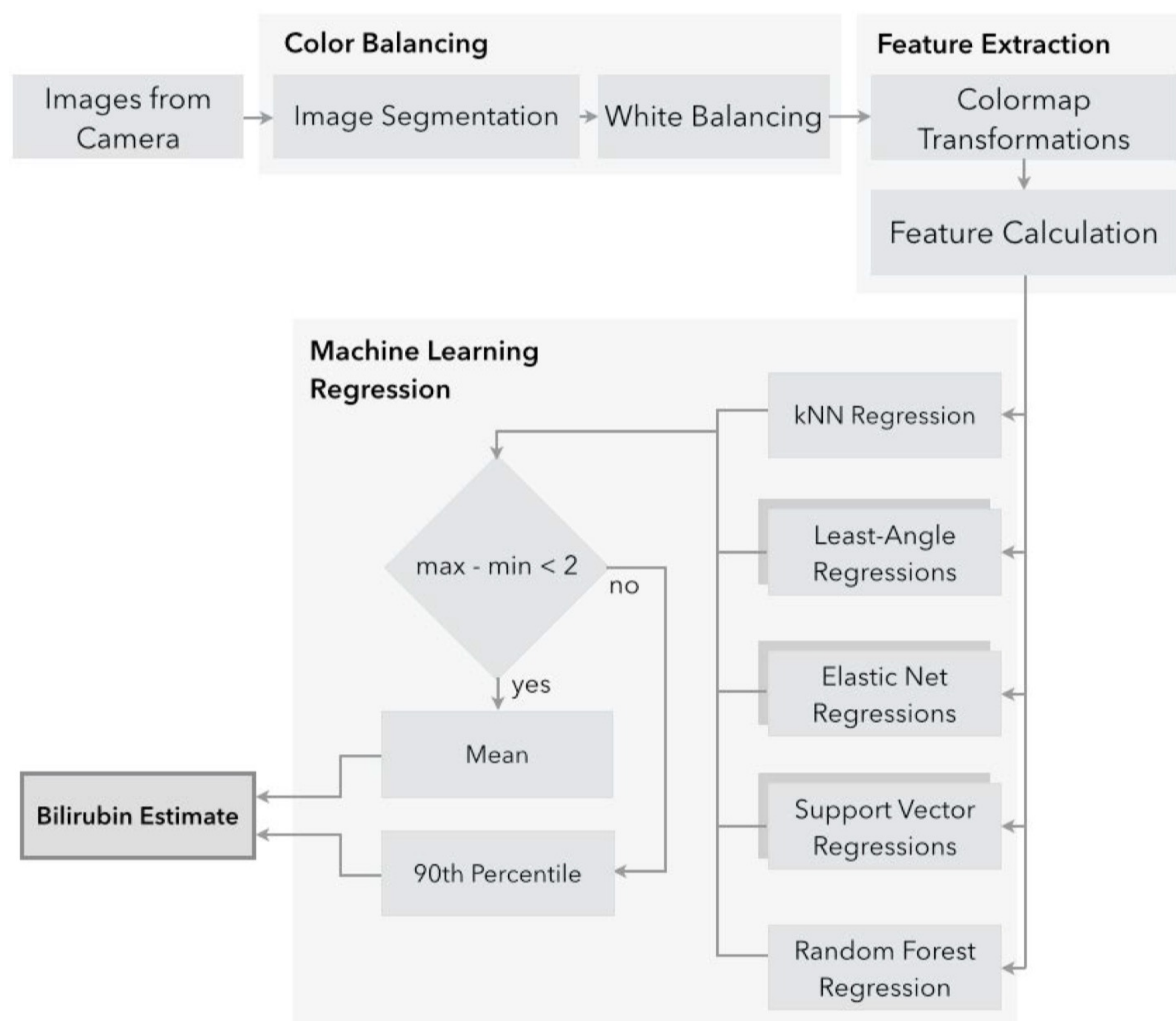
- Bagging
- Boosting

Bagging

- In parallel
- Random Forests



Bagging


Color Balancing
Feature Extraction
Images from Camera
Image Segmentation
White Balancing
Colormap Transformations
Feature Calculation
Machine Learning Regression
kNN Regression
Least-Angle Regressions
Elastic Net Regressions
Support Vector Regressions
Random Forest Regression
max - min < 2
no
yes
Mean
90th Percentile
Bilirubin Estimate

