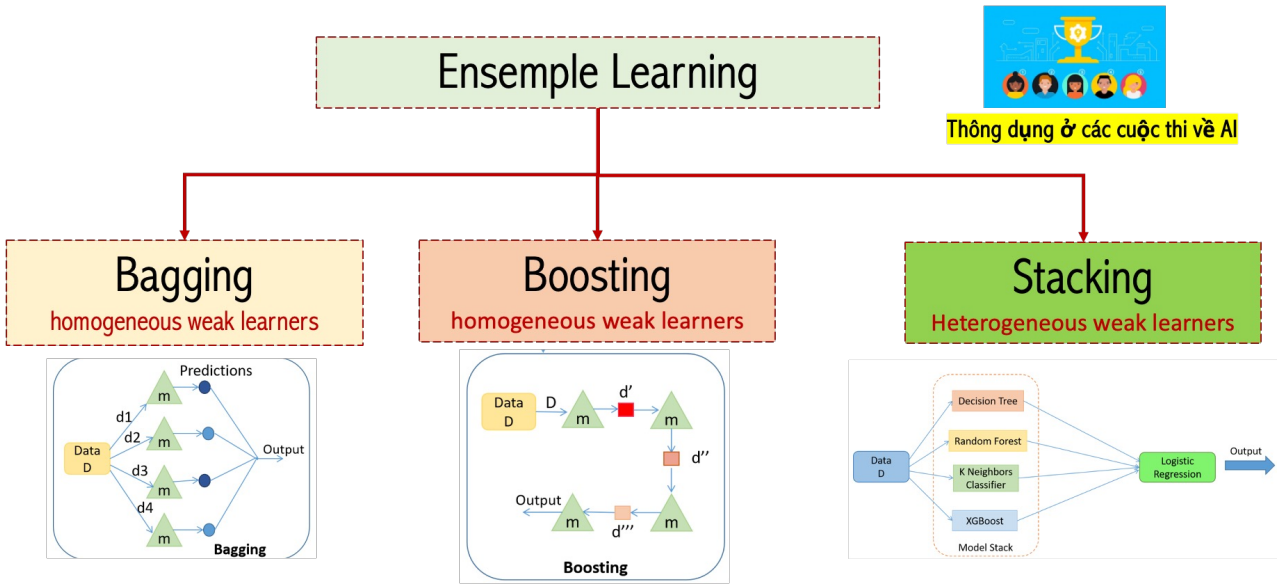
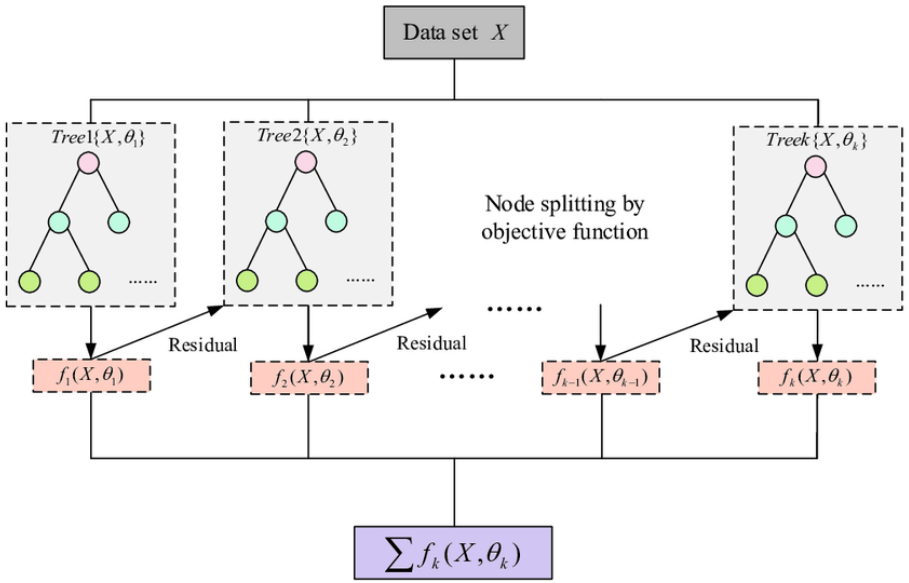
AI VIETNAM
All-in-One Course

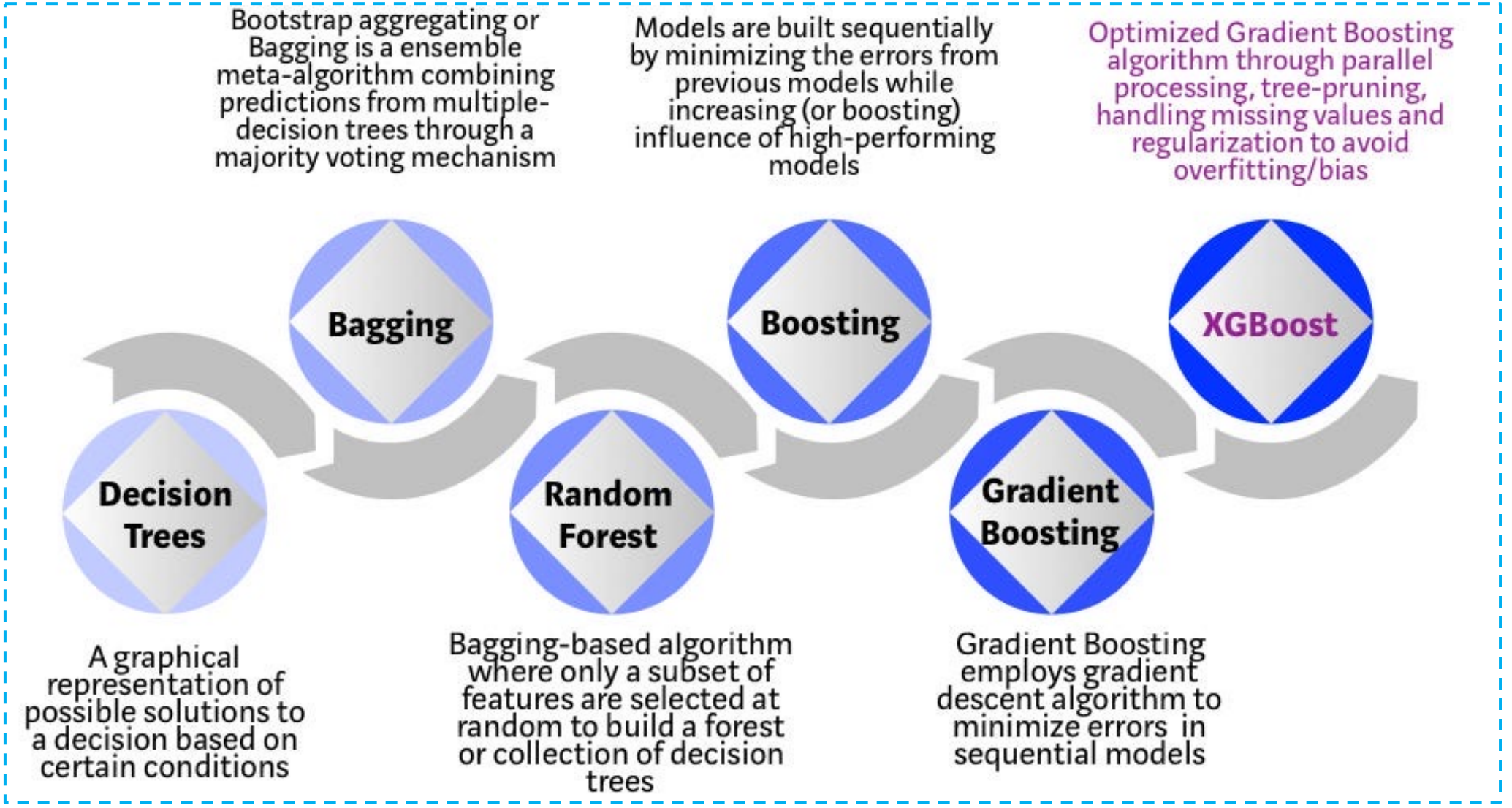
AdaBoost & Gradient Boost





Vinh Dinh Nguyen
PhD in Computer Science

Decision Tree and Its Variance



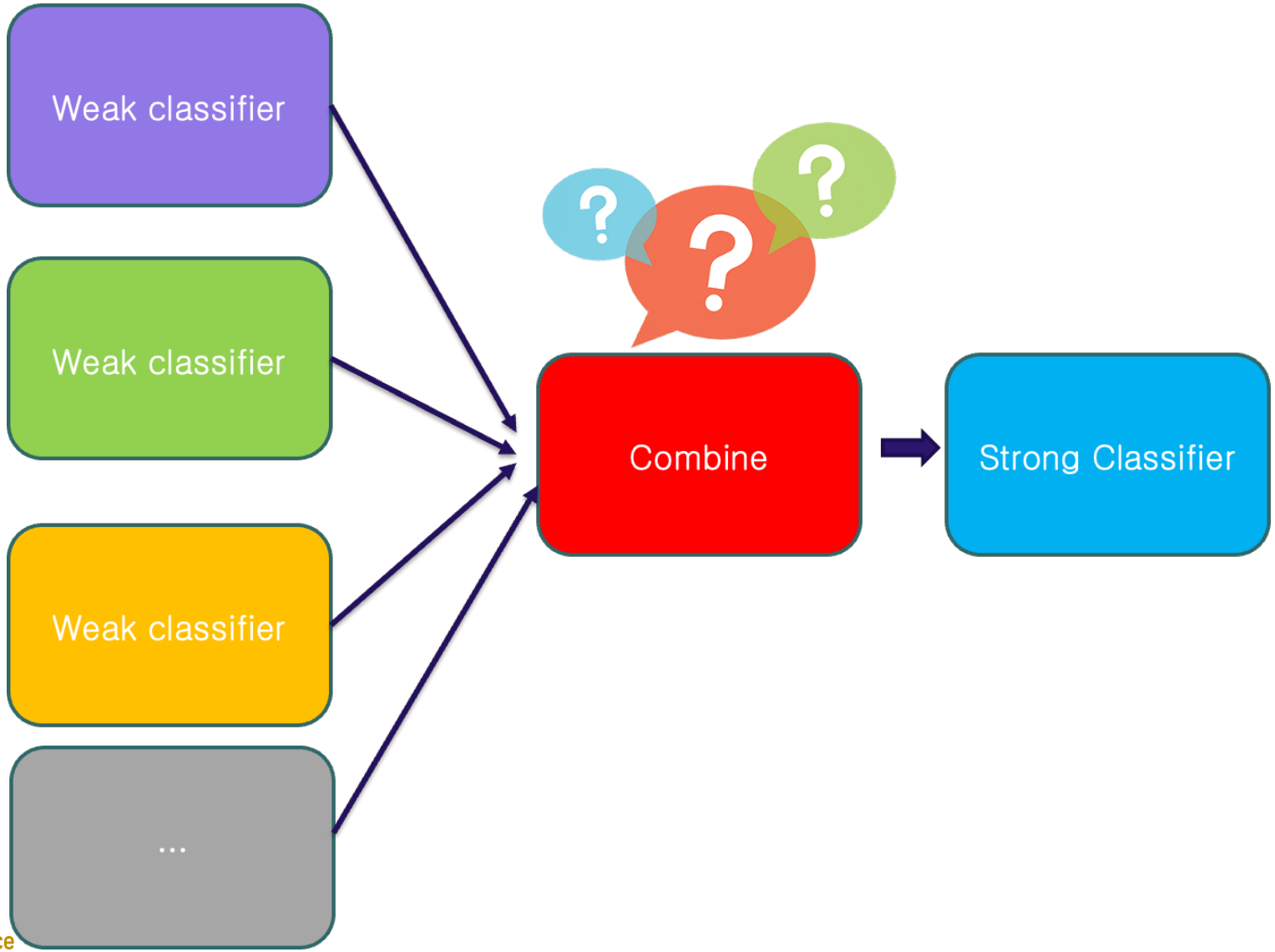
Outline

- Boosting Technique
- AdaBoost and Its Application
- Gradient Boost and Its Application

