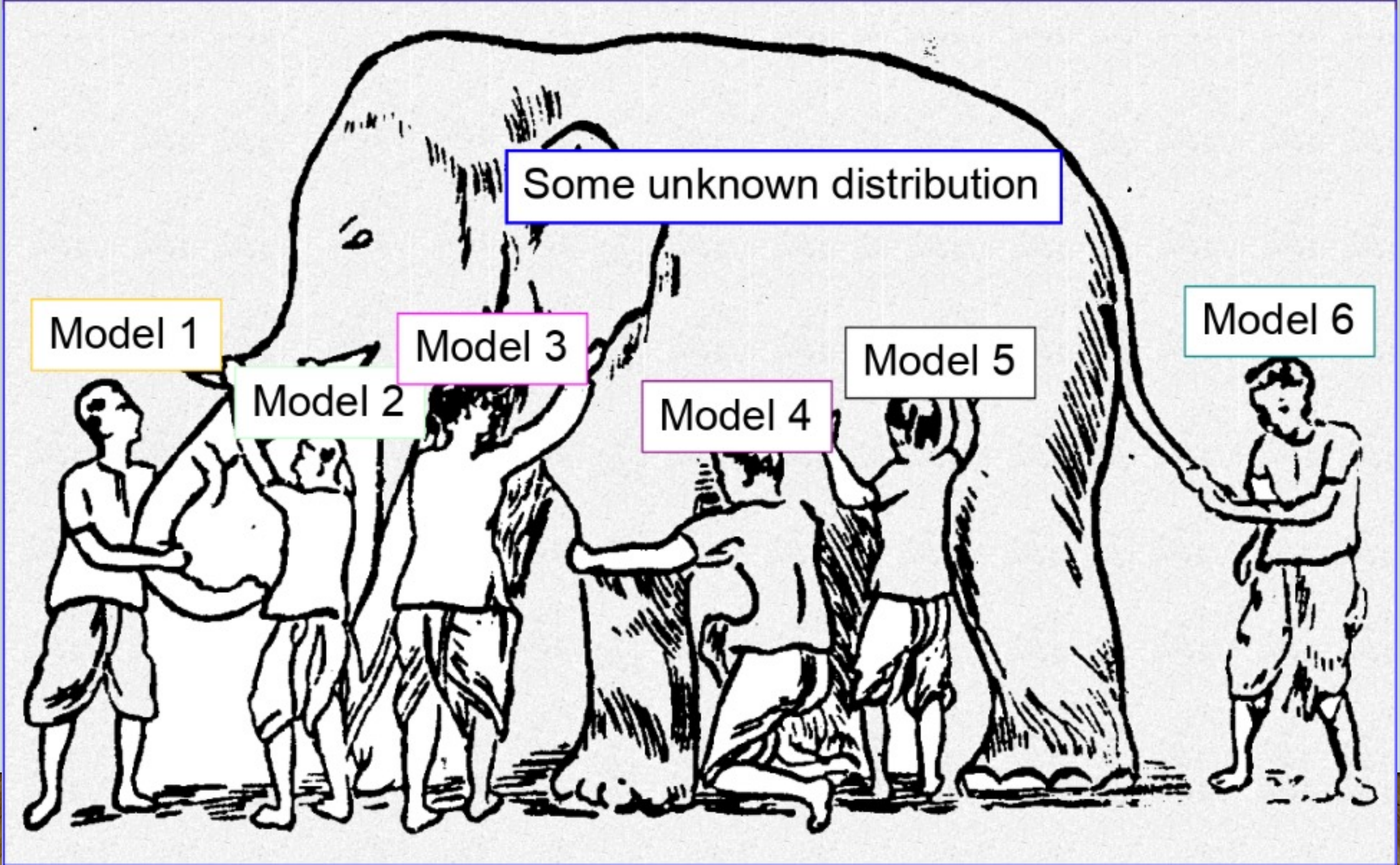
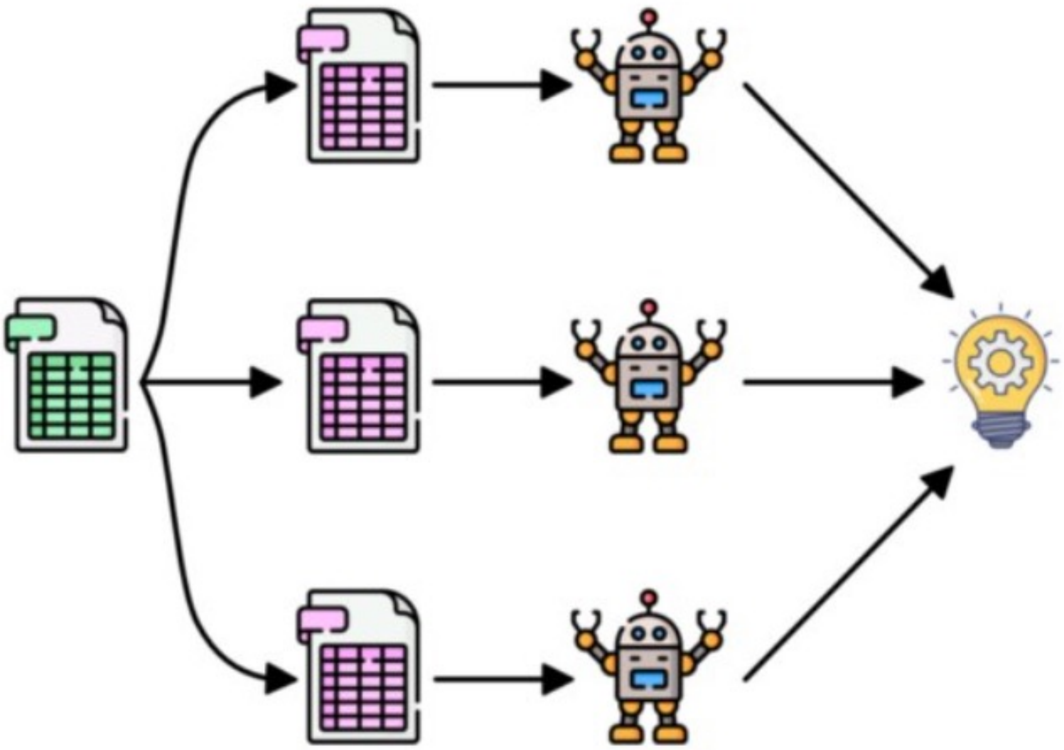
Gradient Boosting