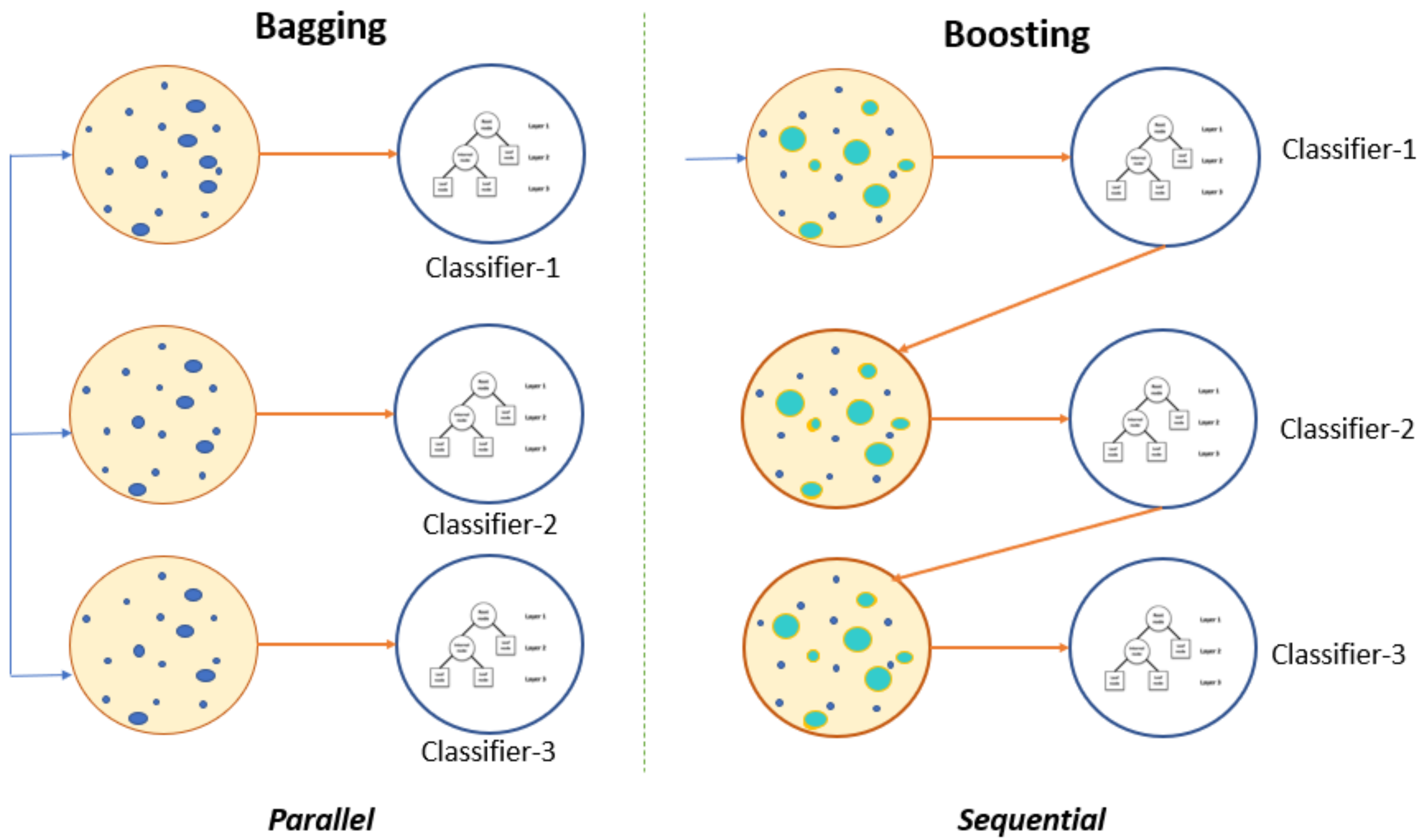


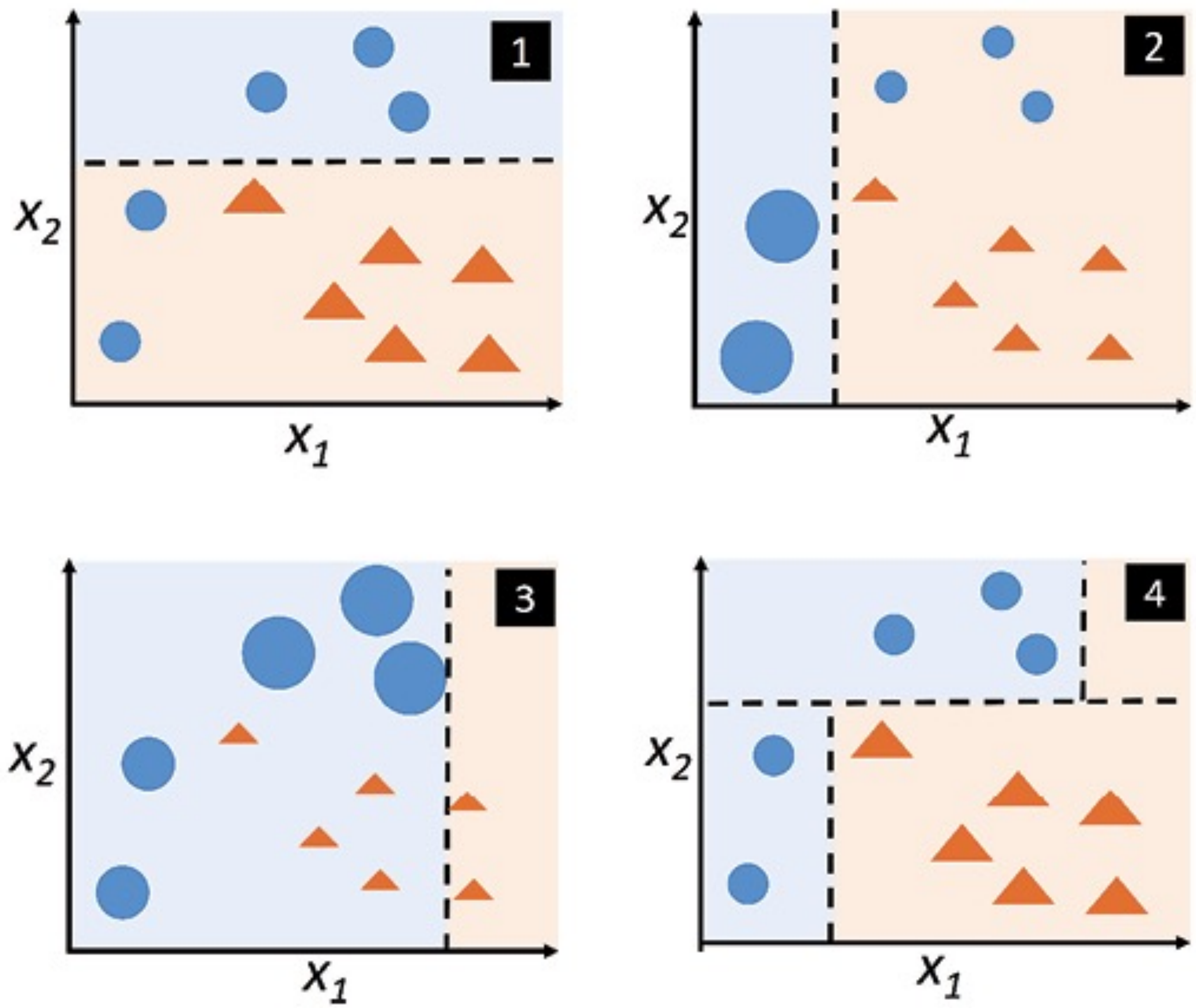
Image Credit: pluralsight.com

Boosting

- Combining *weak classifiers* to make a stronger one
- Sequential

Boosting

- AdaBoost (Adaptive Boosting)
    - Fit a tree
    - Calculate error (weighted)
    - Increase weight of wrongly classified points
    - Train another tree
    - ……….
- In the end, each tree will have a weight, and
- the final prediction is the weighted majority vote from each tree