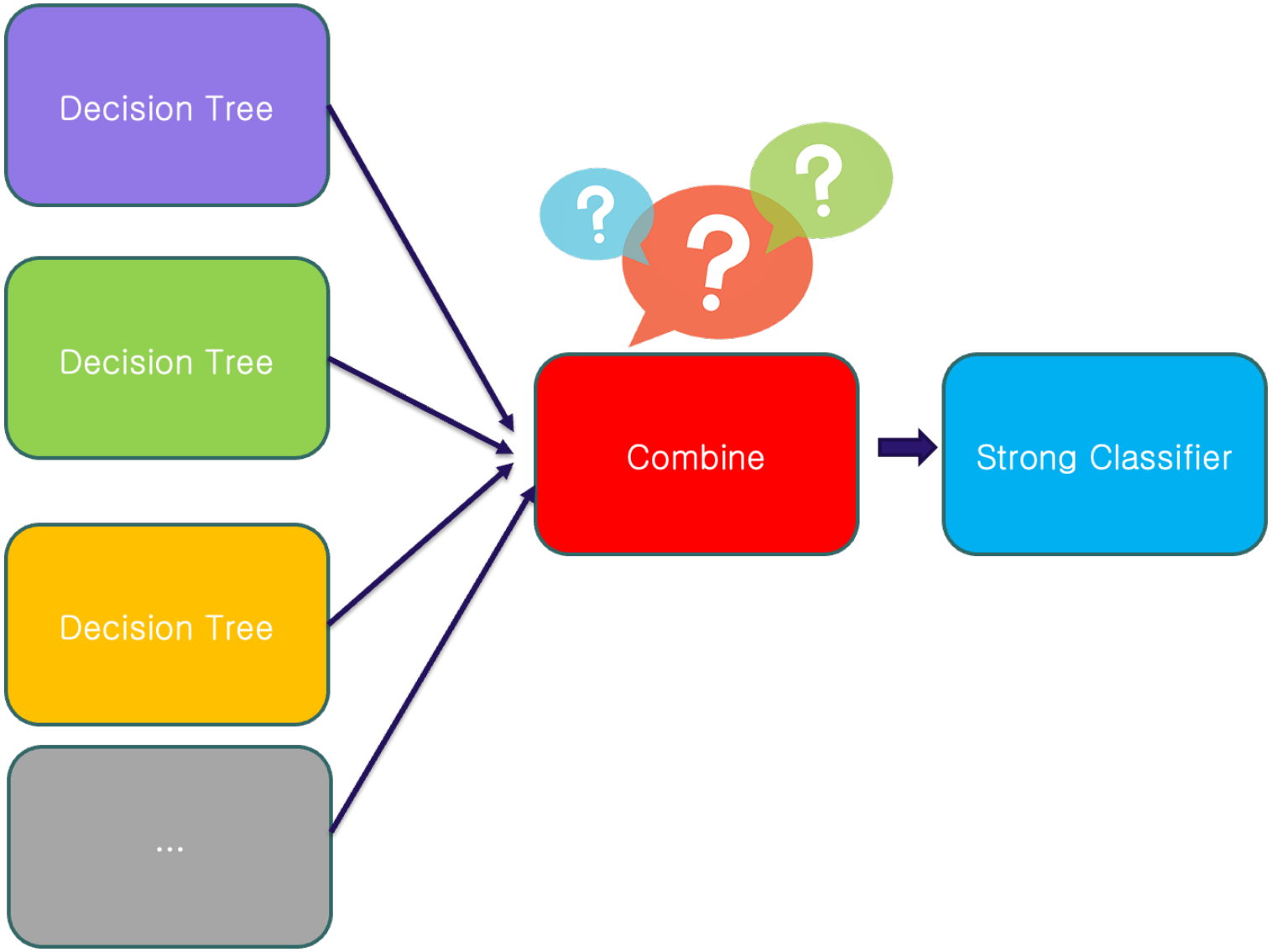
Outline

- Boosting Technique
- AdaBoost and Its Application
- Gradient Boost and Its Application

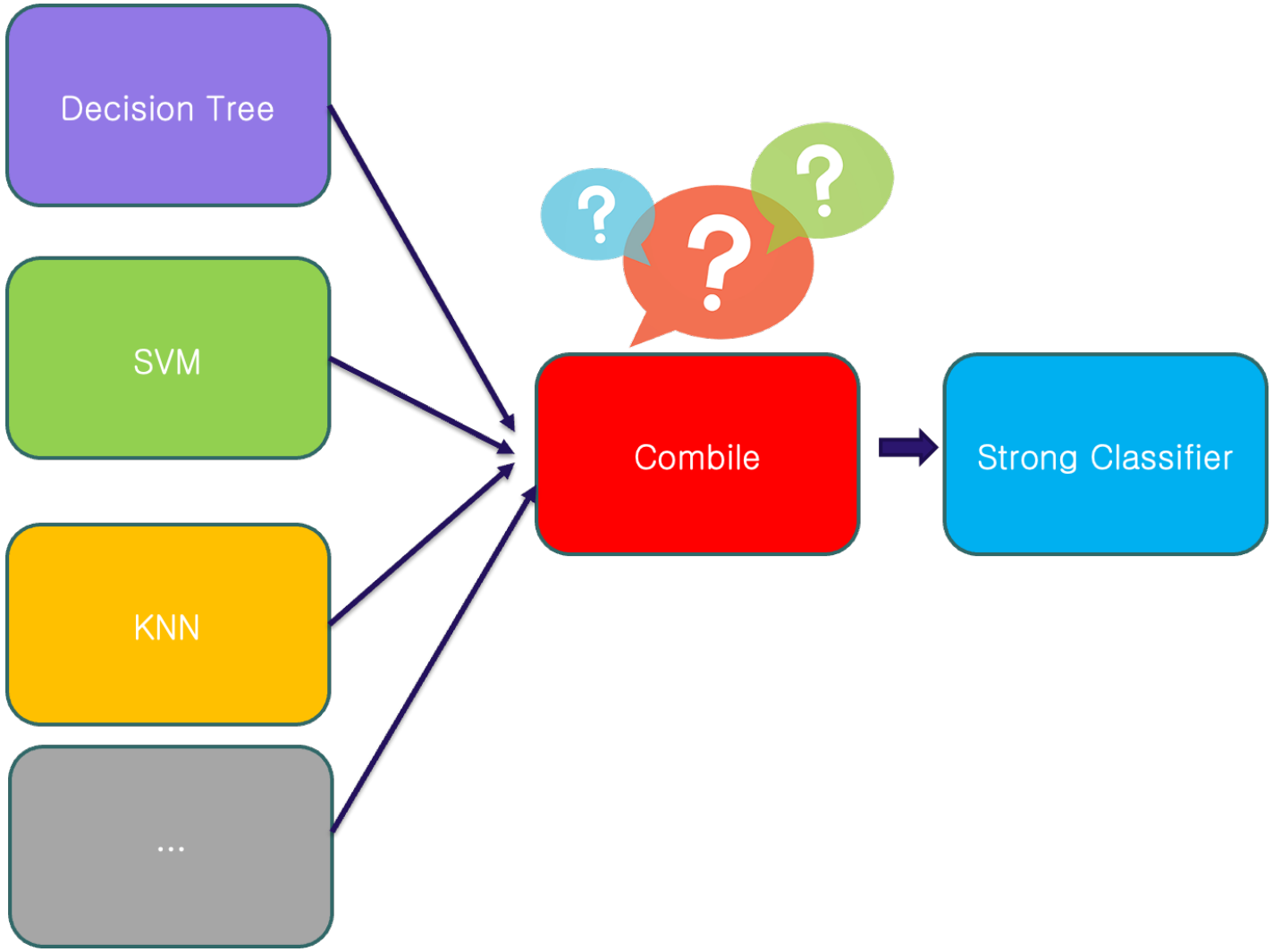
What is an Ensemple Learning?