TUAN NGUYEN

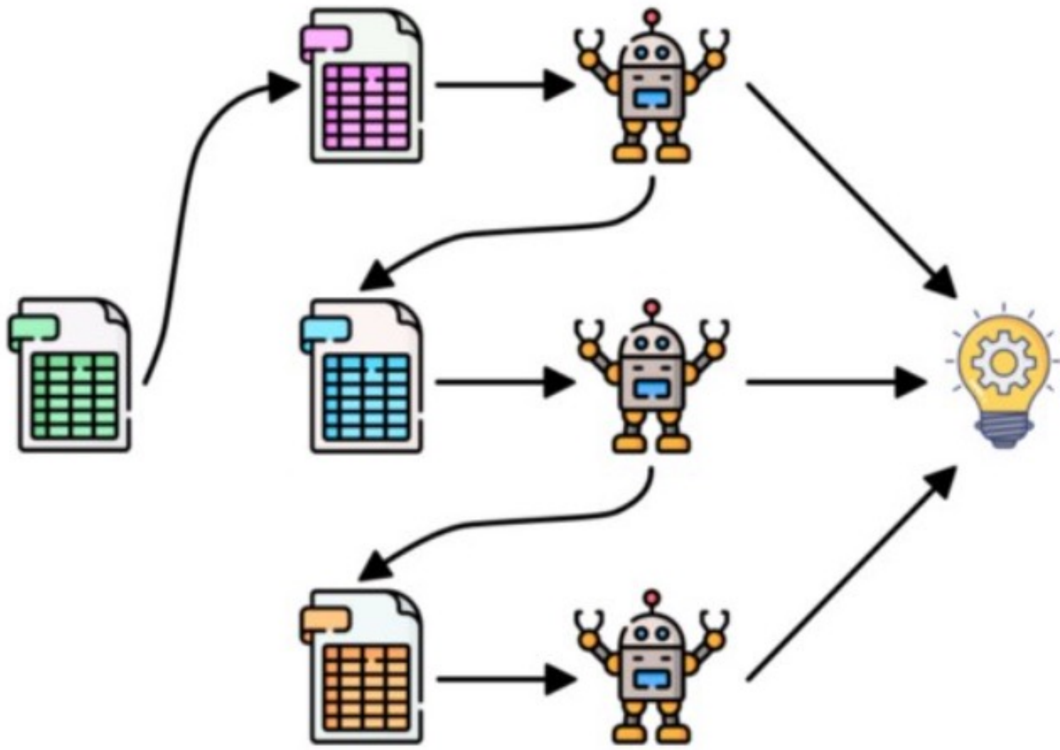

Some unknown distribution
Model 1
Model 2
Model 3
Model 4
Model 5
Model 6

Bagging



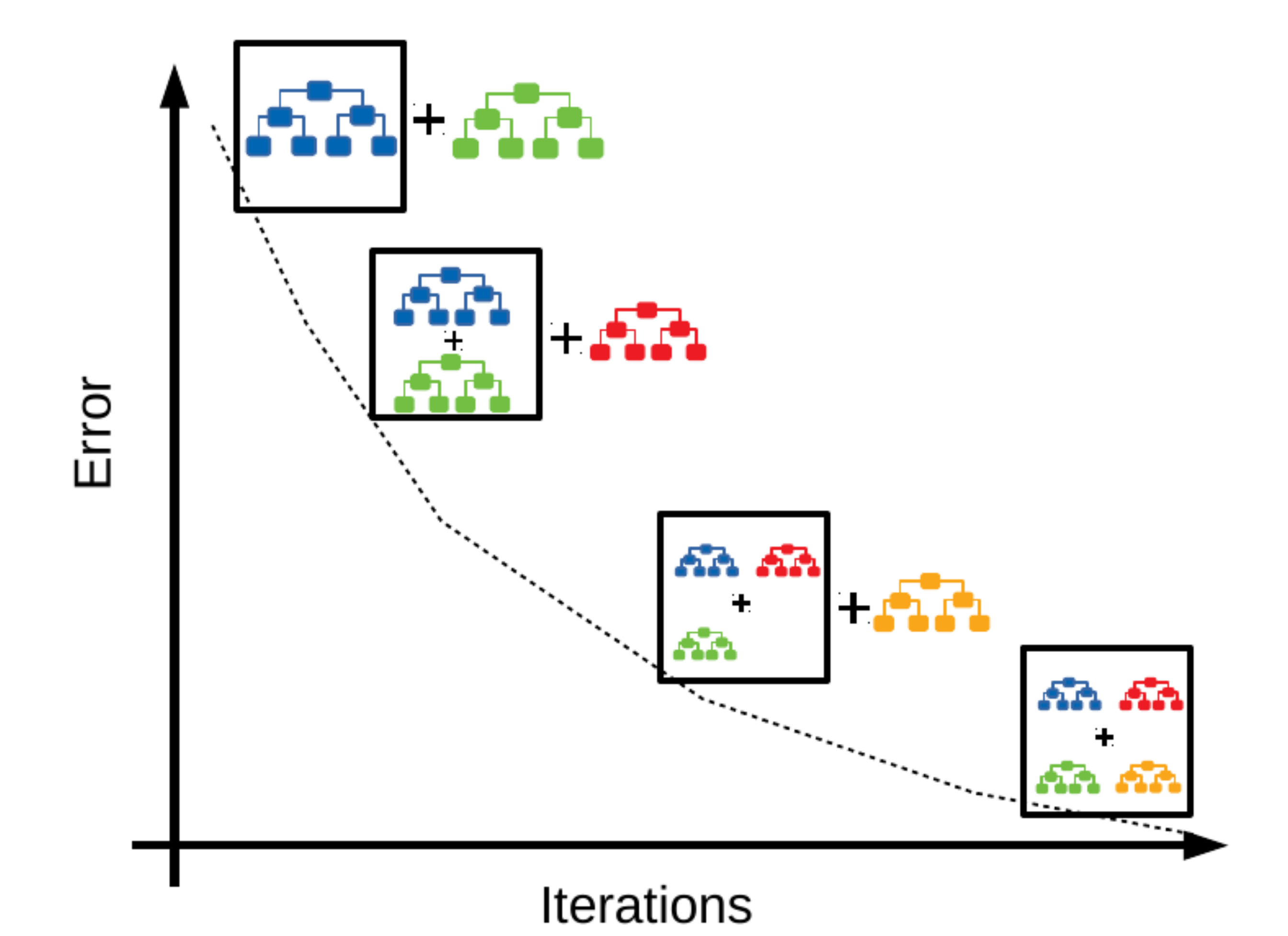
Parallel

Boosting