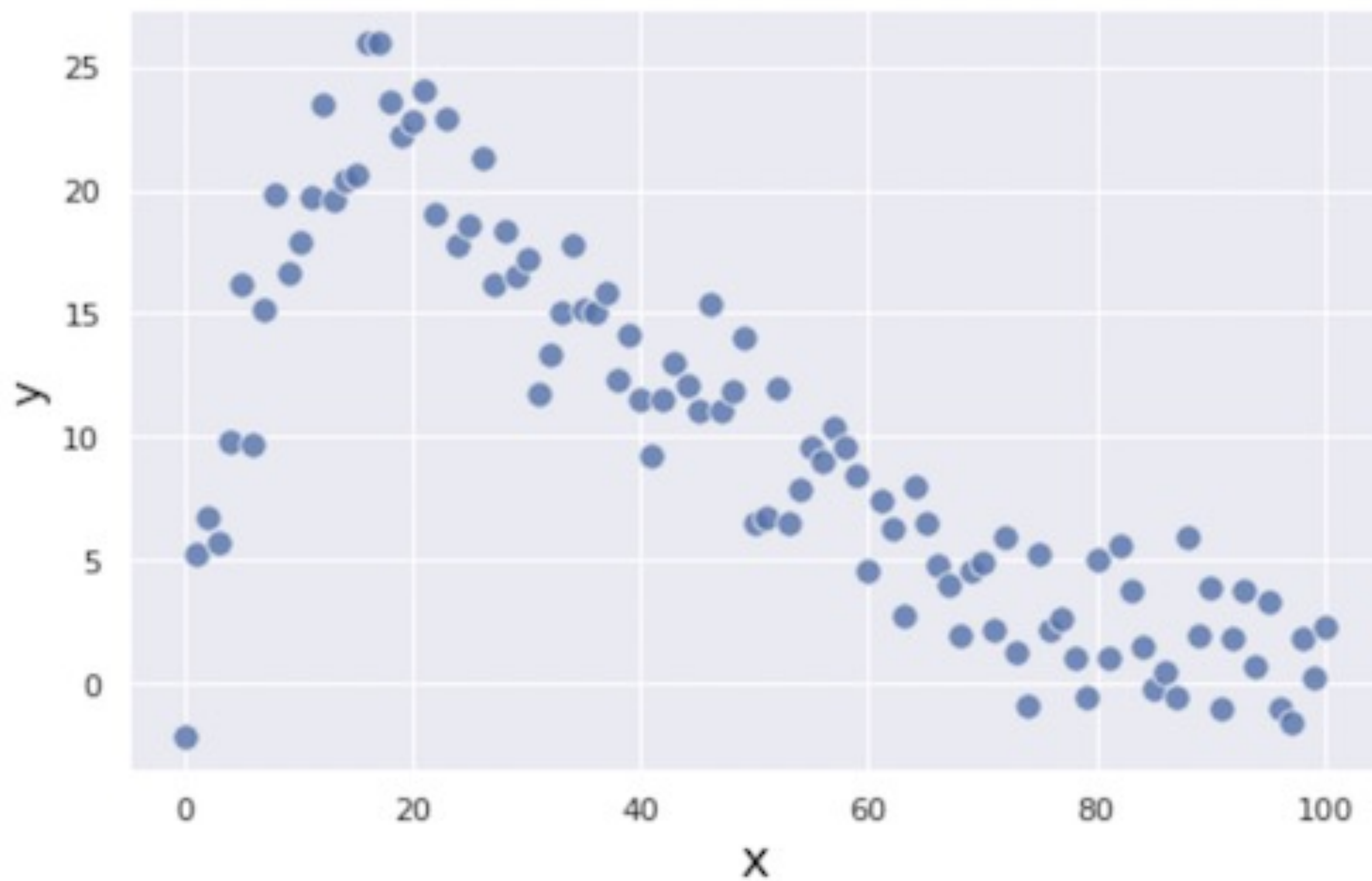


Source: vitalflux.com

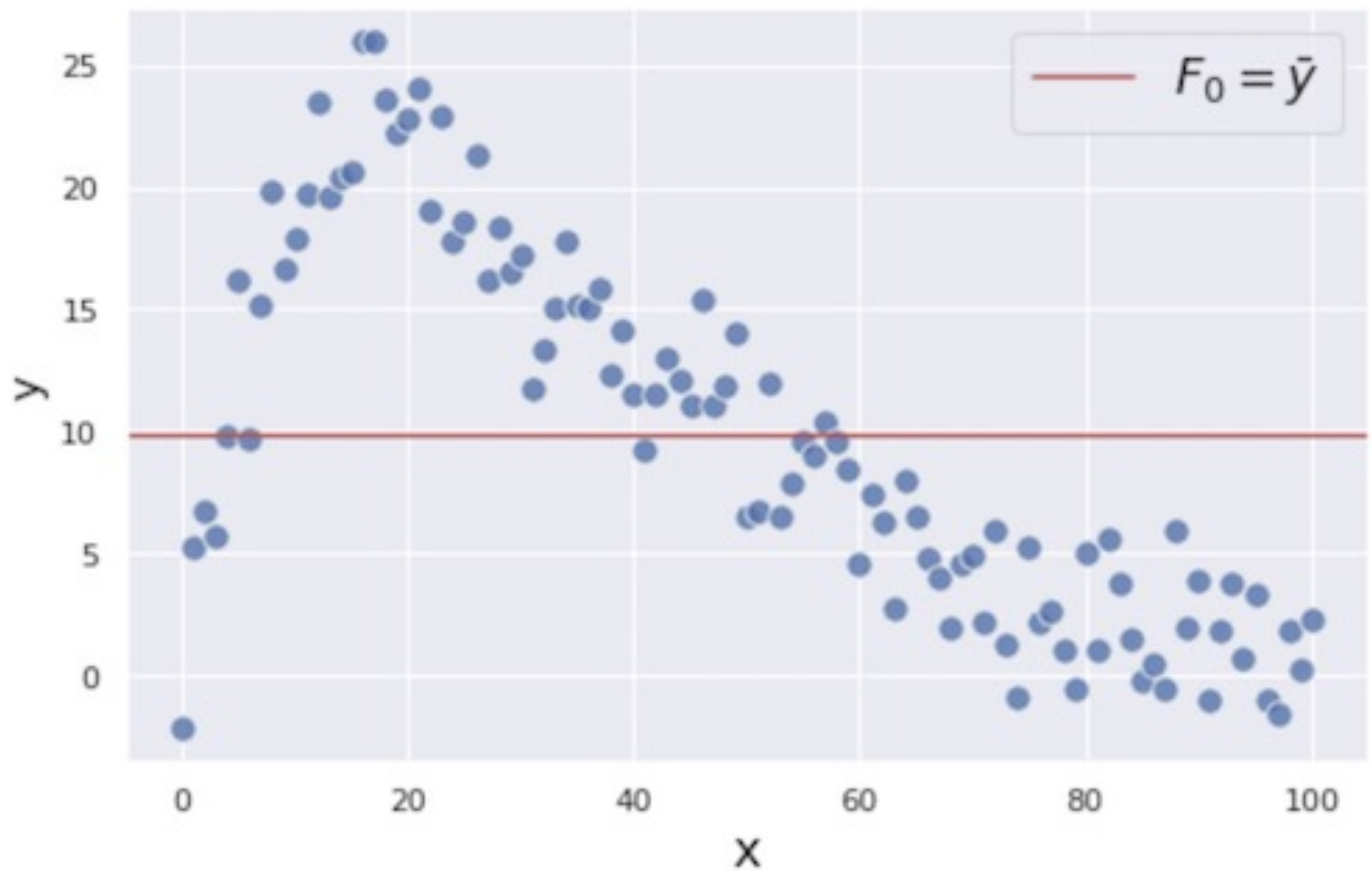
Boosting

- Gradient Boosting
    - Fit a tree