Homogeneous Approach



Heterogeneous Approach

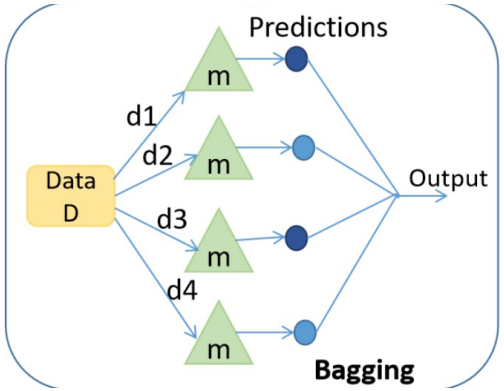


Ensemple Learning Techniques

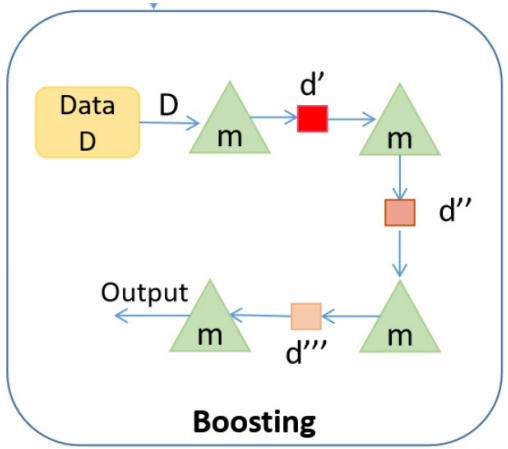
Ensemple Learning

Thông dụng ở các cuộc thi về AI

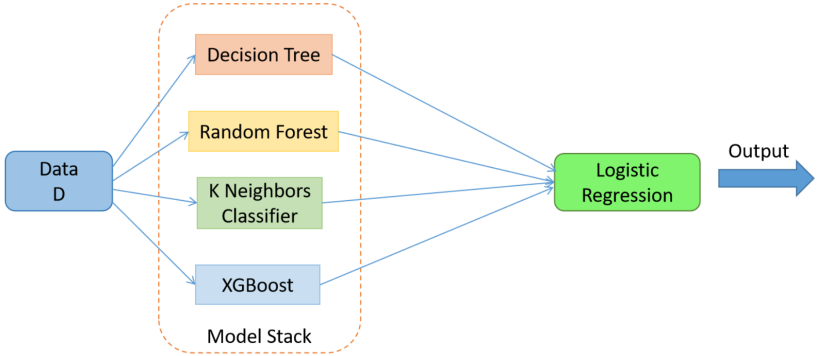
Bagging
homogeneous weak learners



Boosting
homogeneous weak learners



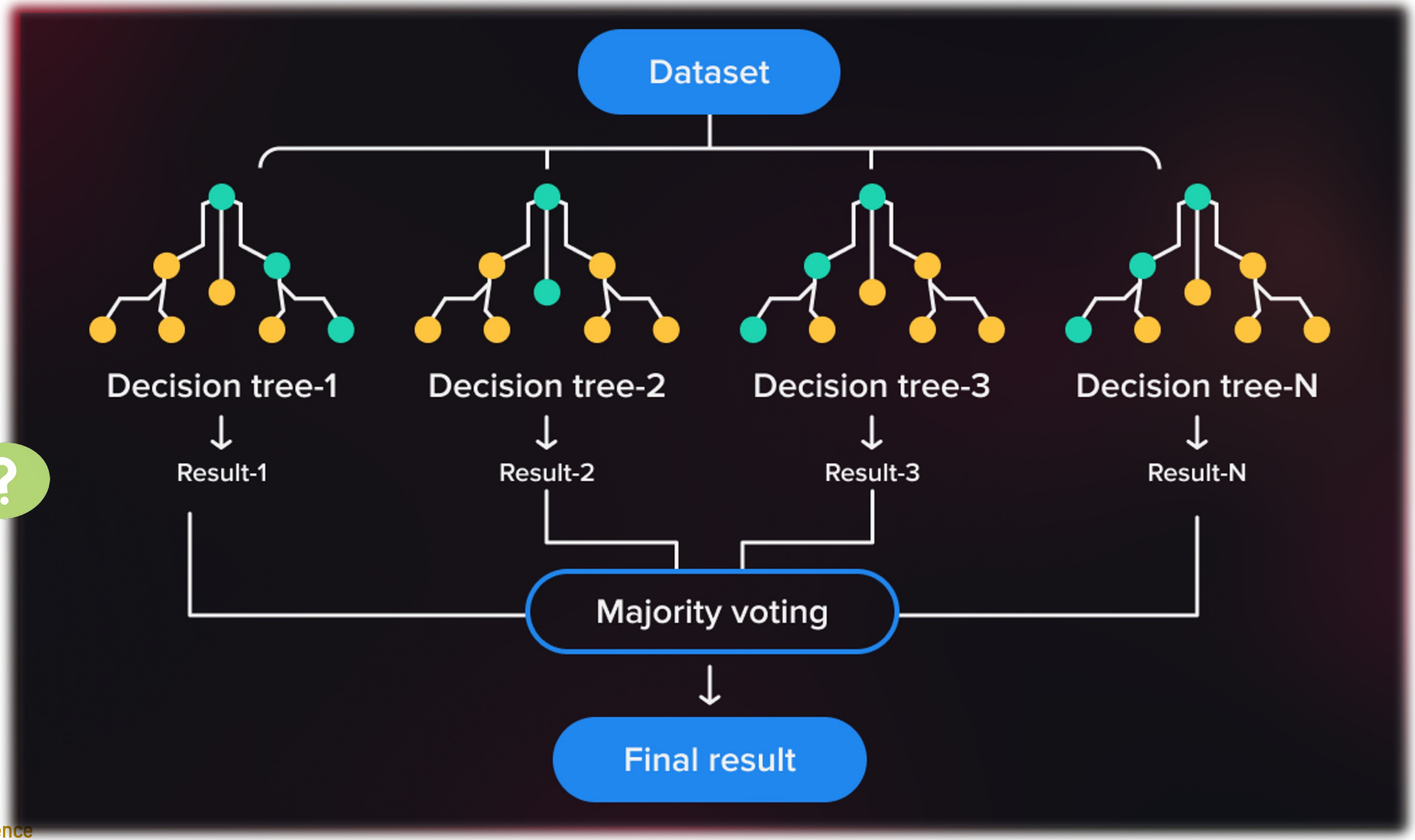
Stacking
Heterogeneous weak learners



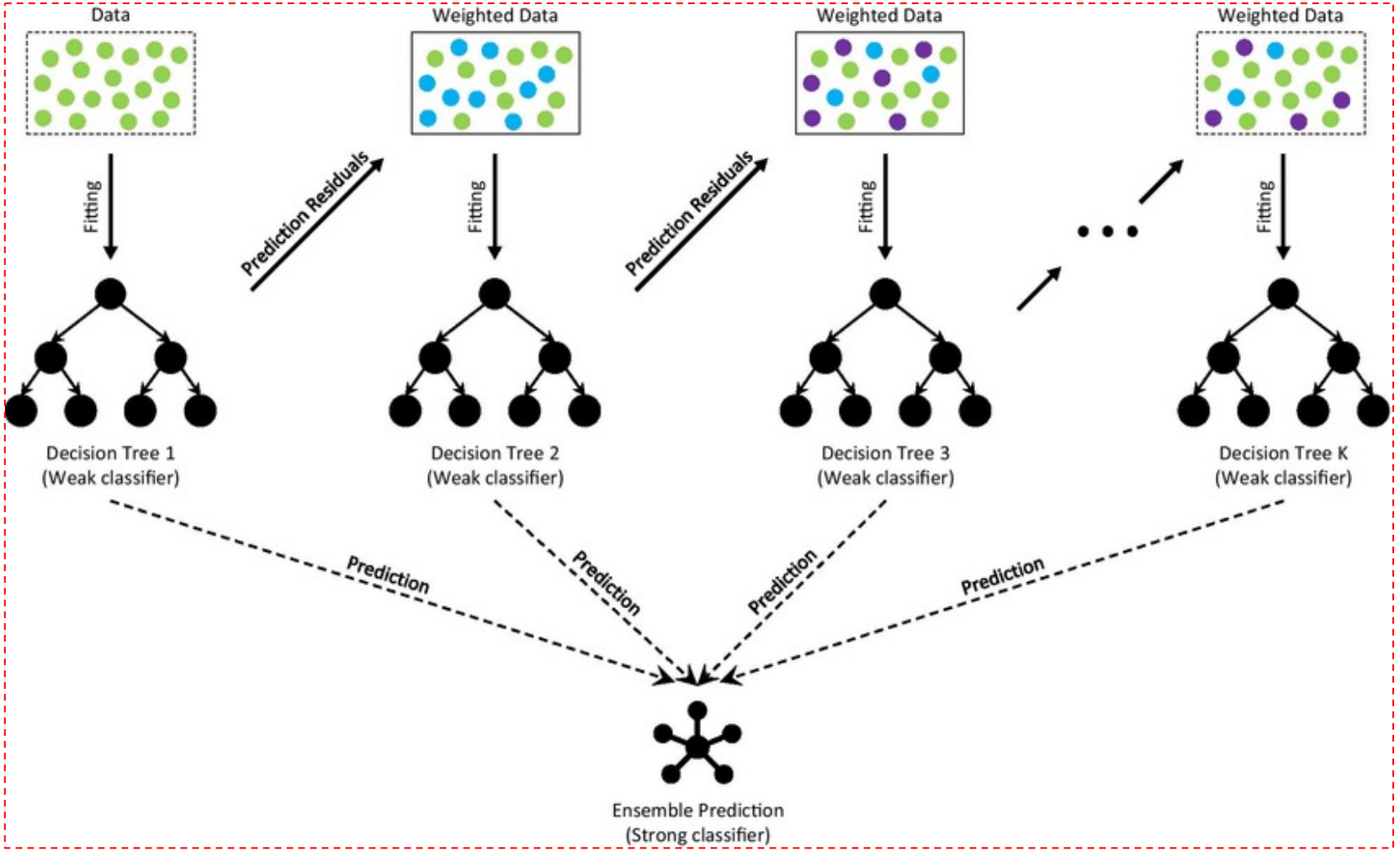
Bagging-based Method

Random Forest

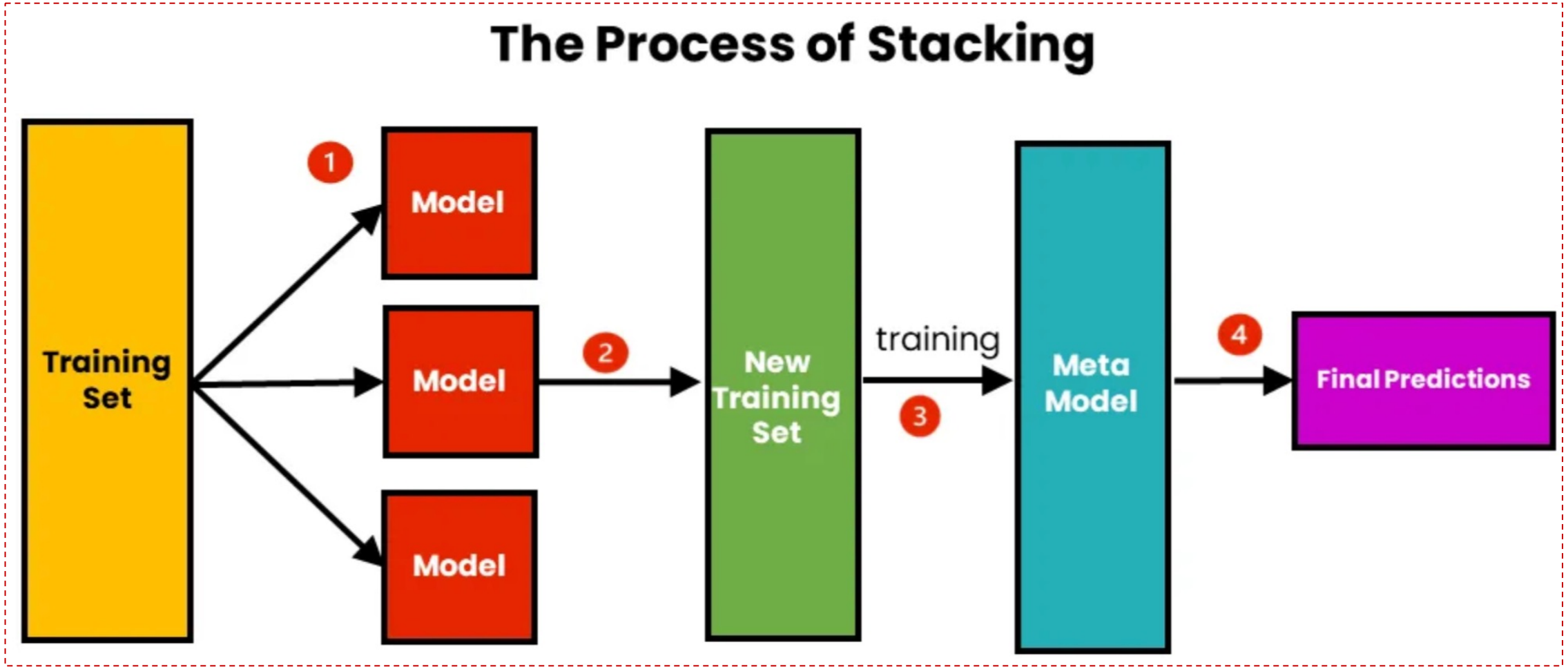
Last Week



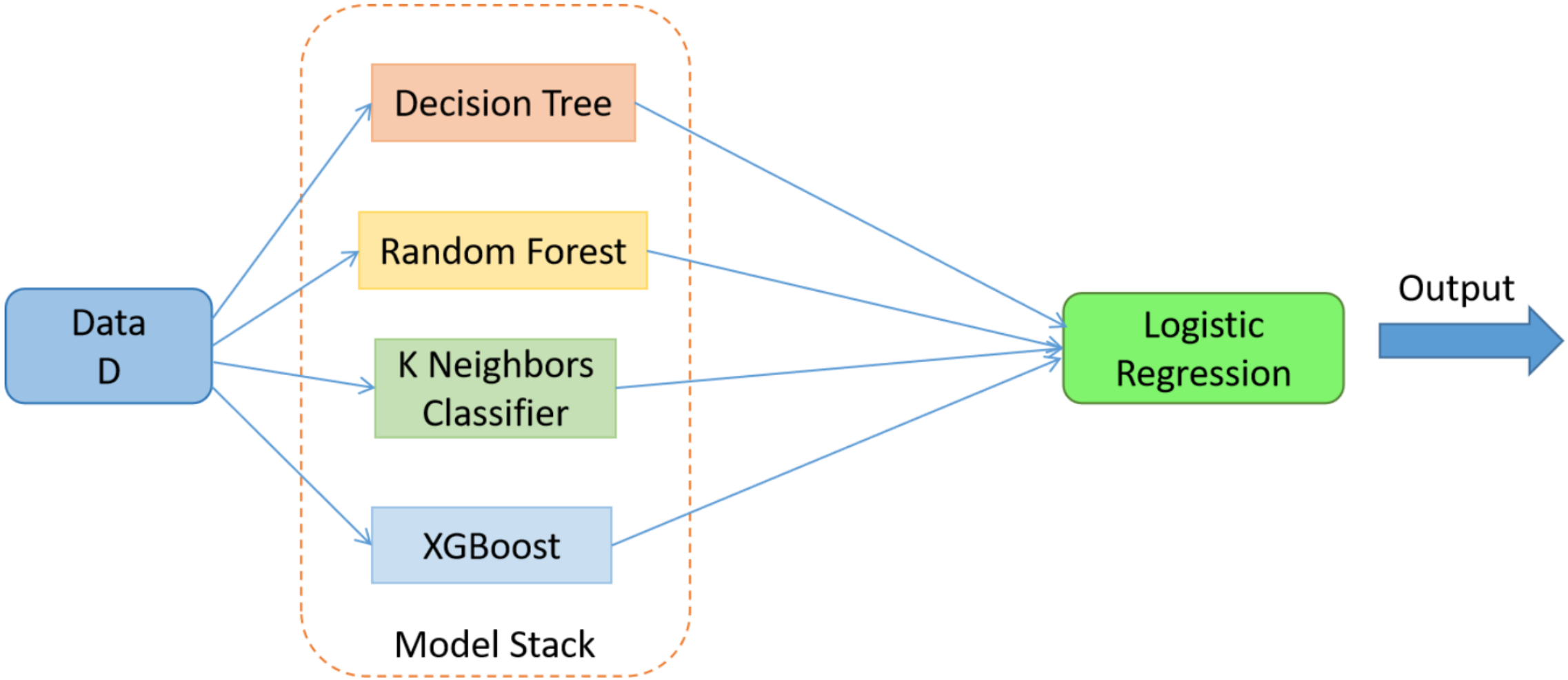
Boosting-Based Method



Stacking-Based Method

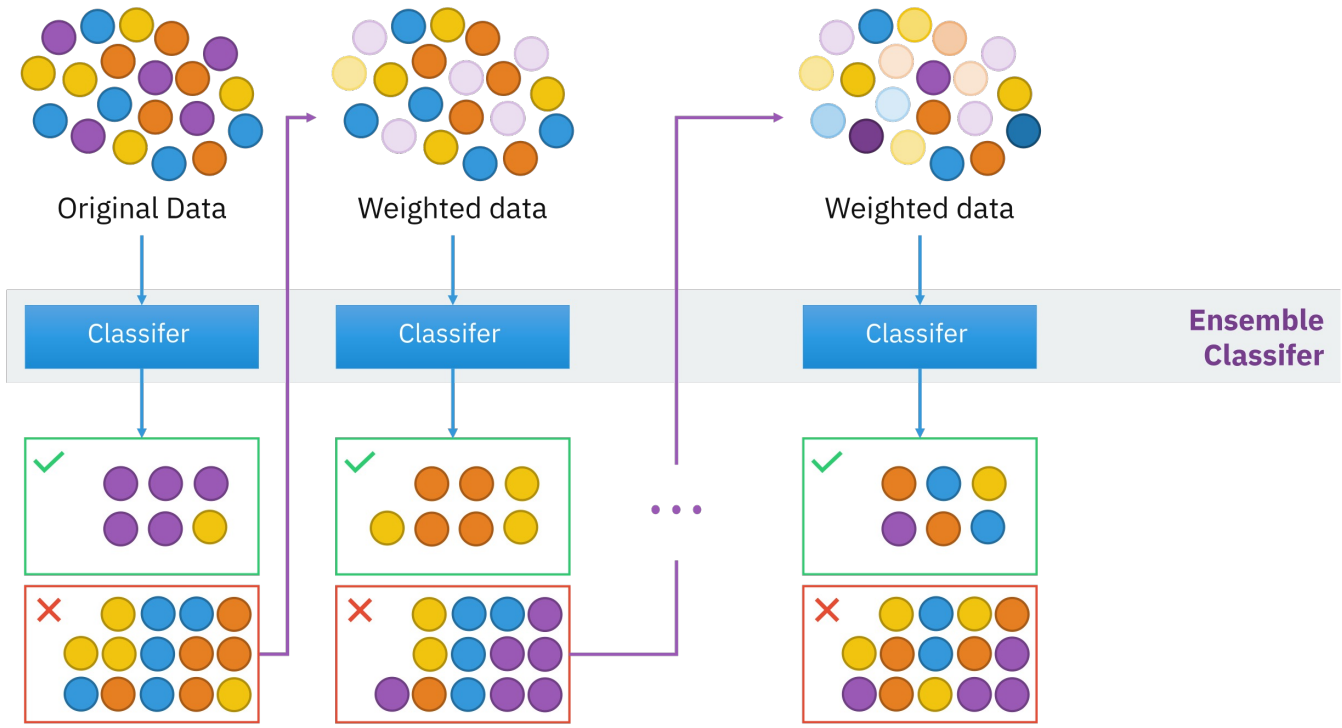


Stacking-Based Method



Boosting Technique

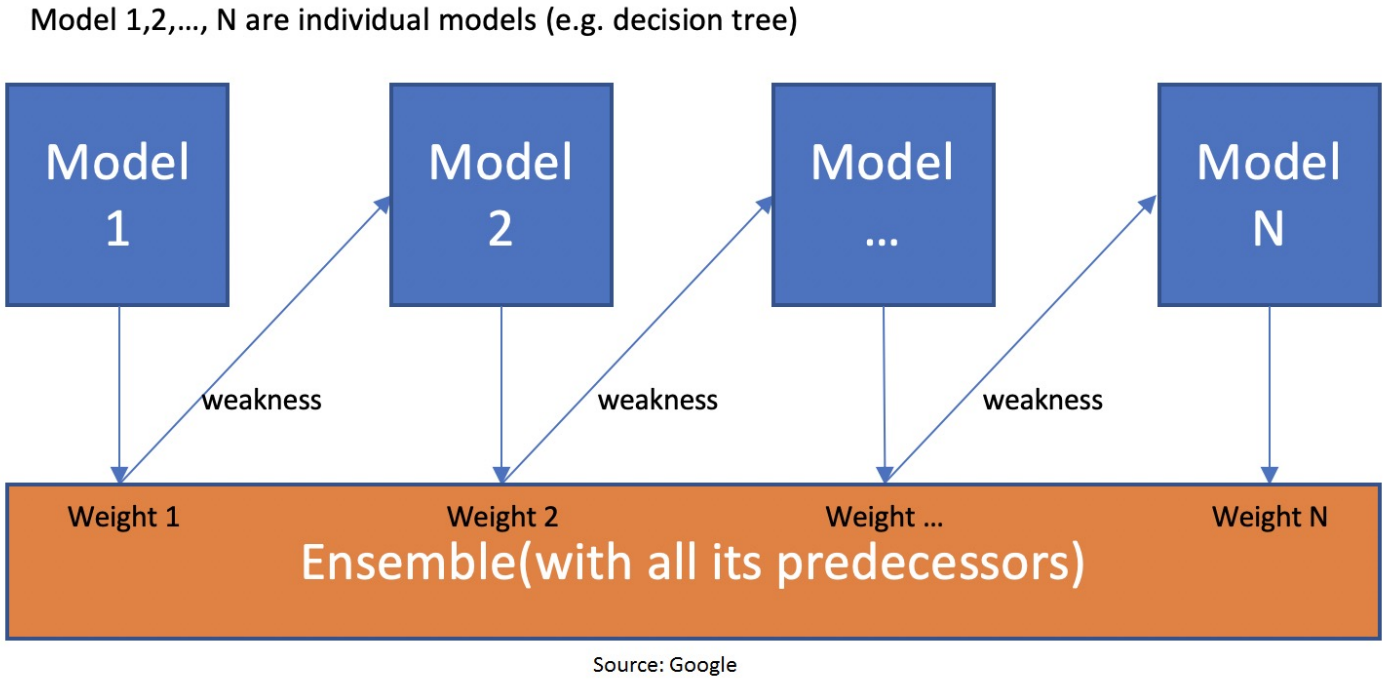
Boosting is an ensemble modelling, technique that attempts to build a strong classifier from the number of weak classifiers



https://en.wikipedia.org/wiki/Boosting_%28machine_learning%29#/media/File:Ensemble_Boosting.svg