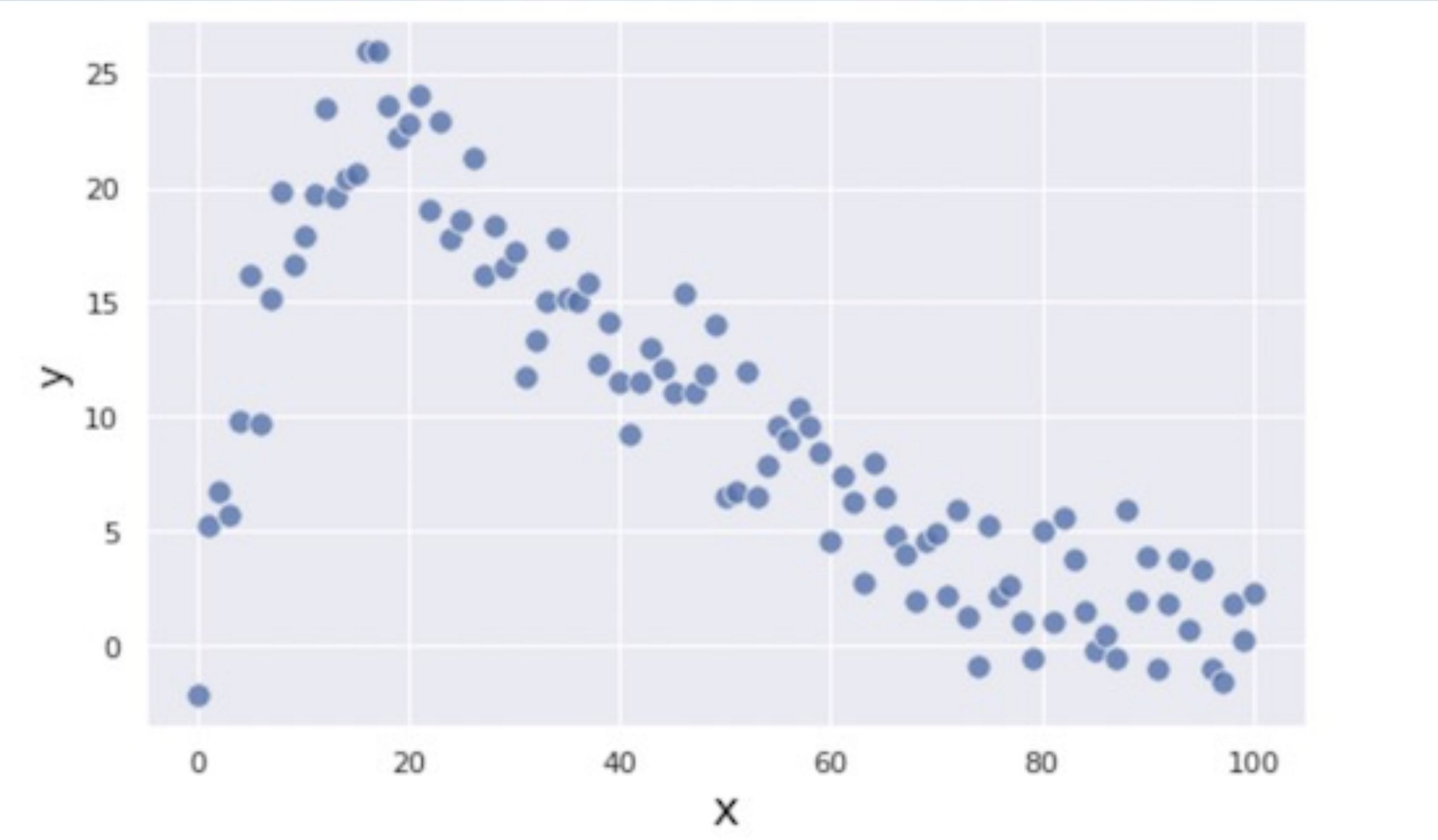


Sequential

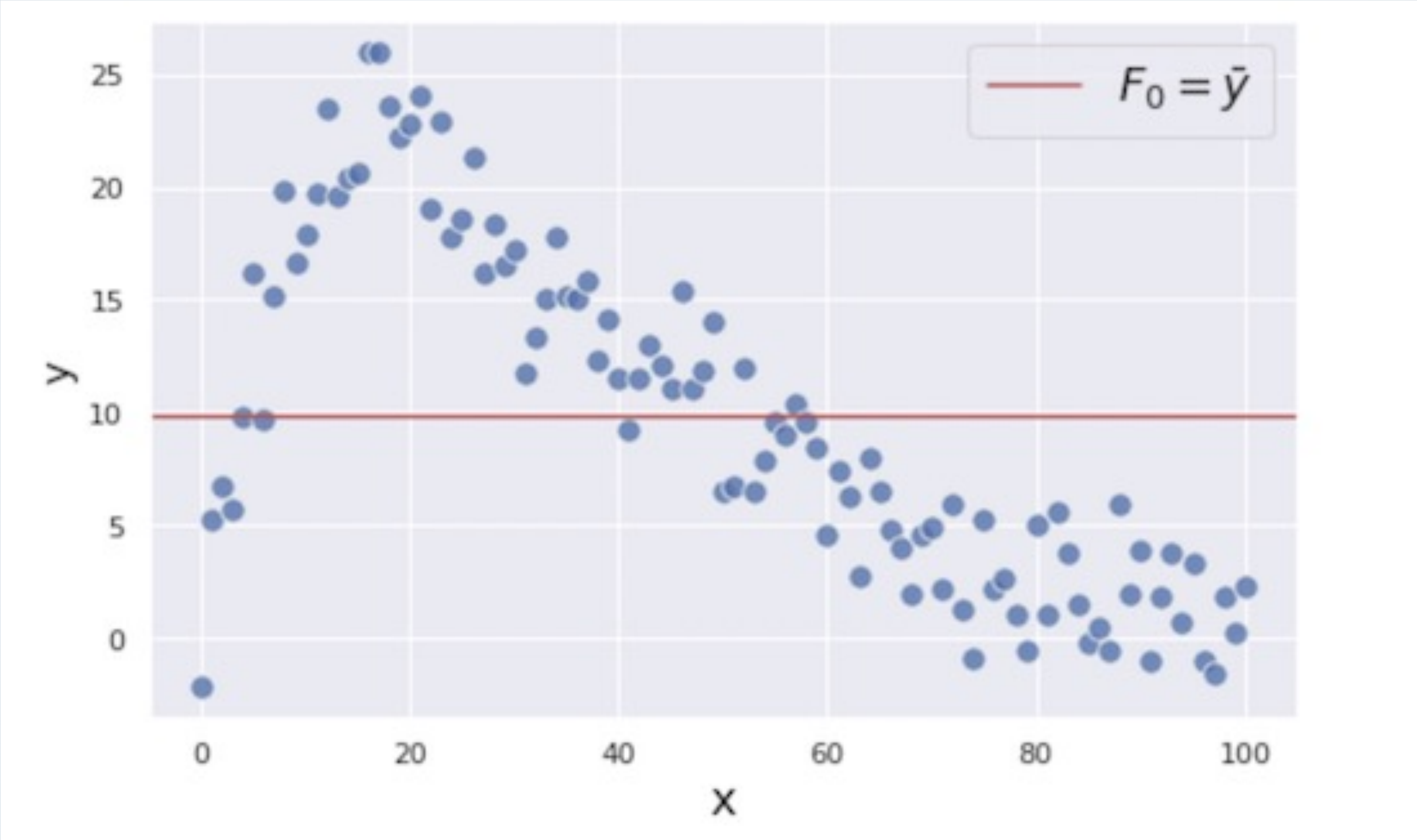
Gradient Boosting



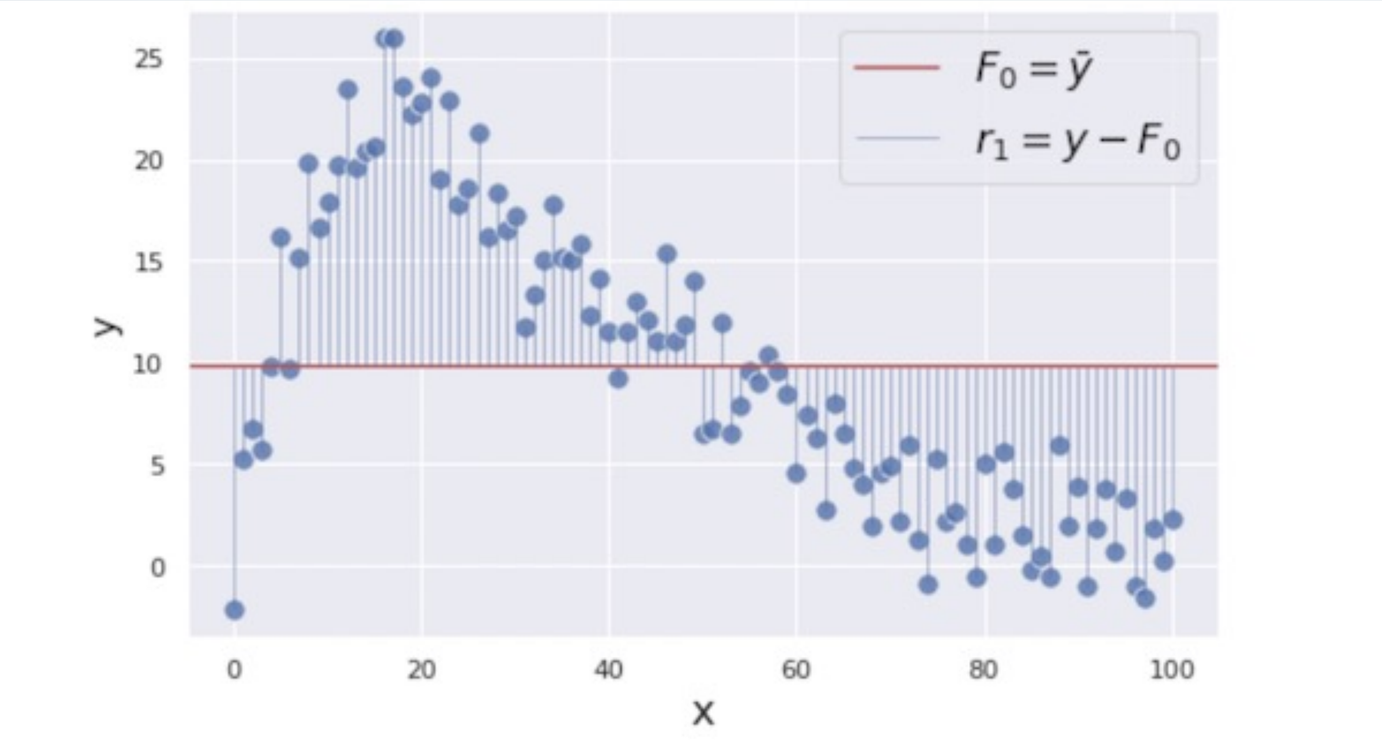
Dataset



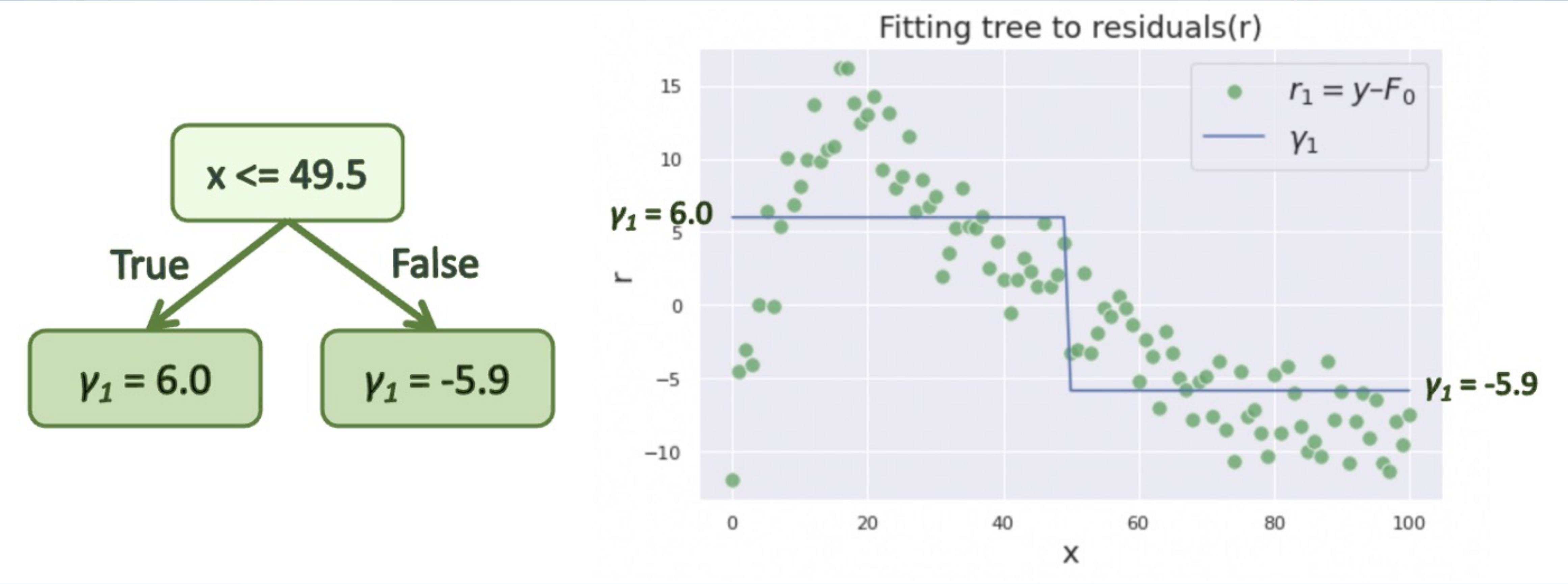