    - Calculate error (weighted)
    - Train on error
    - Train another tree
    - ……….

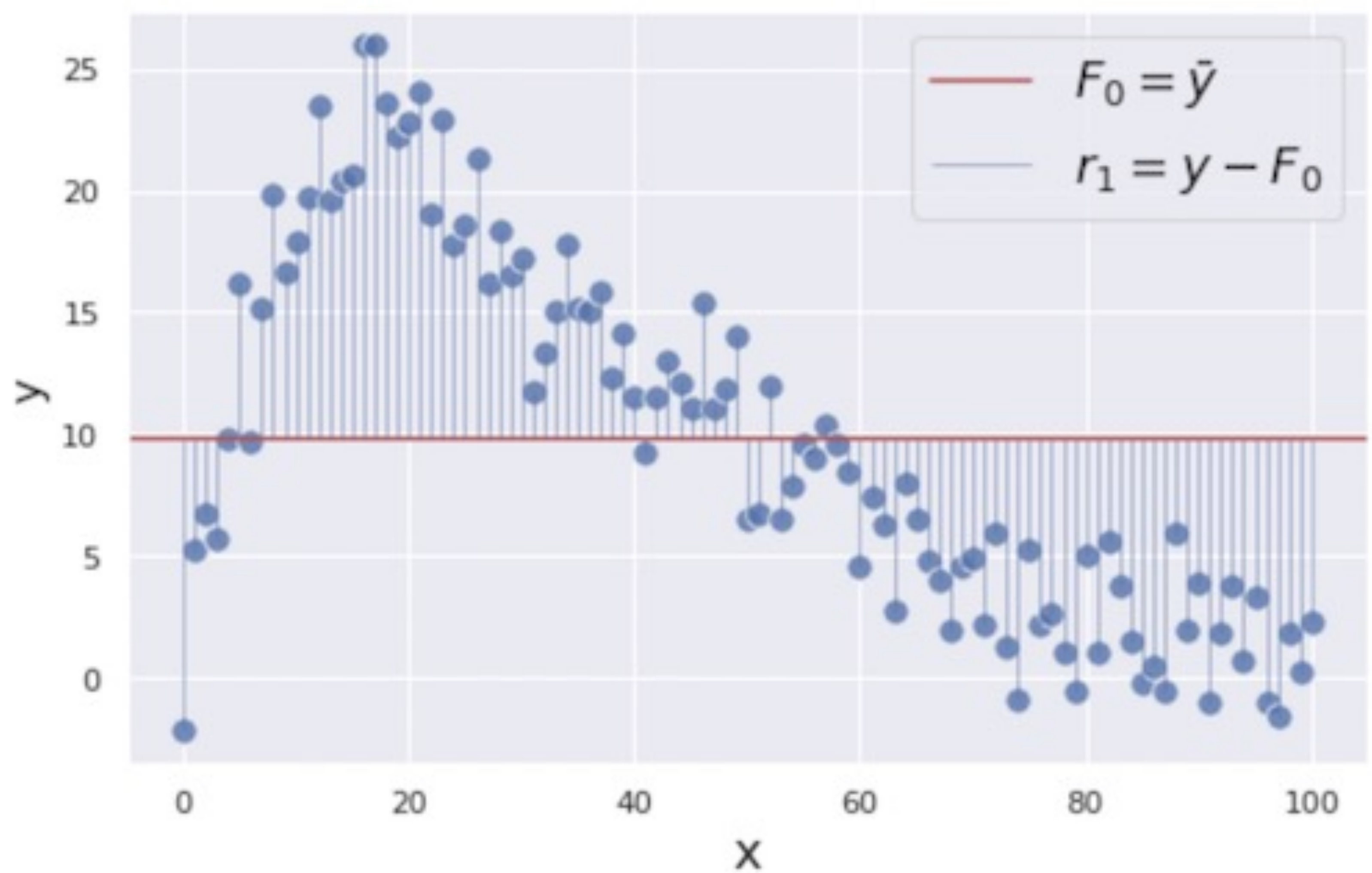
- In the end, add up the predictions of each tree