Boosting Technique

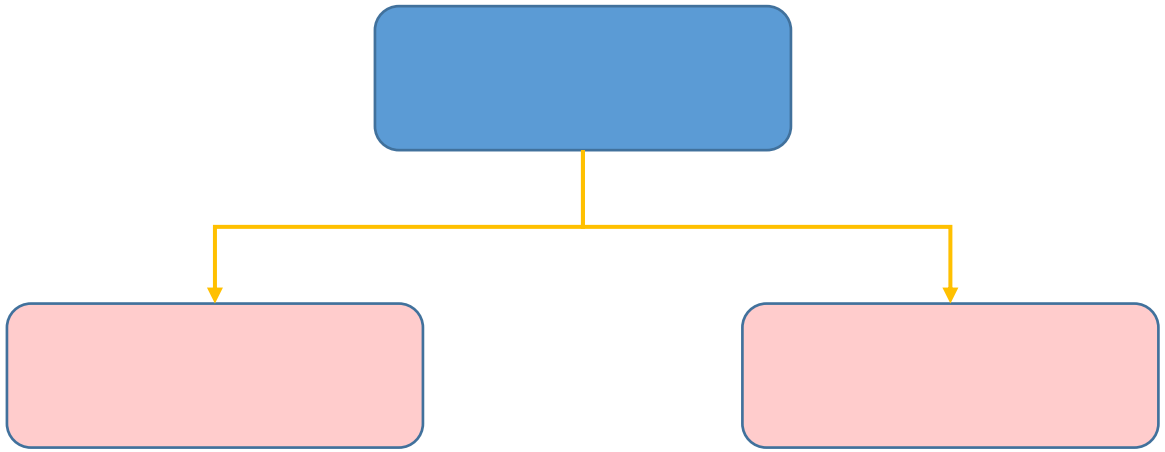
Boosting is an ensemble modelling, technique that attempts to build a strong classifier from the number of weak classifiers



Source: Google

https://en.wikipedia.org/wiki/Boosting_%28machine_learning%29#/media/File:Ensemble_Boosting.svg

Stump Definition

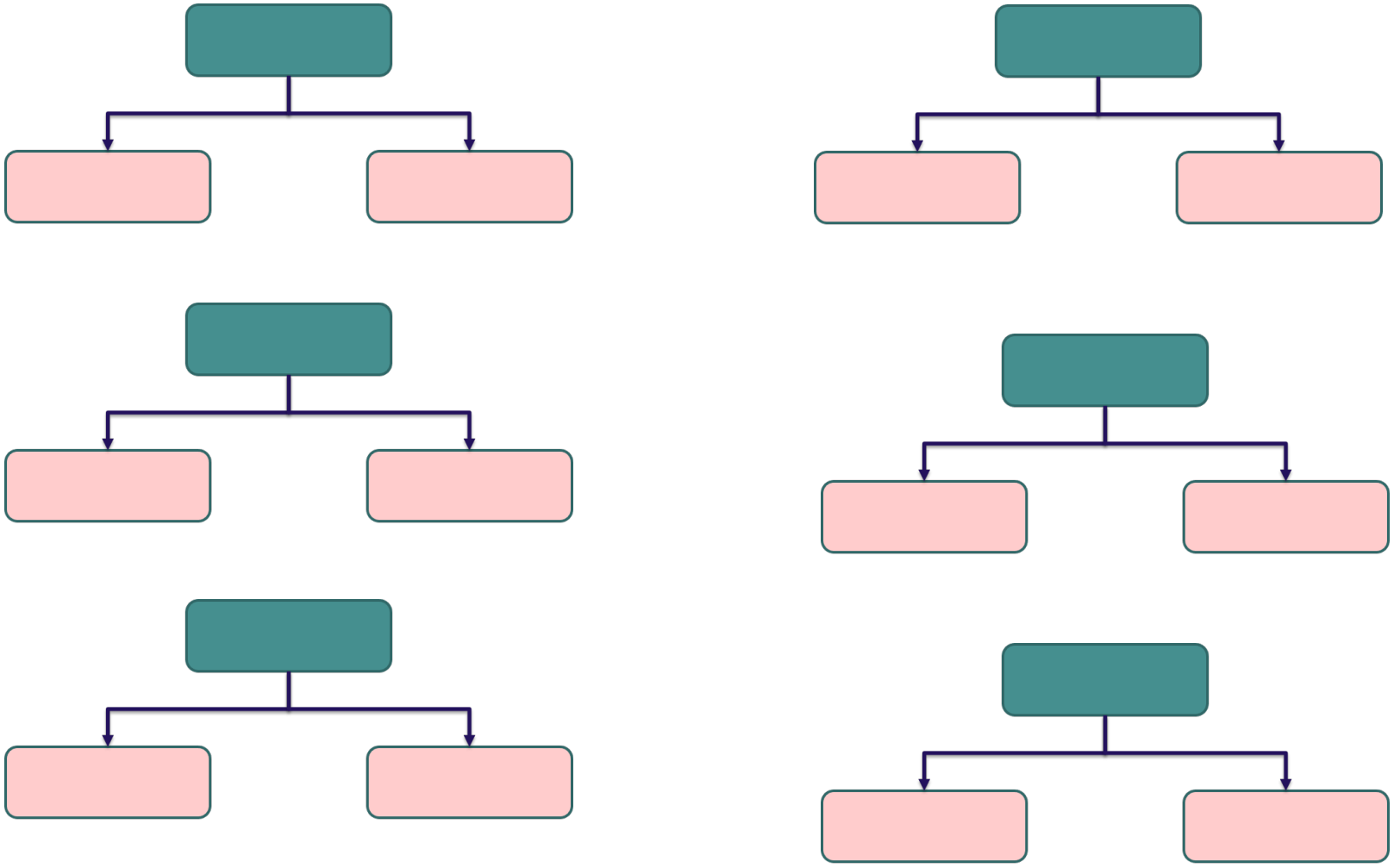


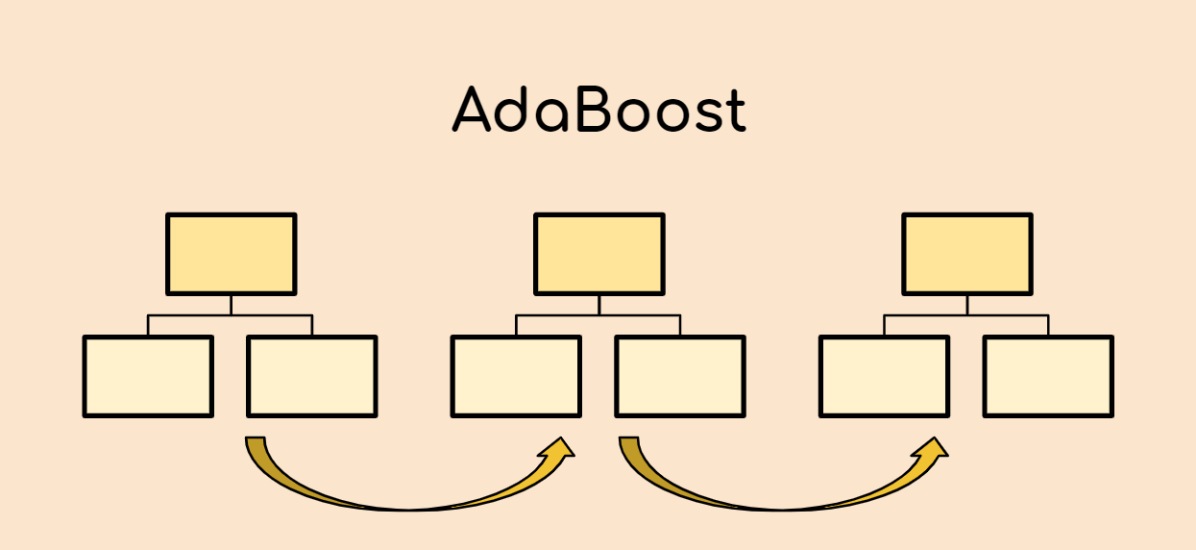
a node with two leaves and this is known as Stump

Outline

- Boosting Technique
- AdaBoost and Its Application
- Gradient Boost and Its Application

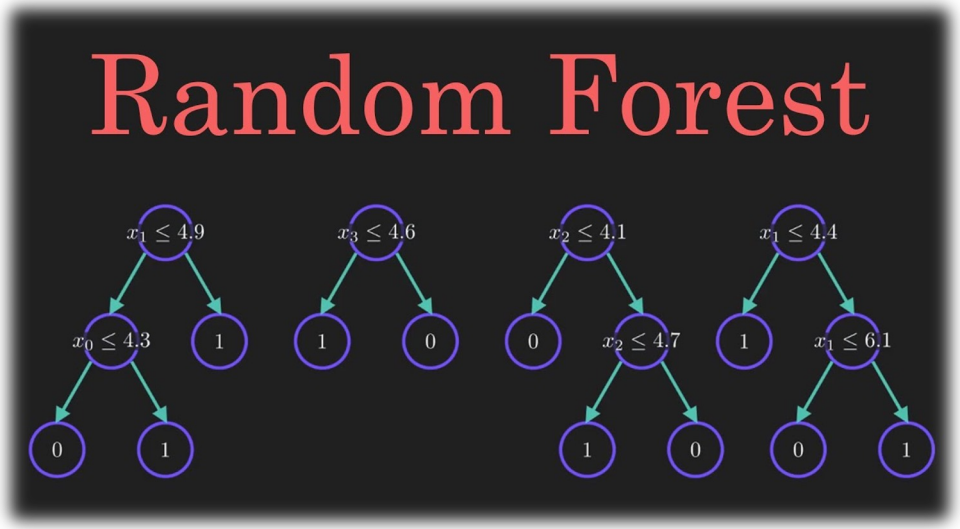
AdaBoost: Forest of Stump




AdaBoost


Make A
DiffeRéNcE


VS

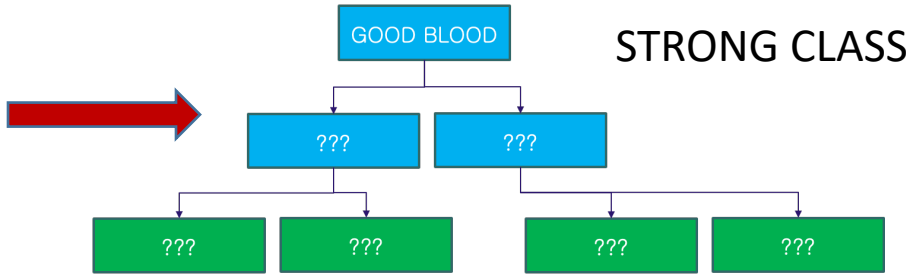

Random Forest
$x_1 \leq 4.9$
$x_0 \leq 4.3$
$x_3 \leq 4.6$
$x_2 \leq 4.1$
$x_2 \leq 4.7$
$x_1 \leq 4.4$
$x_1 \leq 6.1$

Sample Dataset

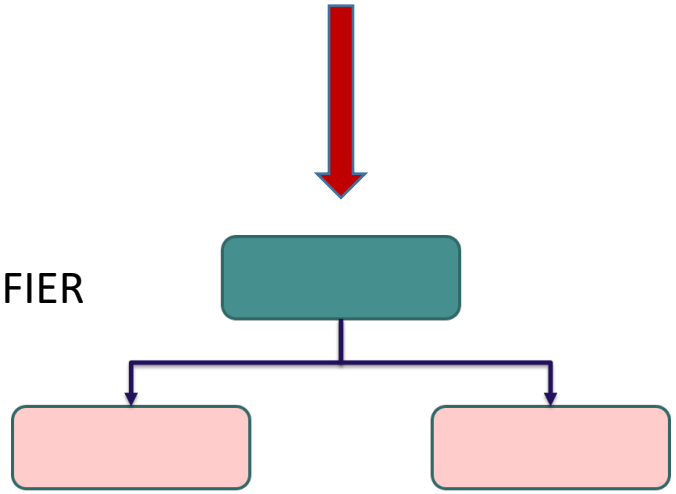
CHEST PAIN	GOOD BLOOD CIRCULATION	BLOCKED ARTERIES	WEIGHT	HEART DISEASE
NO	NO	NO	125	NO
YES	YES	YES	180	YES
YES	YES	NO	210	NO
YES	NO	YES	167	YES