Simple prediction



Residual data



Residual model



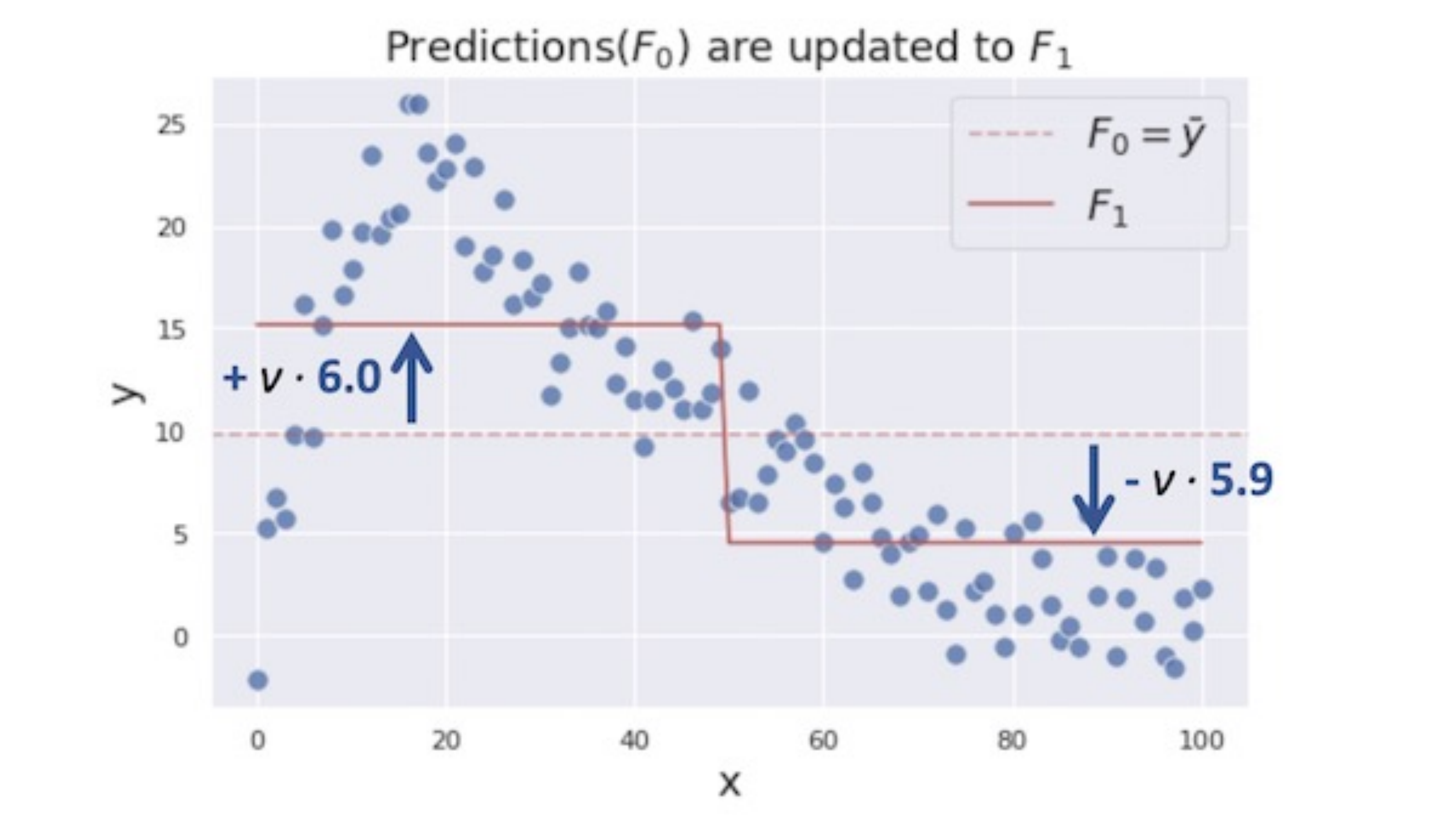
Prediction change

$$F_1 = F_0 + \nu \cdot \gamma_1$$

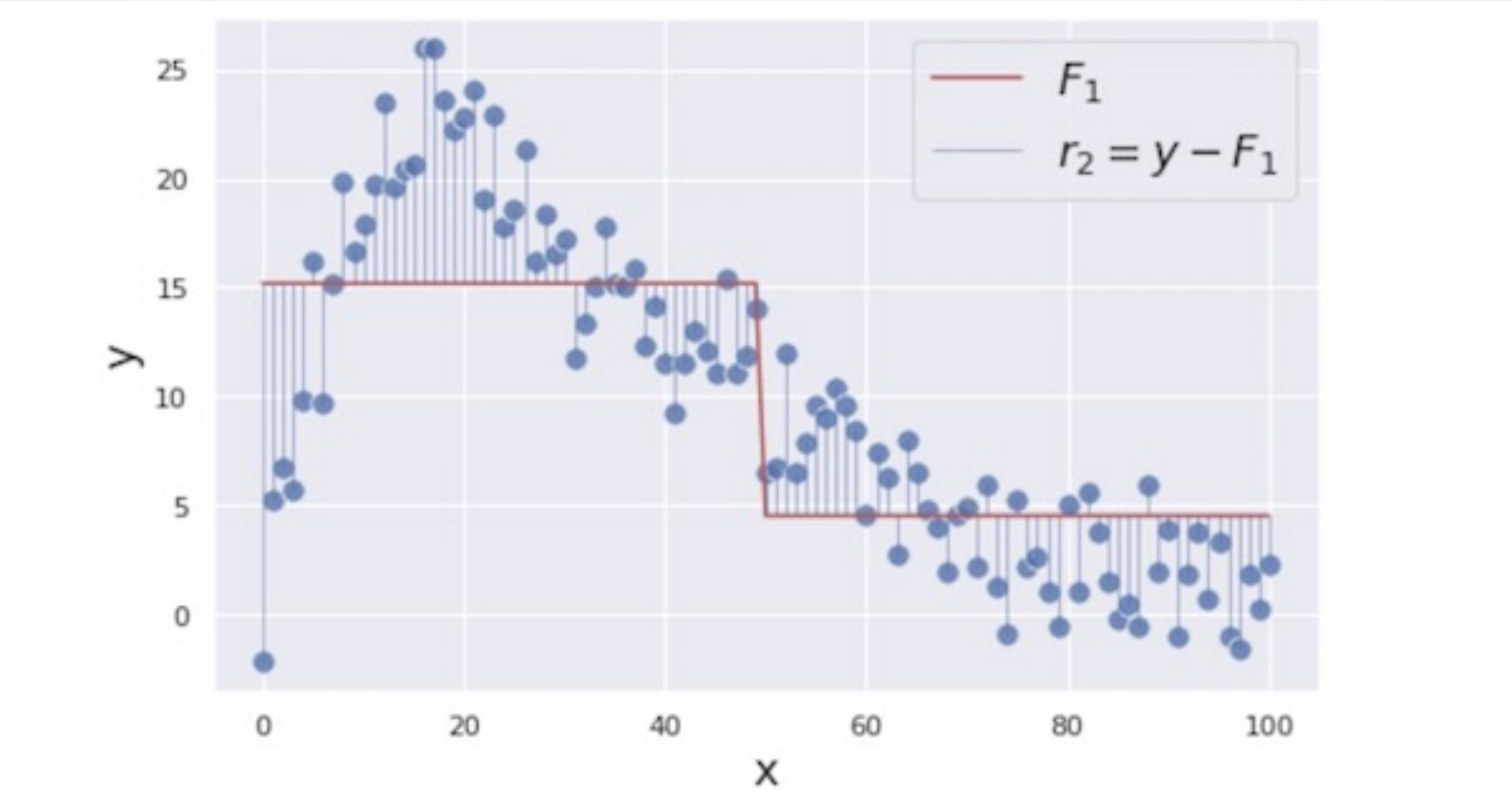
In fact, gradient boosting algorithm does not simply add *γ* to *F* as it makes the model overfit to the training data. Instead, *γ* is scaled down by **learning rate** *ν* which ranges between 0 and 1