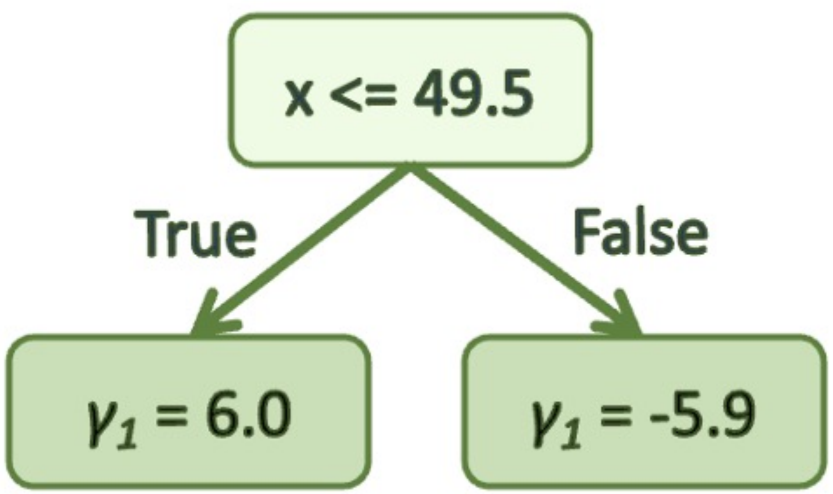
Credit: Tomonori Masui

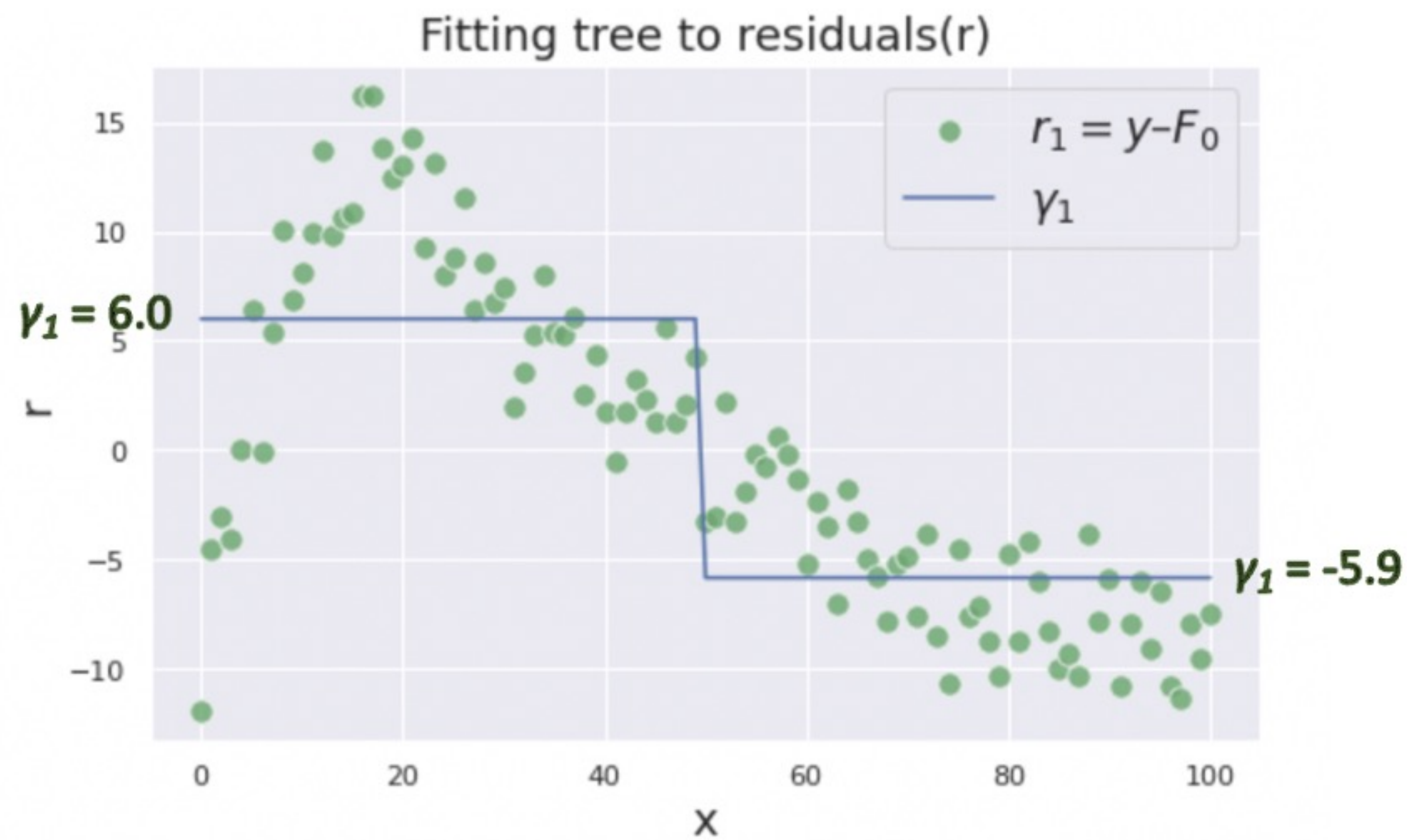


Credit: Tomonori Masui

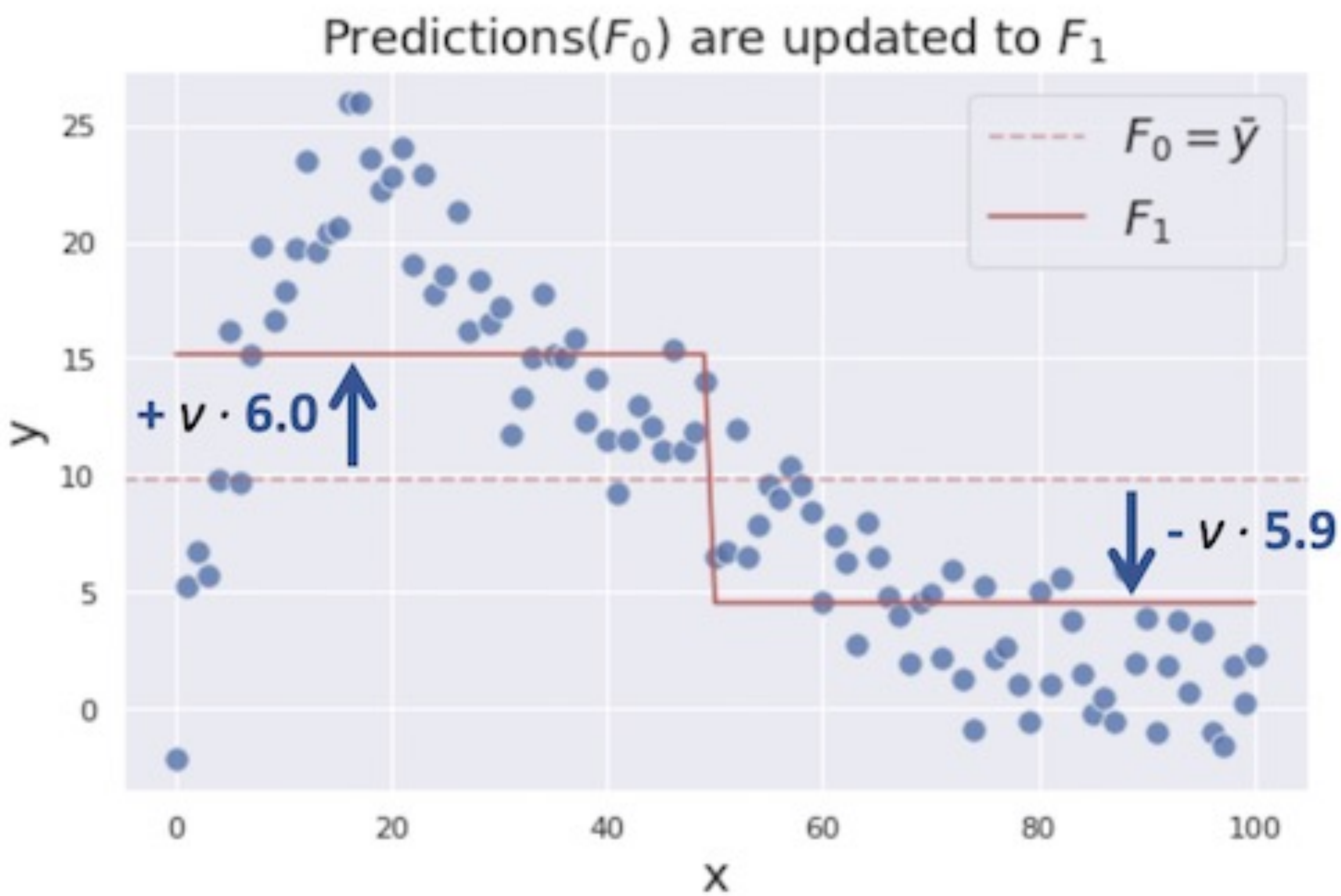