Sample Dataset

CHEST PAIN	GOOD BLOOD CIRCULATION	BLOCKED ARTERIES	WEIGHT	HEART DISEASE
NO	NO	NO	125	NO
YES	YES	YES	180	YES
YES	YES	NO	210	NO
YES	NO	YES	167	YES



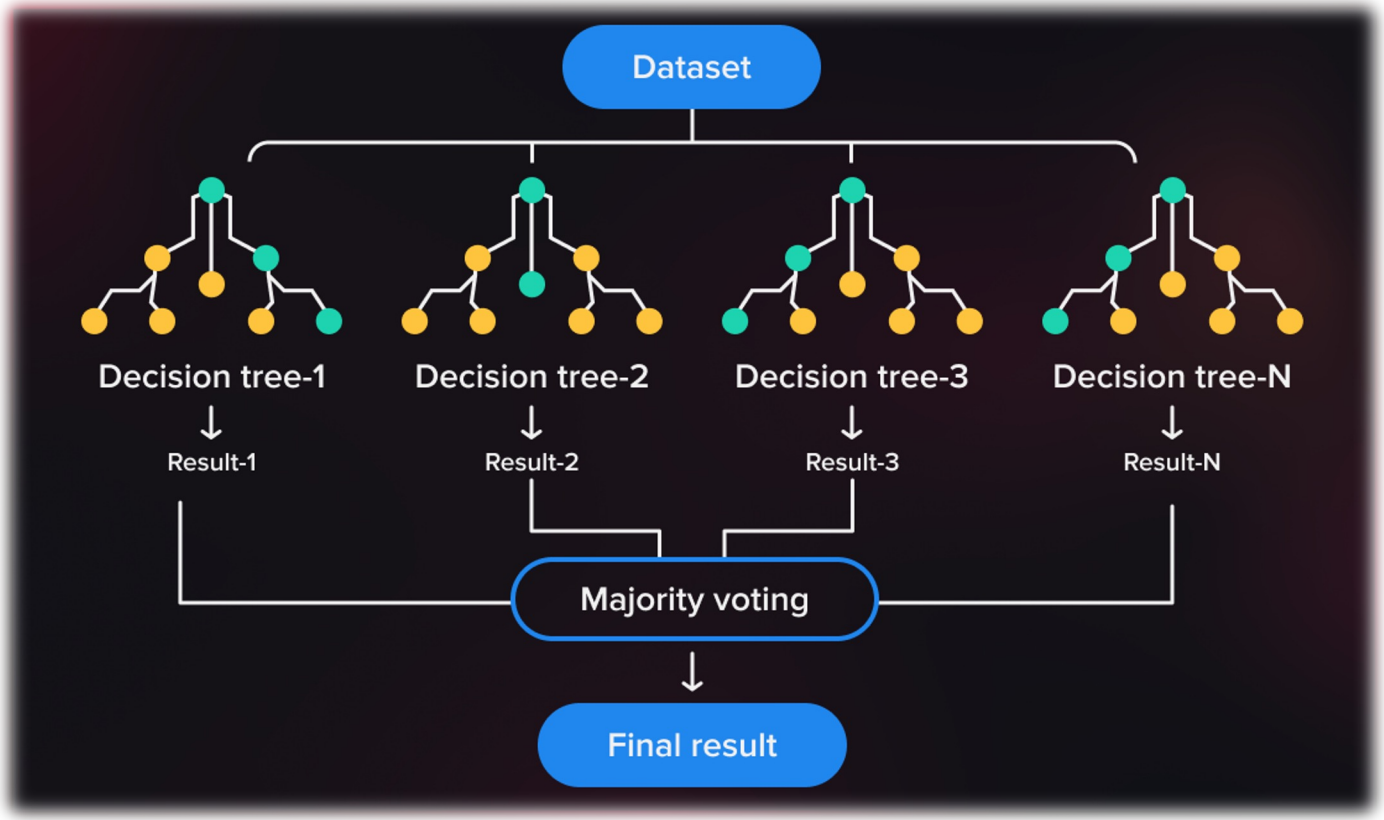
STRONG CLASSIFIER



WEAK CLASSIFIER

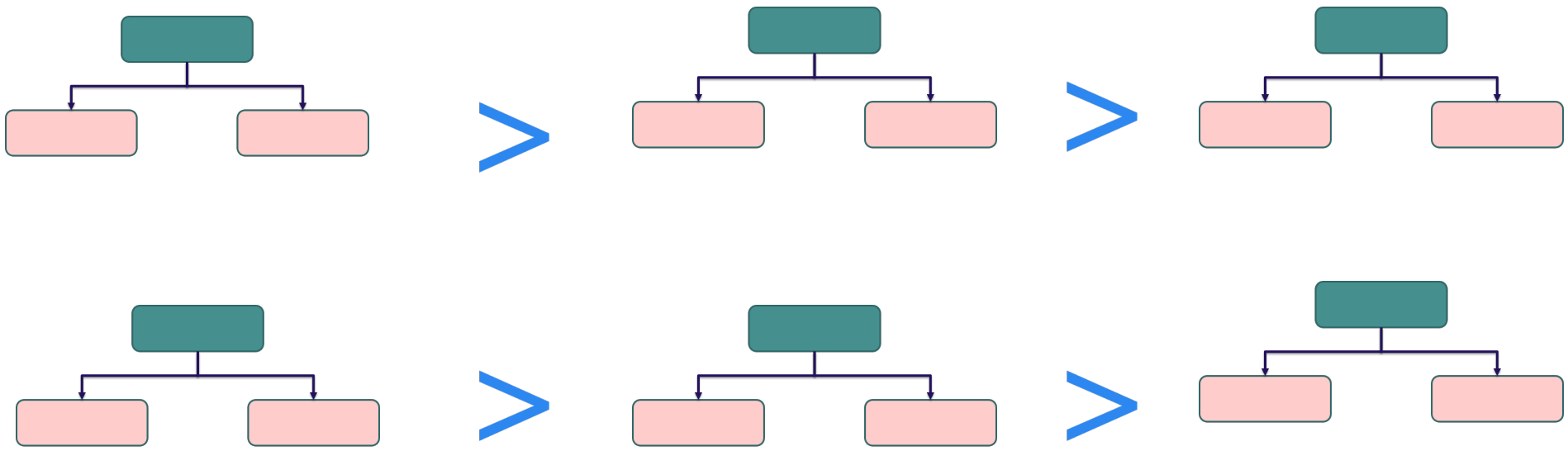
RANDOM FOREST

- Each tree in the random forest has equal votes(weights) on the final decision



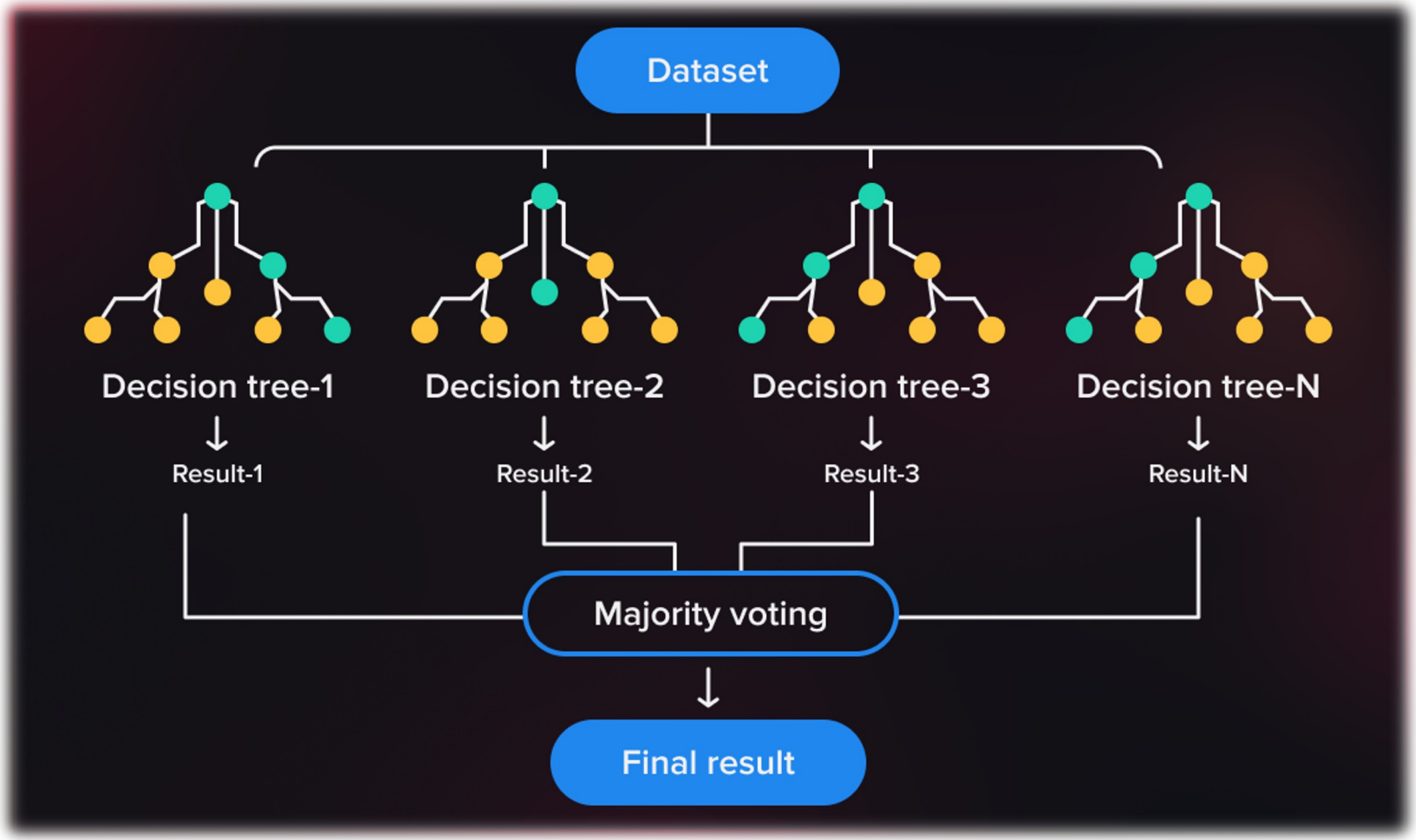
AdaBoost: FOREST OF STUMP

- Stump are not equally weighted in the final decision.
- Stump that create more error will have less contribution in the final decision