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

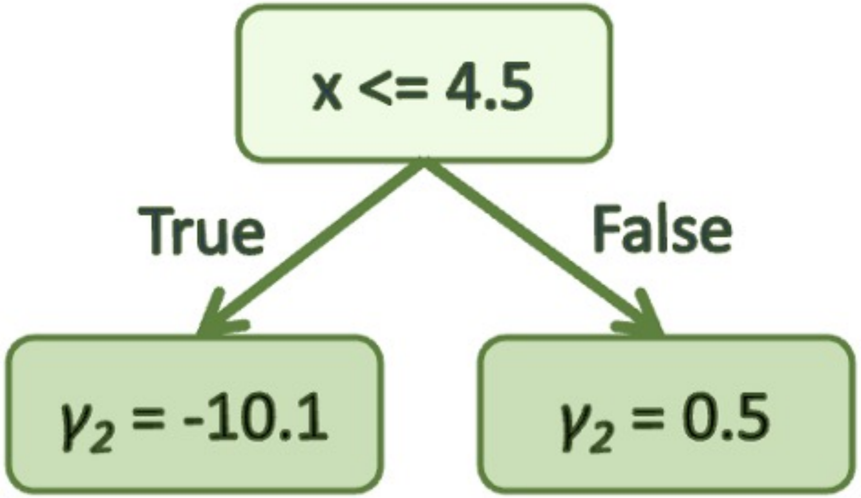
Model prediction update

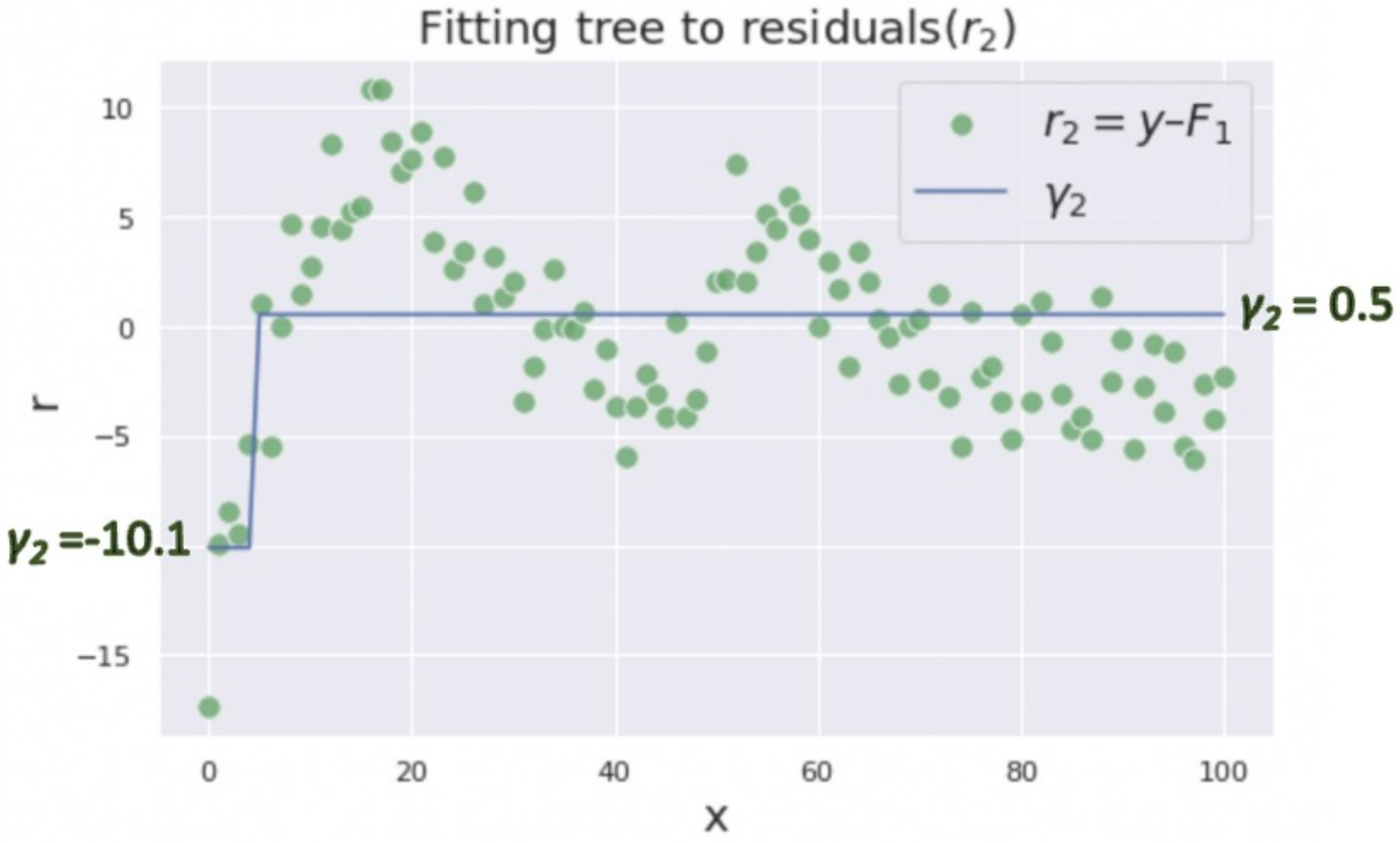