Credit: Tomonori Masui



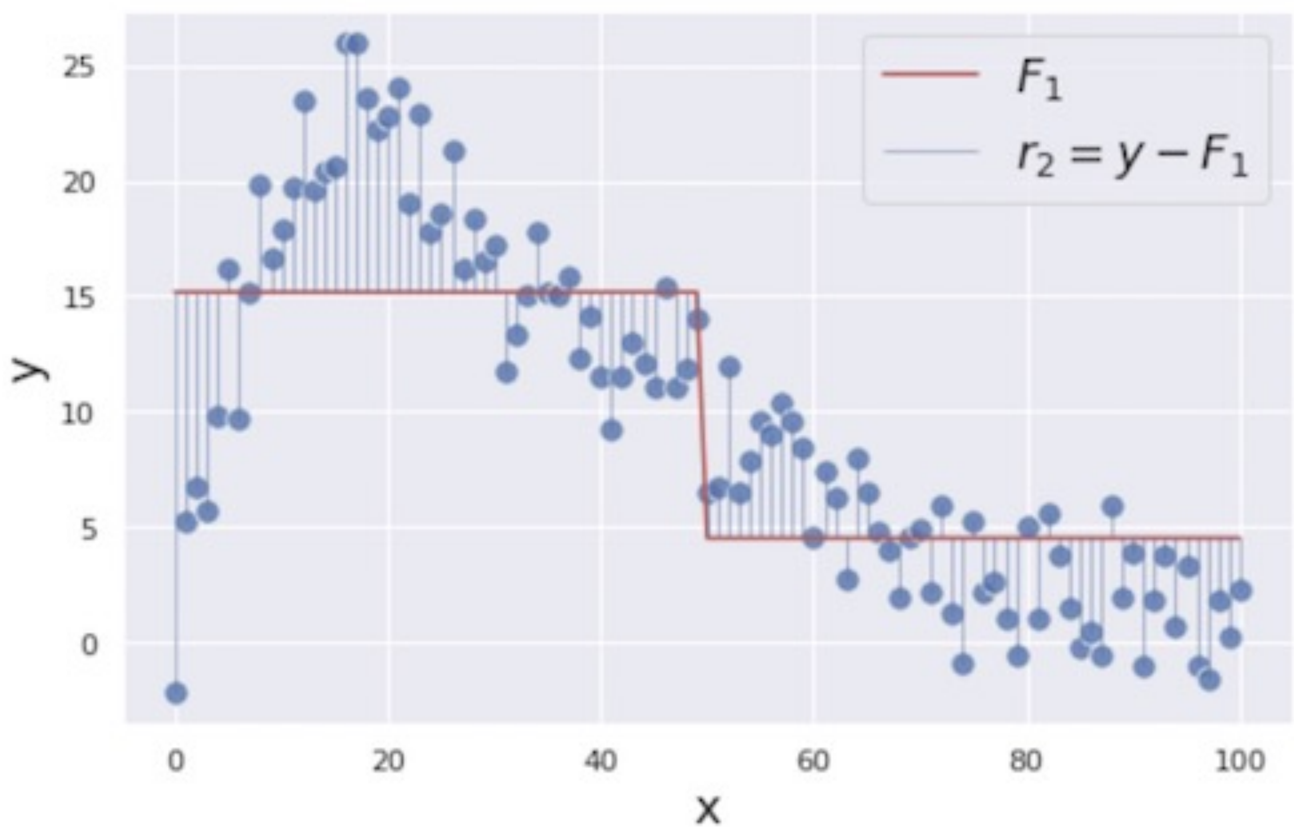


Credit: Tomonori Masui

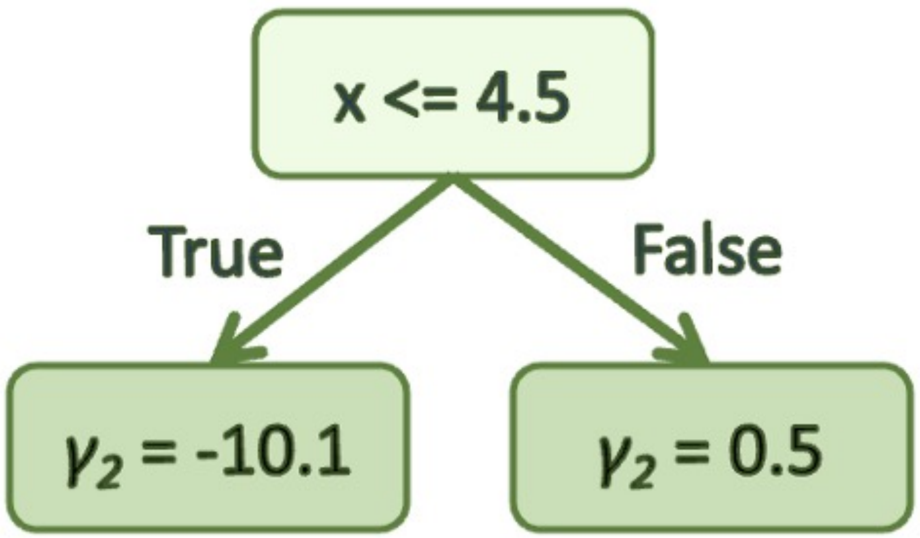