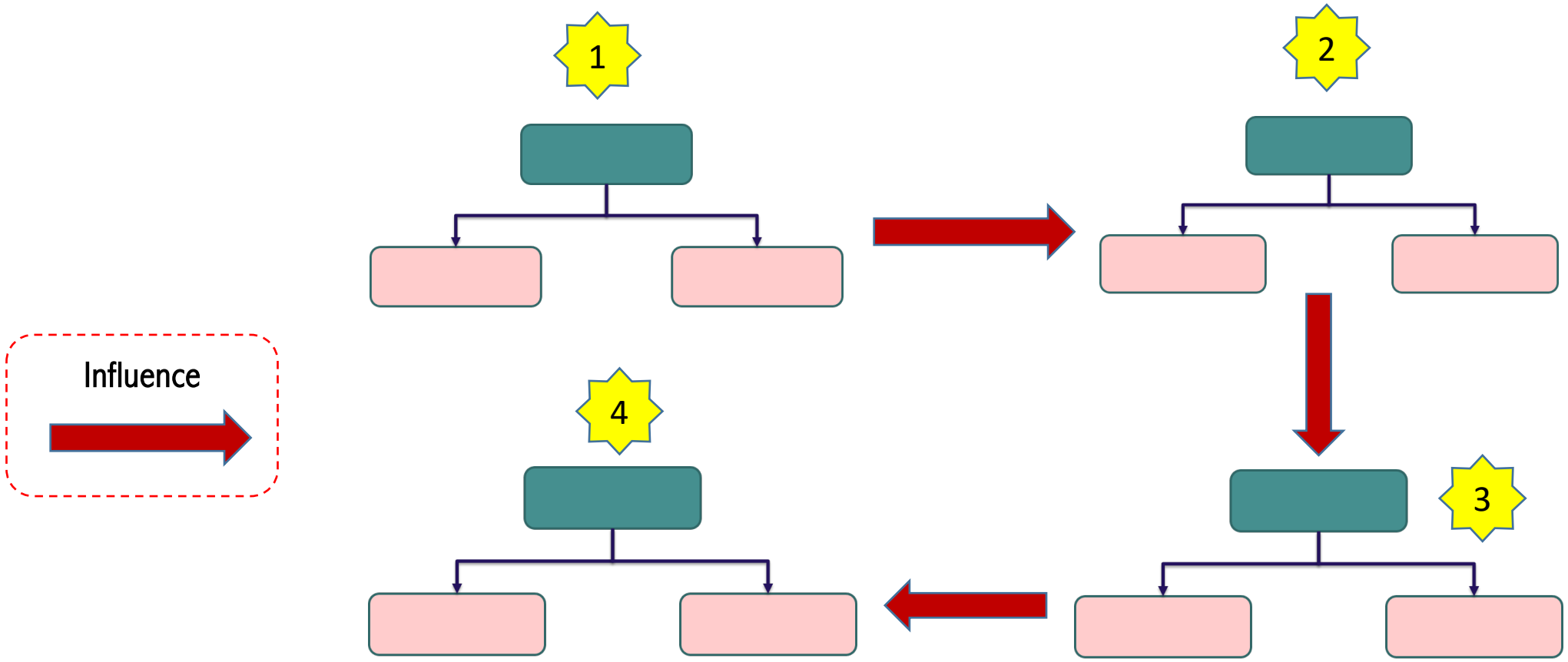


Random Forest

Tree are indepently created



AdaBoost: FOREST OF STUMP



Differences Between RF and AdaBoost

Weak Learners is a ___
AdaBoost combines a lot of___

Stumps have various ___ to the final result

Each stump is created by considering ___

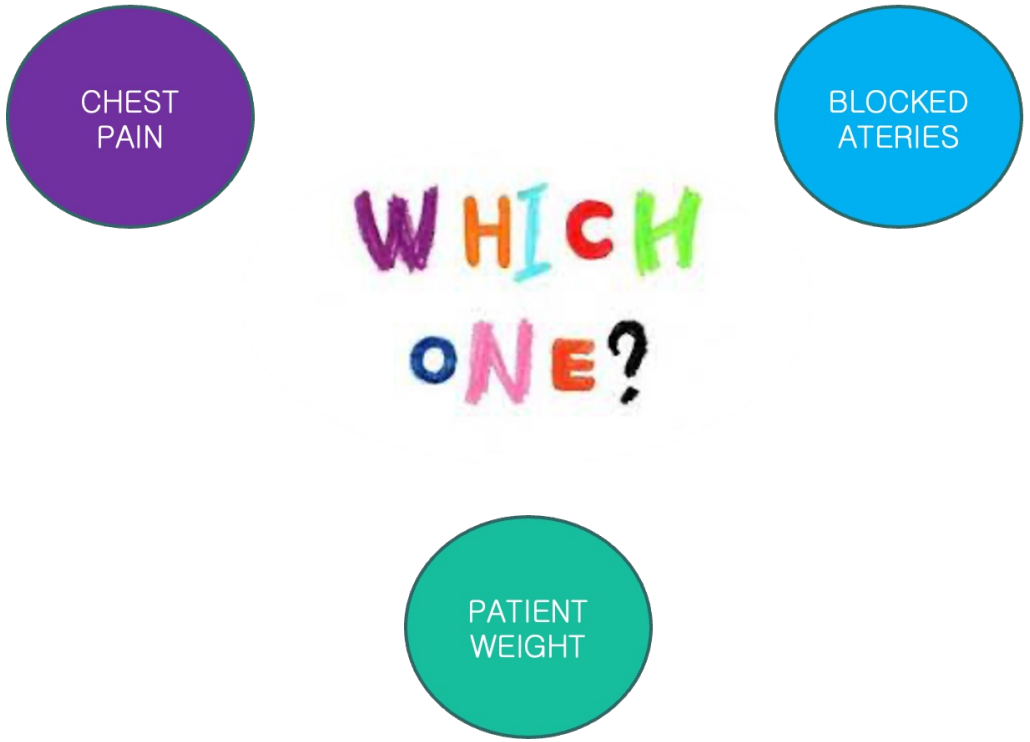
Heart Disease Dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease
Yes	Yes	205	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
No	Yes	156	No
No	Yes	125	No
Yes	No	168	No
Yes	Yes	172	No

Important of sample = Sample weight = 1 / number of samples = 1/8

1st Stump in the Tree

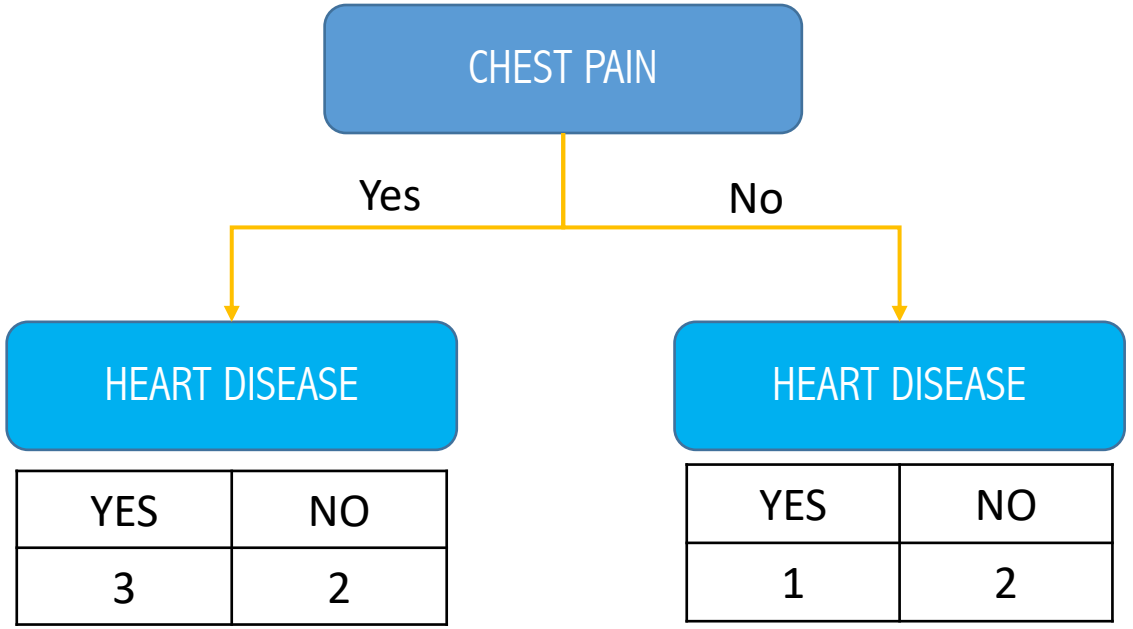
Chest Pain	Blocked Arteries	Patient Weight	Heart Diease
Yes	Yes	205	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
No	Yes	156	No
No	Yes	125	No
Yes	No	168	No
Yes	Yes	172	No



Compute Gini Index For Chest Pain

Chest Pain	Heart Diease	Sample Weight
Yes	Yes	1/8
No	Yes	1/8
Yes	Yes	1/8
Yes	Yes	1/8
No	No	1/8
No	No	1/8
Yes	No	1/8
Yes	No	1/8

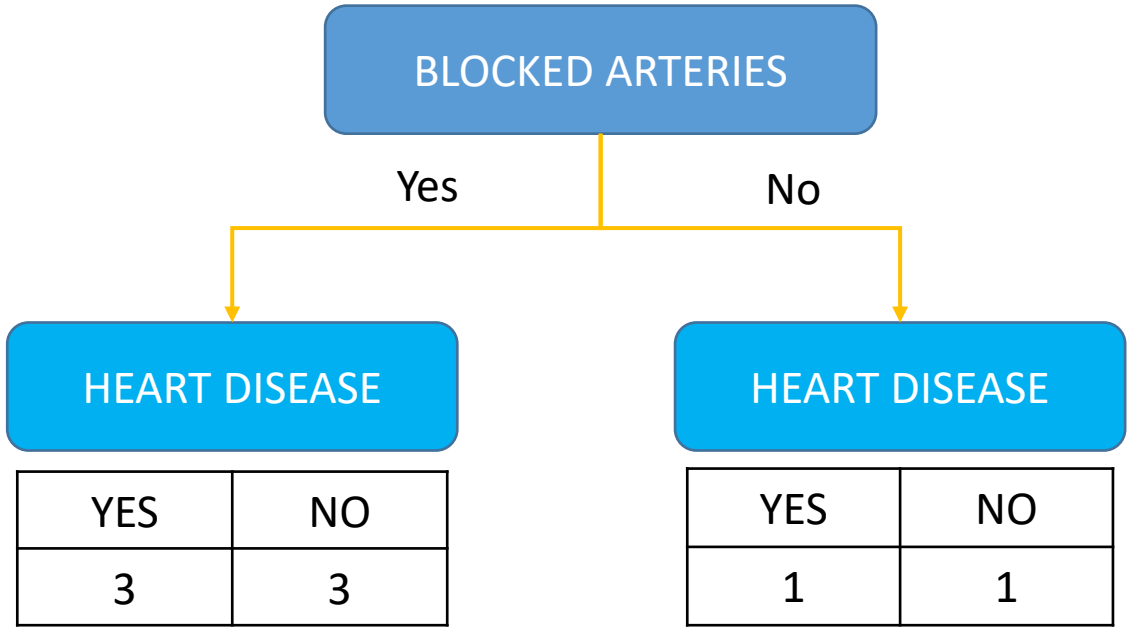
Gini index = 5/8 * (1 – (3/5)^2 – (2/5)^2) + 3/8* (1 - (1/3)^2 - (2/3)^2) = **0.57**



GINI INDEX FOR BLOCKED ARTERIES

Blocked Arteries	Heart Diease	Sample Weight
Yes	Yes	1/8
Yes	Yes	1/8
No	Yes	1/8
Yes	Yes	1/8
Yes	No	1/8
Yes	No	1/8
No	No	1/8
Yes	No	1/8

Gini index = 6/8 * (1 – (3/6)^2 – (3/6)^2) + 2/8* (1 - (1/2)^2 - (1/2)^2) = **0.5**



YES	NO
3	3

YES	NO
1	1

Gini Index for Heart Disease

Patient Weight	Heart Diease	Sample Weight
205	Yes	1/8
180	Yes	1/8
210	Yes	1/8
167	Yes	1/8
156	No	1/8
125	No	1/8
168	No	1/8
172	No	1/8

Gini index = =4/8 * (1-(1/4)^2 - (3/4)^2) + 4/8* (1-(1/4)^2 -(3/4)^2) = 0.375

192.5

195

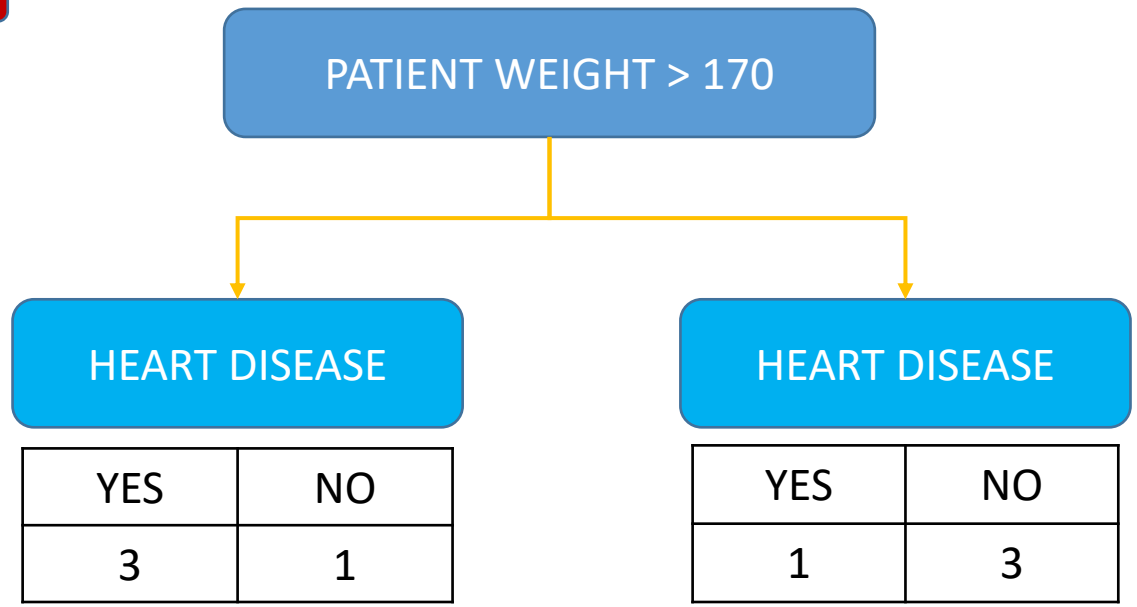
188.5

161.5

140.5

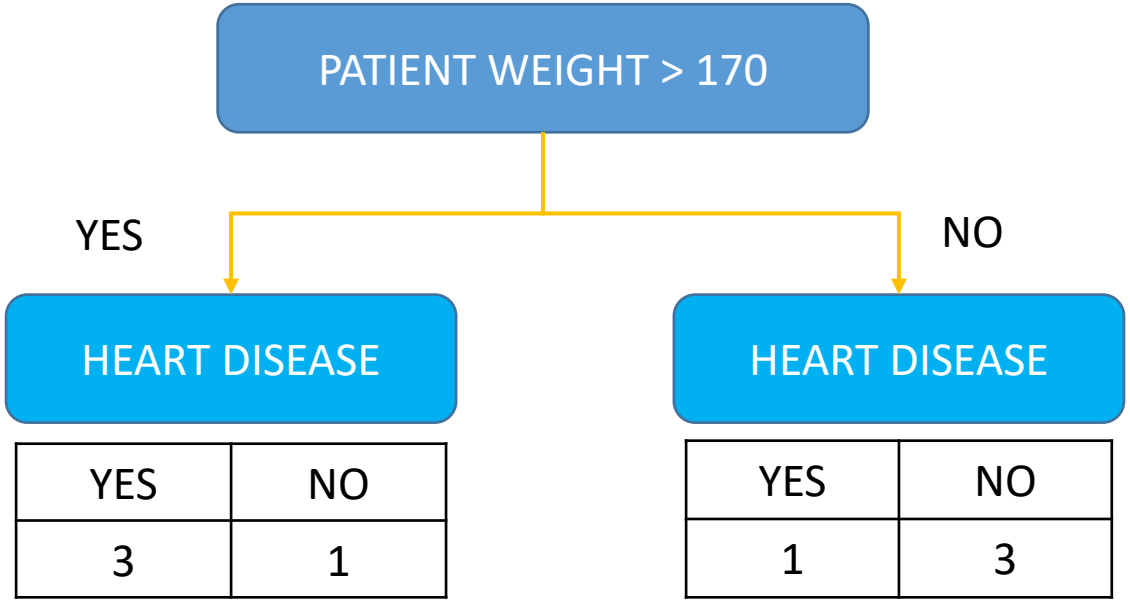
146.5

170



Amount of Say

How was this stump contribute to the final decision (classification)?