Residual update

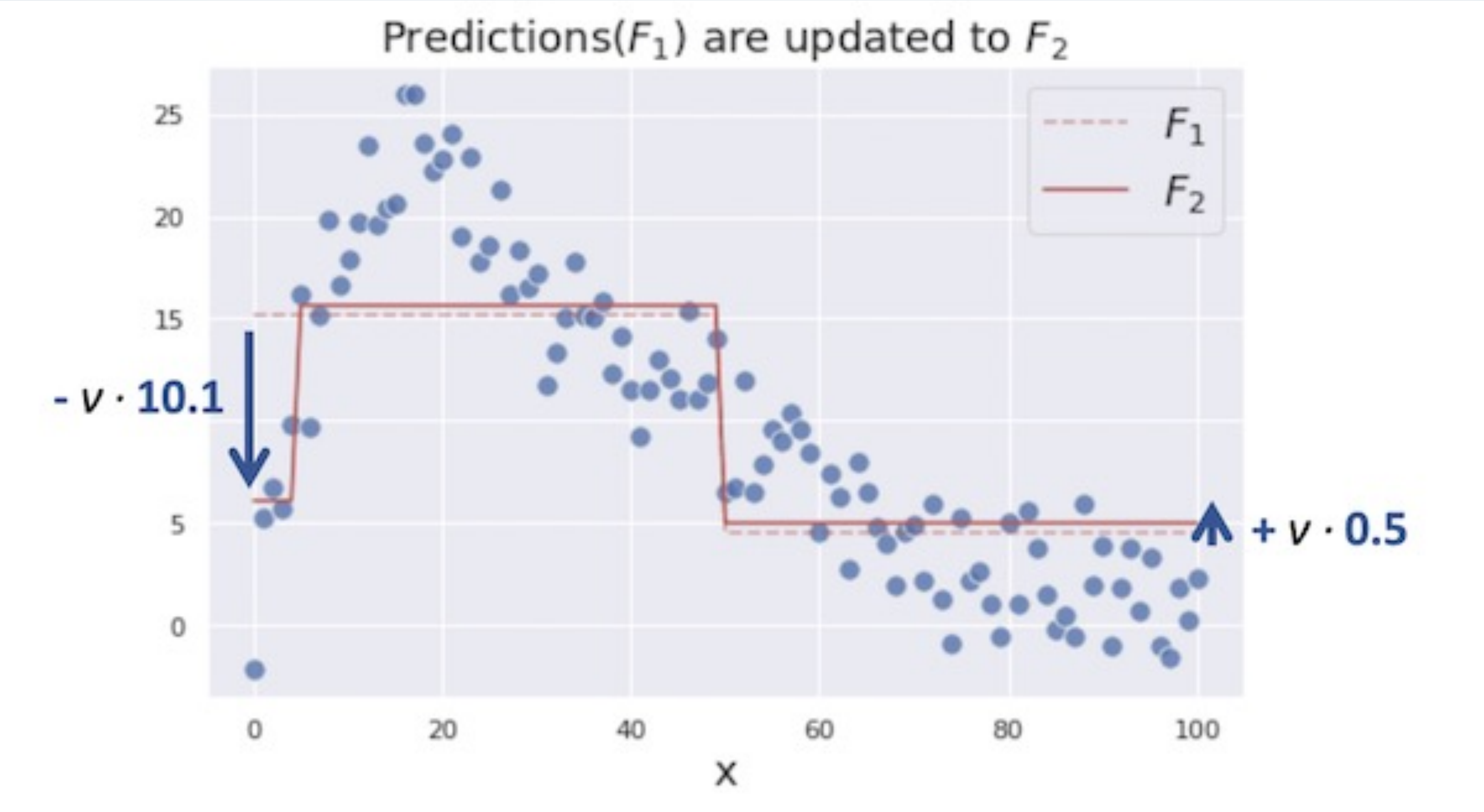


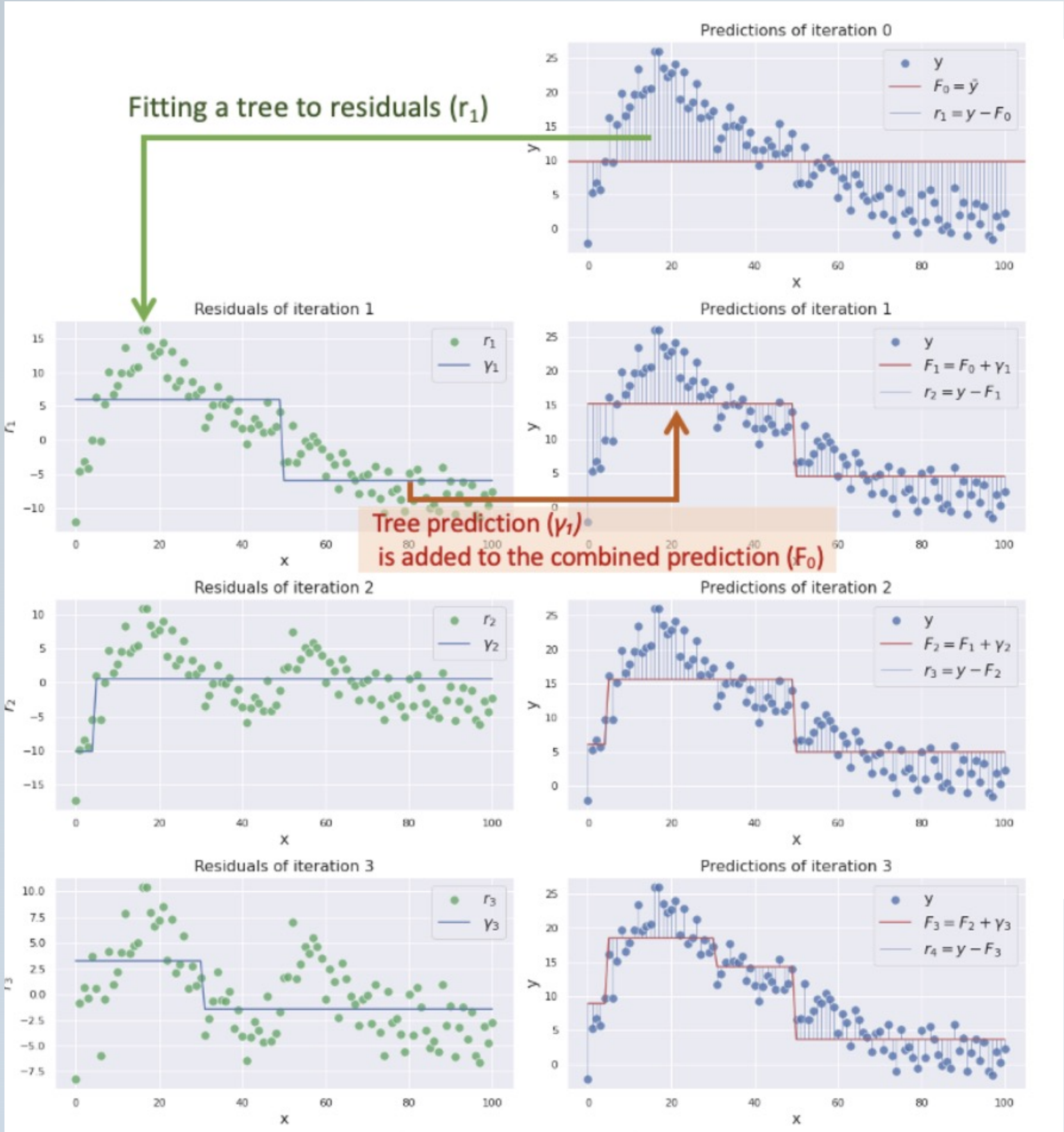
Residual model (2)