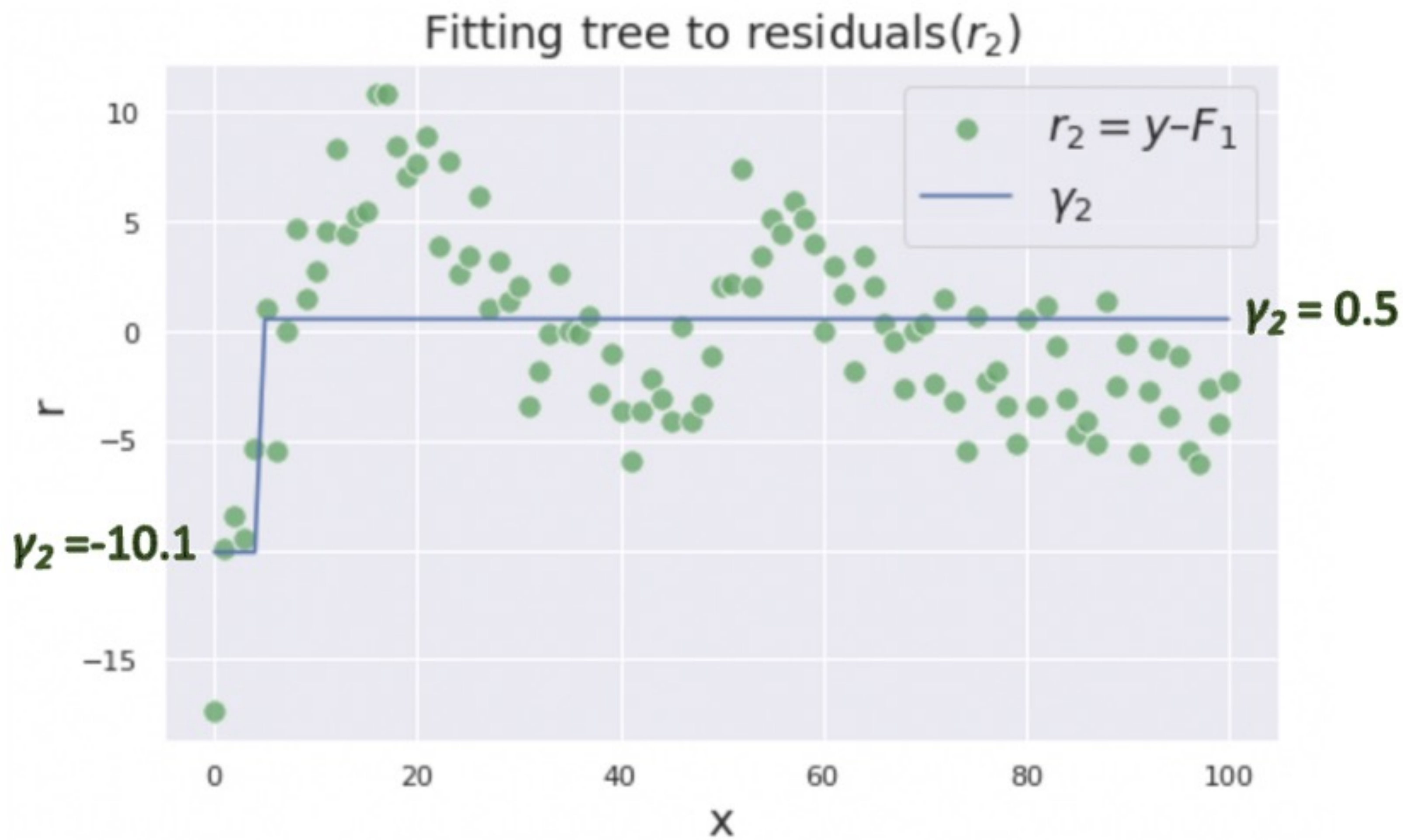
$$F_1 = \begin{cases} F_0 + \nu \cdot 6.0 & if\ x \leq 49.5 \\ F_0 - \nu \cdot 5.9 & otherwise \end{cases}$$