YES	NO
3	1

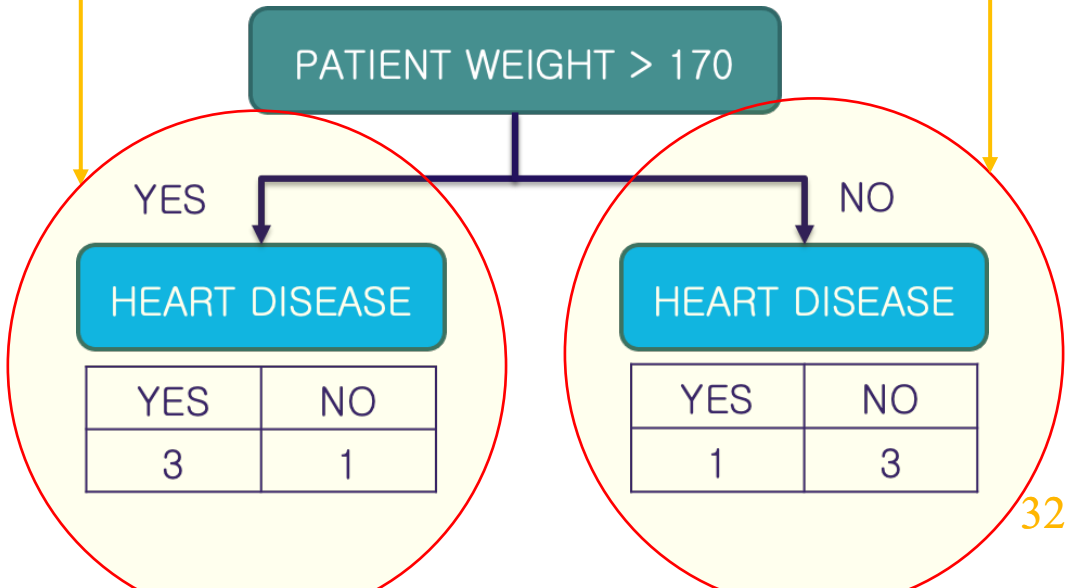
YES	NO
1	3

$$\text{Amount of Say} = \frac{1}{2}\log\left(\frac{1 - \text{Total Error}}{\text{Total Error}}\right)$$

Amount of Say: Patient Weight

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Sample Weight
Yes	Yes	205	Yes	1/8
No	Yes	180	Yes	1/8
Yes	No	210	Yes	1/8
Yes	Yes	167	Yes	1/8
No	Yes	156	No	1/8
No	Yes	125	No	1/8
Yes	No	168	No	1/8
Yes	Yes	172	No	1/8

- Total error is equal to the sum of the weights of the incorrect classified
- Amount of say = 1/2*log((1-2/8) / (2/8)) = 0.55



Amount of Say: Weight of The Tree

Probability Vs. Odds

Your friend went fishing 10 times a month

- Caught a fish 4 times
- Failed to catch 6 times

What is the *probability* and *odds* of getting a Fish for lunch?

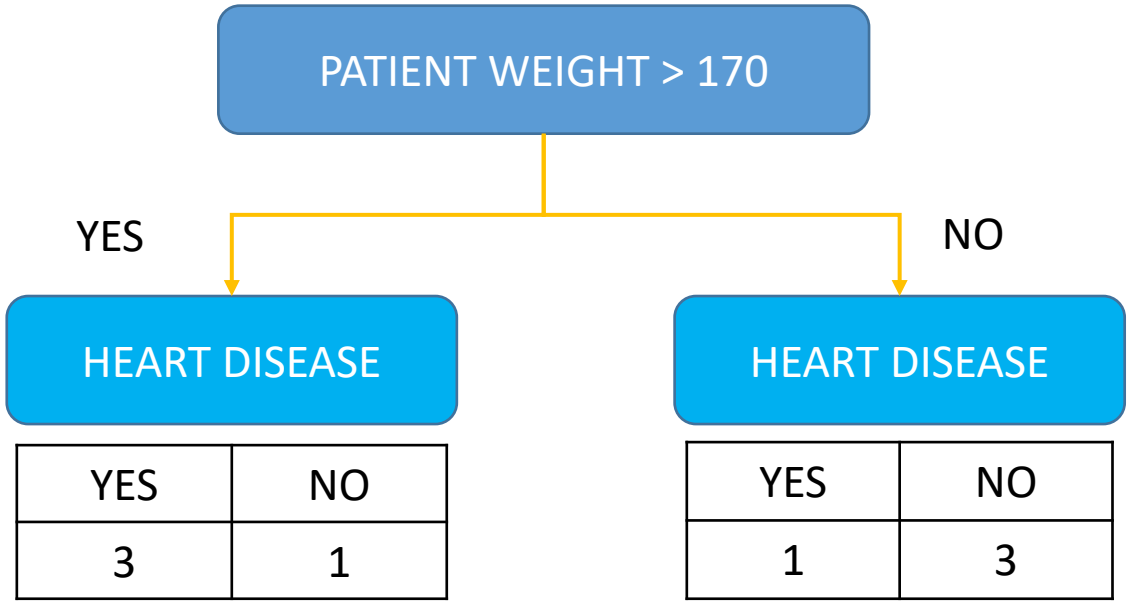
$$\text{Probability} = \frac{\text{Chance for catching fish}}{\text{Total chances}} = \frac{4}{10} = 0.4$$

$$\text{Odds} = \frac{\text{Chance for catching fish}}{\text{Chance for not catching fish}} = \frac{4}{6} = 0.67$$

$$\text{Odds} = \frac{\text{Probability of catching fish}}{\text{Probability of not catching fish}} = \frac{4/10}{6/10} = 0.67$$

$$\text{Odds} = \frac{1 - \text{Probability of not catching fish}}{\text{Probability of not catching fish}} = \frac{4/10}{6/10} = 0.67$$

Amount of Say: Weight of The Tree



YES	NO
3	1

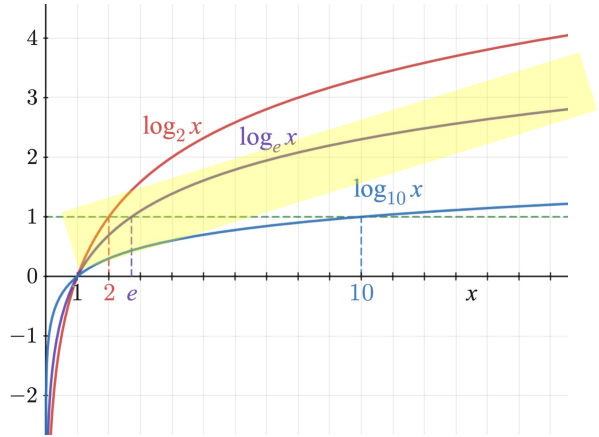
YES	NO
1	3

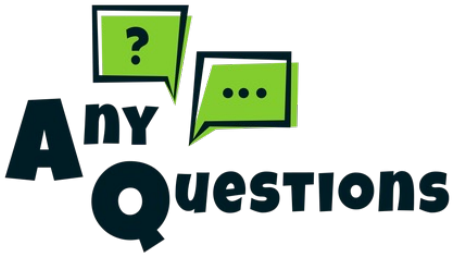
$$\text{Amount of Say} = \frac{1}{2}\log\left(\frac{1 - \text{Total Error}}{\text{Total Error}}\right)$$

$$\text{Odds} = \frac{1 - \text{Probability of not catching fish}}{\text{Probability of not catching fish}} = \frac{4/10}{6/10} = 0.67$$

$$\text{Odds} = \frac{1 - \text{Probability of incorrect prediction}}{\text{Probability of incorrect prediction}}$$

Amount of says = Odds or Amount of says = ½ x log(Odds)

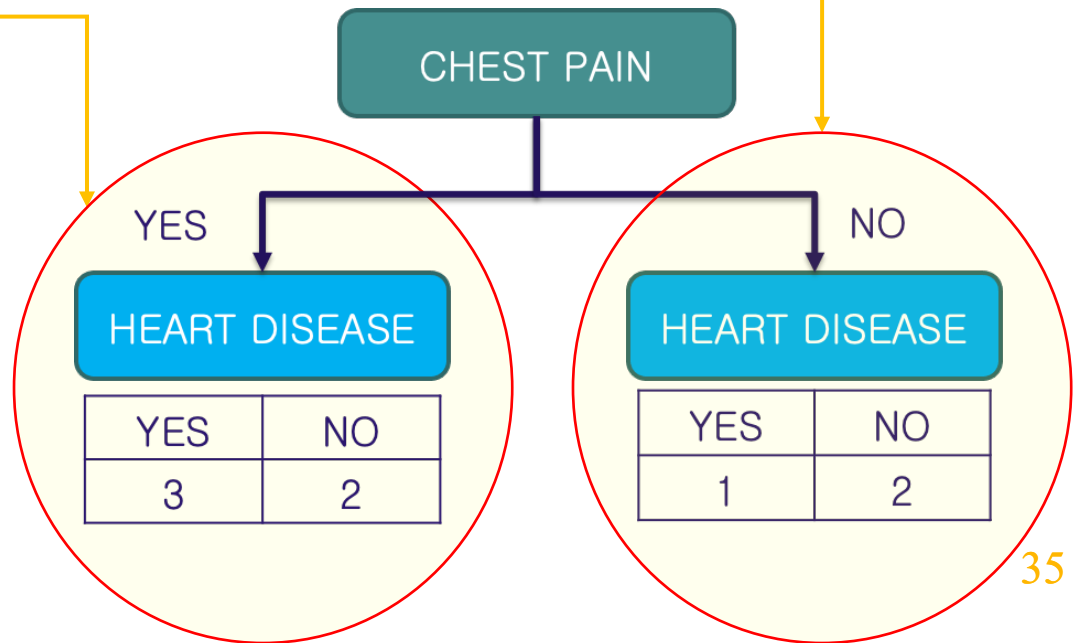




Why?

Amount of say: Chest Pain

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Sample Weight
Yes	Yes	205	Yes	1/8
No	Yes	180	Yes	1/8
Yes	No	210	Yes	1/8
Yes	Yes	167	Yes	1/8
No	Yes	156	No	1/8
No	Yes	125	No	1/8
Yes	No	168	No	1/8
Yes	Yes	172	No	1/8

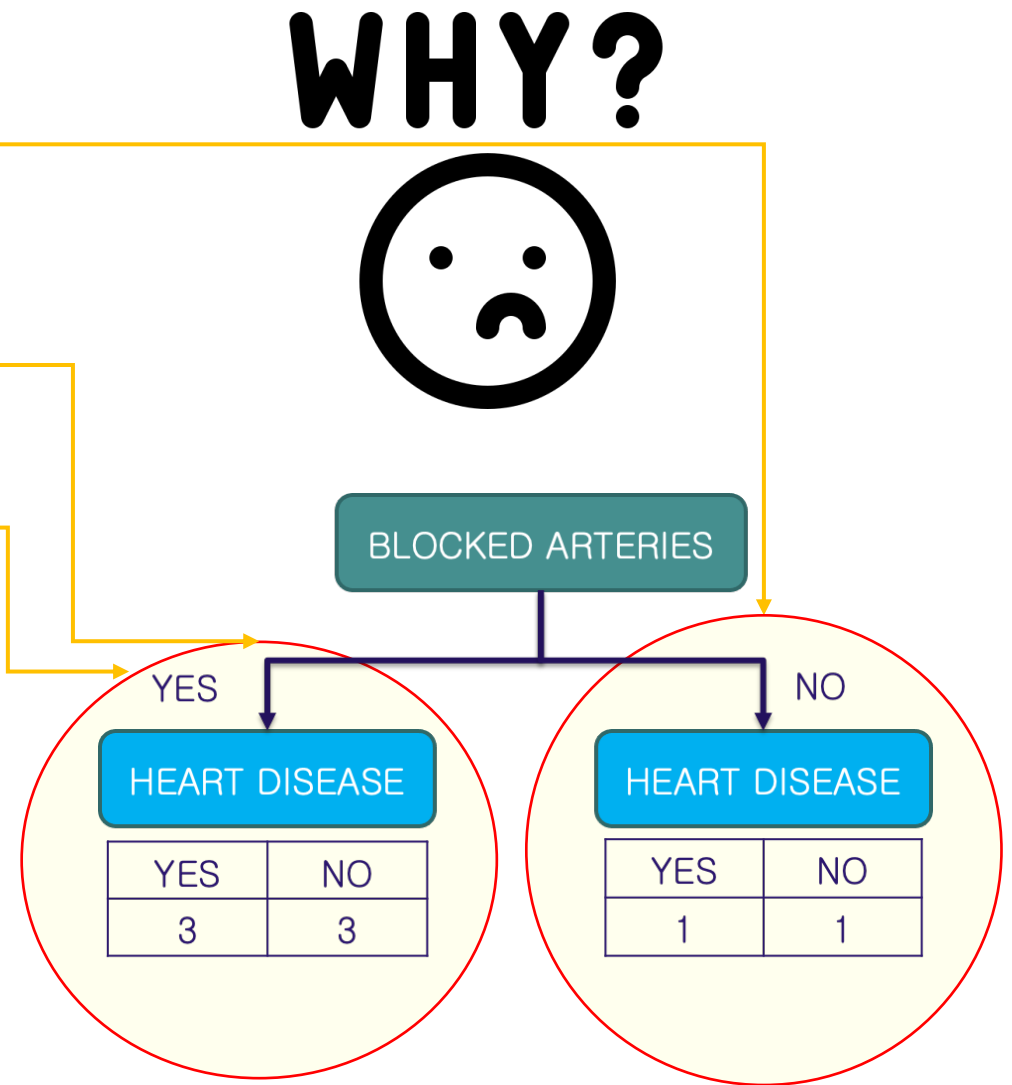


- Total error is equal to the sum of the weights of the incorrect classified
- Amount of say = 1/2*log((1-3/8) / (3/8)) = 0.25

$$\log(\text{Odds}) = \log\left(\frac{1 - \text{Probability of incorrect prediction}}{\text{Probability of incorrect prediction}}\right)$$