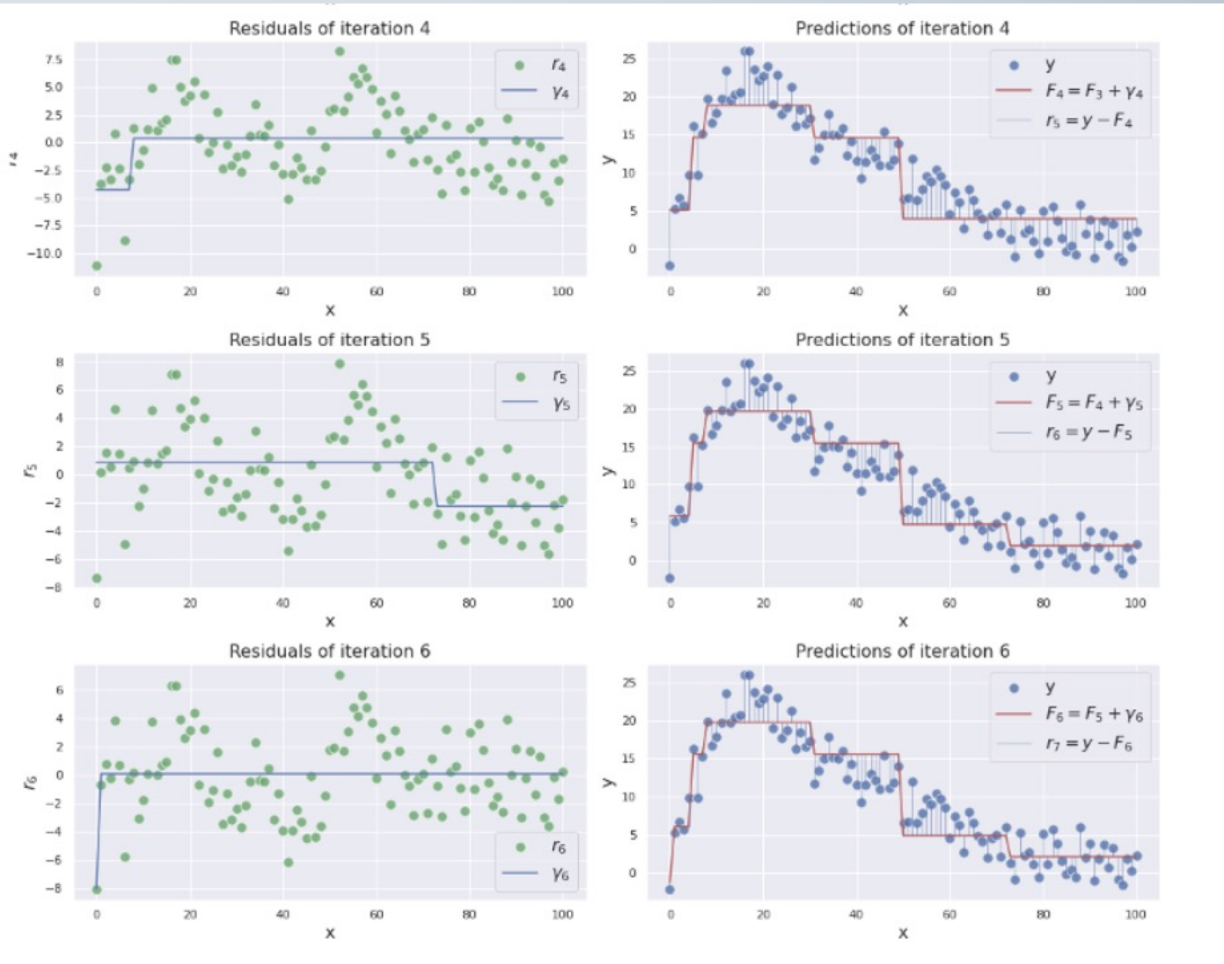


Model prediction update





Fitting a tree to residuals (r_1)
Tree prediction (γ_1) is added to the combined prediction (F_0)
Predictions of iteration 0
y
$F_0 = \hat{y}$
$r_1 = y - F_0$
Residuals of iteration 1
r_1
γ_1
Predictions of iteration 1
$F_1 = F_0 + \gamma_1$
$r_2 = y - F_1$
Residuals of iteration 2
r_2
γ_2
Predictions of iteration 2
$F_2 = F_1 + \gamma_2$
$r_3 = y - F_2$
Residuals of iteration 3
r_3
γ_3
Predictions of iteration 3
$F_3 = F_2 + \gamma_3$
$r_4 = y - F_3$
x



Residuals of iteration 4
r_4
γ_4
Predictions of iteration 4
y
$F_4 = F_3 + \gamma_4$
$r_5 = y - F_4$
Residuals of iteration 5
r_5
γ_5
Predictions of iteration 5
$F_5 = F_4 + \gamma_5$
$r_6 = y - F_5$
Residuals of iteration 6
r_6
γ_6
Predictions of iteration 6
$F_6 = F_5 + \gamma_6$
$r_7 = y - F_6$
x

Algorithms

Gradient Boosting Algorithm

1. Initialize model with a constant value:

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma)$$

2. for $m = 1 \text{ } to \text{ } M$:

2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i,\ F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad for \ i = 1,\ldots,n$

2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\ldots,J_m$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma) \quad for \ j = 1,\ldots,J_m$

2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

More…

- https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502
- https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-2-classification-d3ed8f56541e