Credit: Tomonori Masui

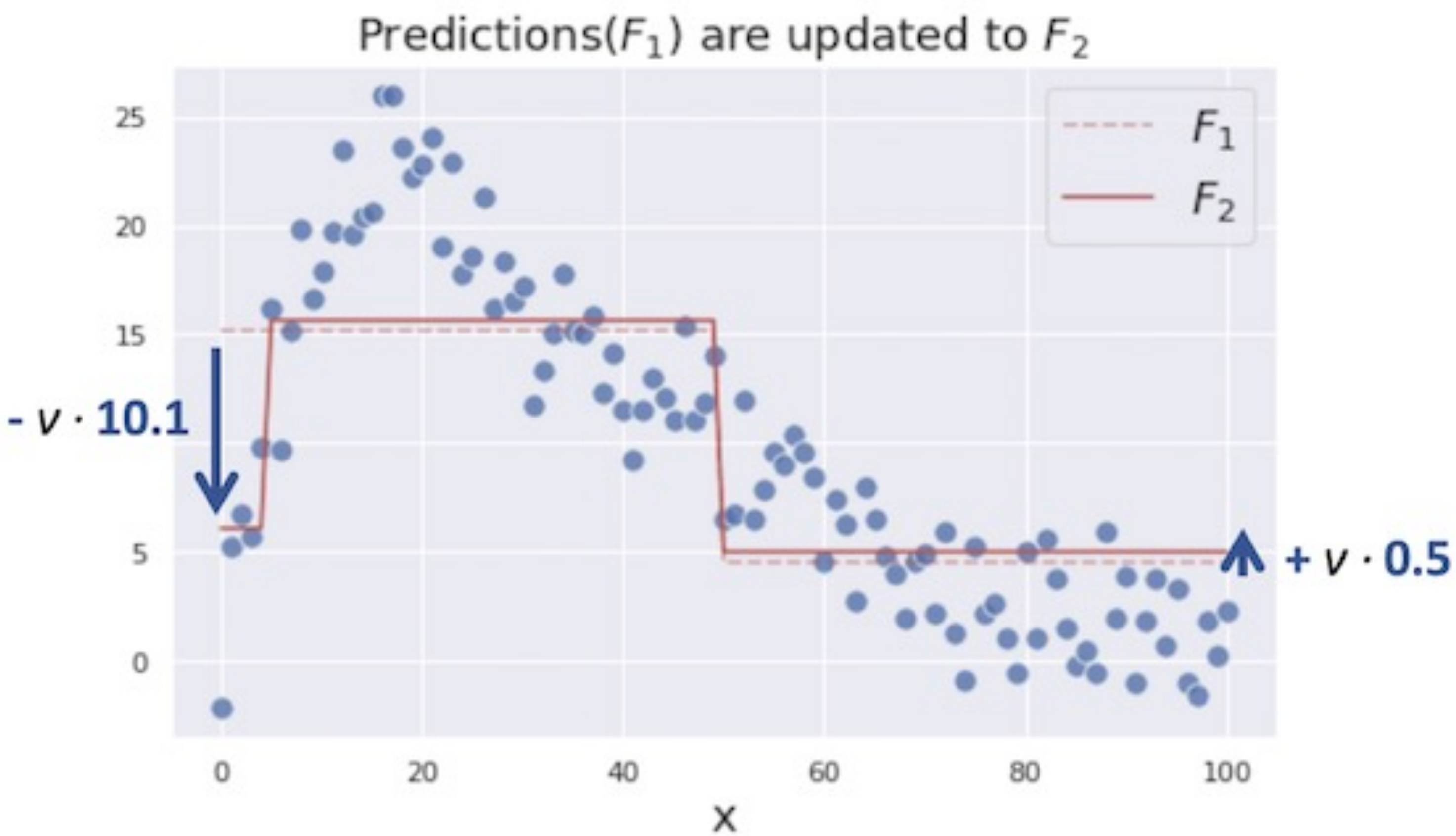


Credit: Tomonori Masui

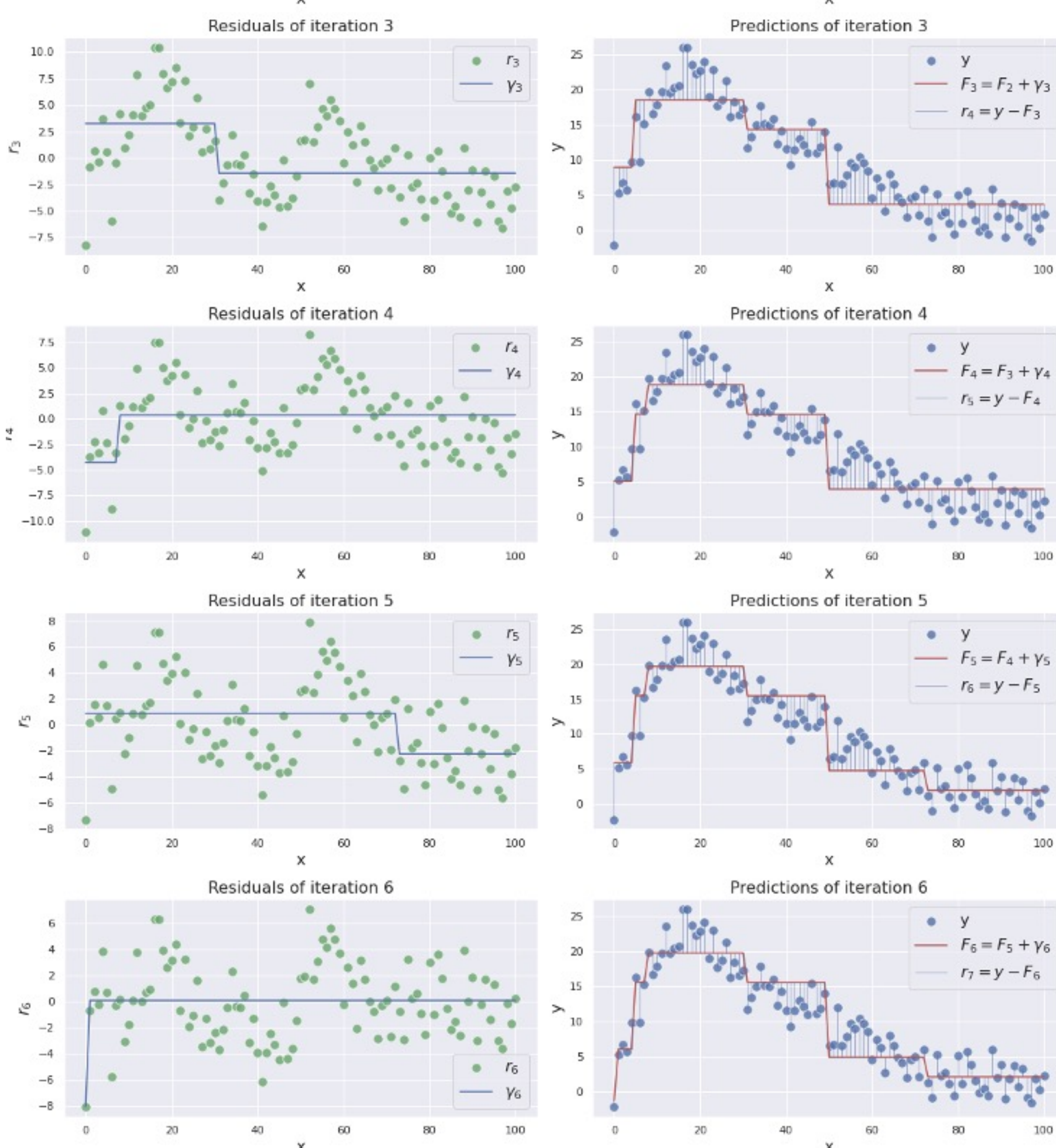




Credit: Tomonori Masui



Credit: Tomonori Masui



Credit: Tomonori Masui