Amout of Say: Blocked Arteries

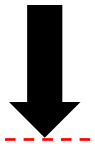
Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Sample Weight
Yes	Yes	205	Yes	1/8
No	Yes	180	Yes	1/8
Yes	No	210	Yes	1/8
Yes	Yes	167	Yes	1/8
No	Yes	156	No	1/8
No	Yes	125	No	1/8
Yes	No	168	No	1/8
Yes	Yes	172	No	1/8



- Total error is equal to the sum of the weights of the incorrect classified
- Amount of say = 1/2*log((1-4/8) / (4/8)) = 0

Assumptions

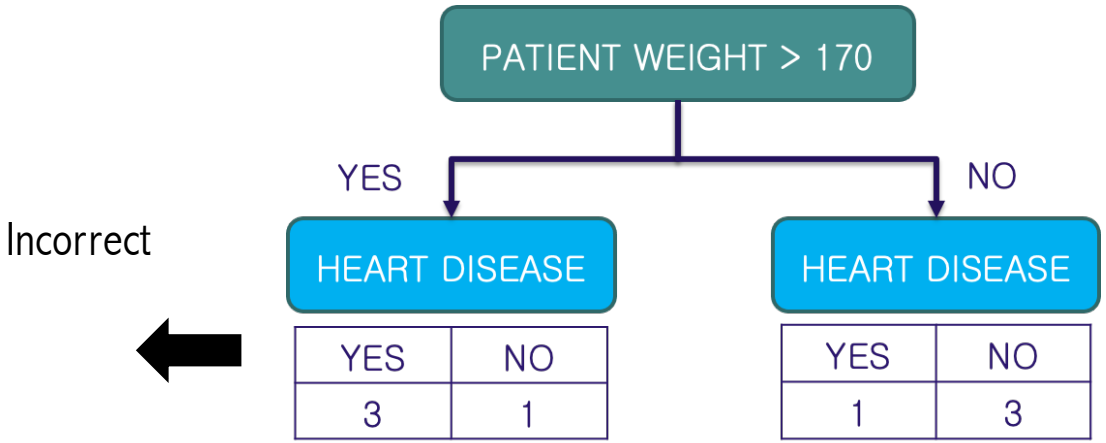
Known: weight cho các sample dự đoán sai được sử dụng để tính "Amount of Say" cho từng stump hiện tại.



Unknown: Tiếp theo, chúng ta cần làm thế nào để sử dụng thông tin các weight của sample dự đoán sai này để xây dựng stump tiếp và khắc phục các dự đoán sai này

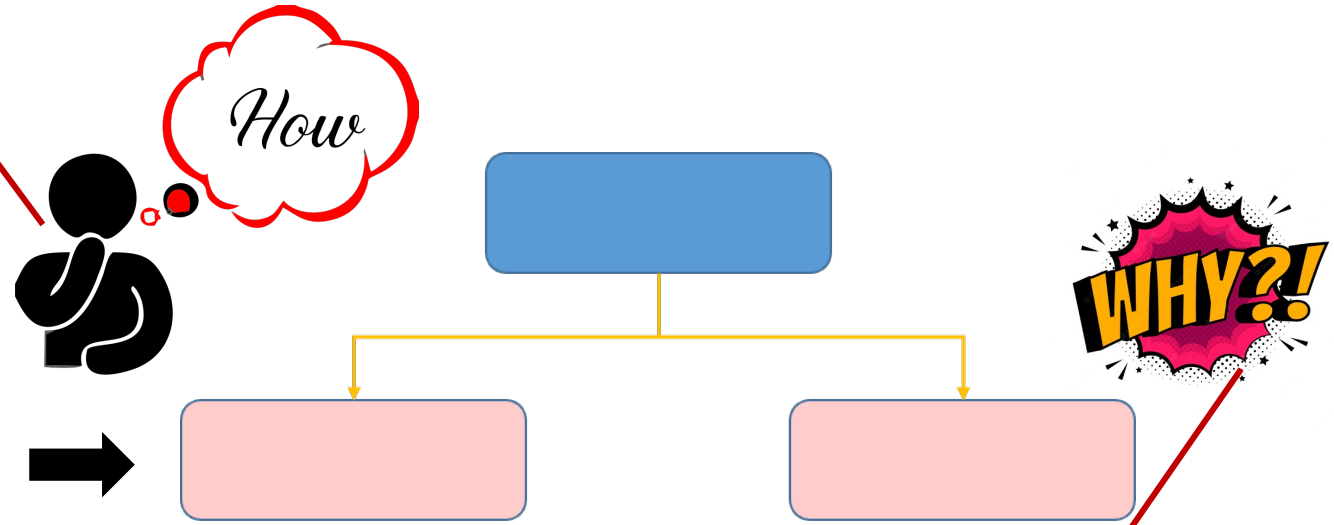
Idea: Improved Bootstrapped Dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease
Yes	Yes	205	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
No	Yes	156	No
No	Yes	125	No
Yes	No	168	No
Yes	Yes	172	No



Create new dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Disease
Yes	Yes	205	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
Yes	Yes	167	Yes
Yes	Yes	167	Yes
Yes	Yes	172	No
Yes	Yes	172	No

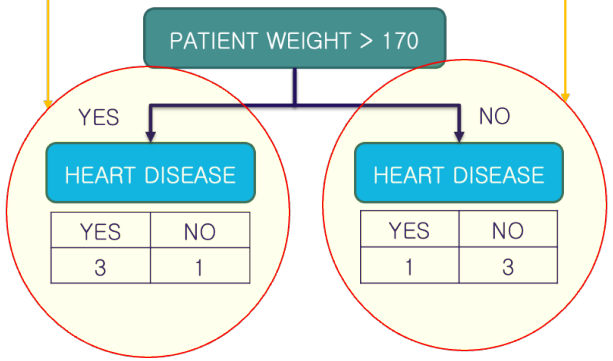


How to build next Stump

Increase the sample weights of samples that were incorrectly classified and decrease sample weights of samples that were correctly classified. Label {-1, 1}

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Sample Weight
Yes	Yes	205	Yes	1/8
No	Yes	180	Yes	1/8
Yes	No	210	Yes	1/8
Yes	Yes	167	Yes	1/8
No	Yes	156	No	1/8
No	Yes	125	No	1/8
Yes	No	168	No	1/8
Yes	Yes	172	No	1/8

- Total error is equal to the sum of the weights of the incorrect classified
- Amount of say = 1/2*log((1-2/8) / (2/8)) = 0.55



$$\text{New Sample Weight} = \text{sample weight} \times e^{\text{amount of say}}$$

New sample weight = 1/8 * e^{0.55} = 0.22

```
def update_weights_formular1(w_i, alpha, y, y_pred):
    result = w_i * np.exp(-alpha * y * y_pred)
    w_norm = result / np.sum(result)
    return w_norm
```

How to build next Stump

Increase the sample weights of samples that were incorrectly classified and decrease the sample weights of samples that were correctly classified. Label {-1, 1}

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease
Yes	Yes	205	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
No	Yes	156	No
No	Yes	125	No
Yes	No	168	No
Yes	Yes	172	No

$$\text{New Sample Weight} = \text{sample weight} \times e^{\text{-amount of say}}$$

New sample weight = 1/8 * e^{-0.55} = 0.07

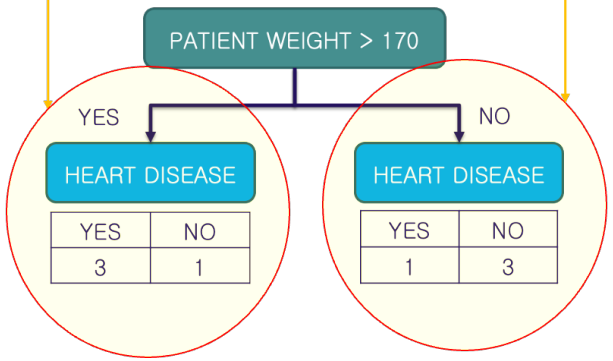
```
def update_weights_formular1(w_i, alpha, y, y_pred):
    result = w_i * np.exp(-alpha * y * y_pred)
    w_norm = result / np.sum(result)
    return w_norm
```

How to build next Stump

Increase the sample weights of samples that were incorrectly classified and keep the sample weights of samples that were correctly classified. Label {0, 1}

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Sample Weight
Yes	Yes	205	Yes	1/8
No	Yes	180	Yes	1/8
Yes	No	210	Yes	1/8
Yes	Yes	167	Yes	1/8
No	Yes	156	No	1/8
No	Yes	125	No	1/8
Yes	No	168	No	1/8
Yes	Yes	172	No	1/8

- Total error is equal to the sum of the weights of the incorrect classified
- Amount of say = 1/2*log((1-2/8) / (2/8)) = 0.55



$$w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))],$$

New sample weight = 1/8 * e^{0.55*1} = 0.22

```
def update_weights_formular2(w_i, alpha, y, y_pred):
    result = w_i * np.exp(alpha * (
        np.not_equal(y, y_pred)).astype(int))
    w_norm = result / np.sum(result)
    return w_norm
```

How to build next Stump

Increase the sample weights of samples that were incorrectly classified and keep the sample weights of samples that were correctly classified. Label {0, 1}

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease
Yes	Yes	205	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
No	Yes	156	No
No	Yes	125	No
Yes	No	168	No
Yes	Yes	172	No

$$w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))],$$

New sample weight = 1/8 * e^{0.55*0} = 0.125

```
def update_weights_formular2(w_i, alpha, y, y_pred):
    result = w_i * np.exp(alpha * (
        np.not_equal(y, y_pred)).astype(int))
    w_norm = result / np.sum(result)
    return w_norm
```

New Sample Weight

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Sample Weight	New Weight	Normal Weight
Yes	Yes	205	Yes	1/8	0.07	0.08
No	Yes	180	Yes	1/8	0.07	0.08
Yes	No	210	Yes	1/8	0.07	0.08
Yes	Yes	167	Yes	1/8	0.22	0.25
No	Yes	156	No	1/8	0.07	0.08
No	Yes	125	No	1/8	0.07	0.08
Yes	No	168	No	1/8	0.07	0.08
Yes	Yes	172	No	1/8	0.22	0.25
Sum				~1.0	0.86	~1.0

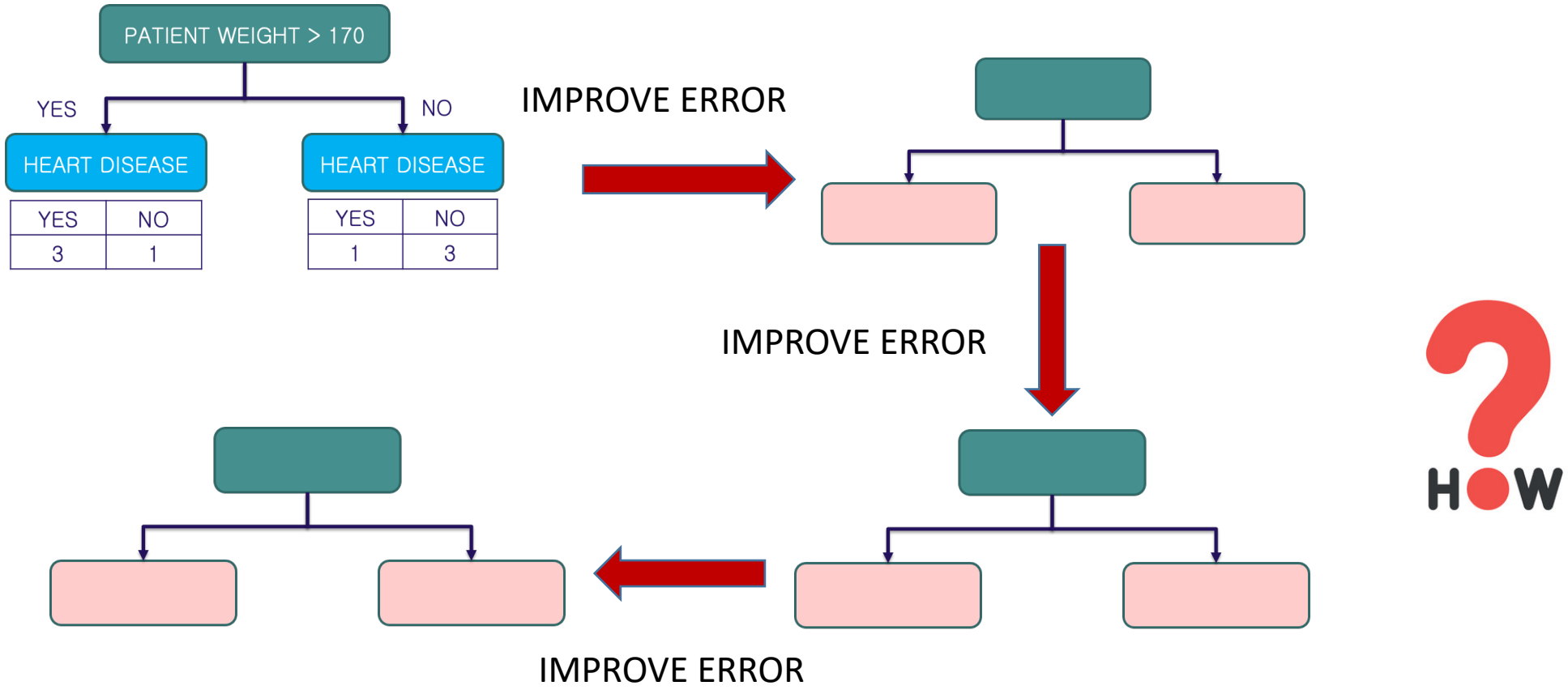
$$\text{New Sample Weight} = \text{sample weight} \times e^{\text{amount of say}}$$

Update

New Sample Weight

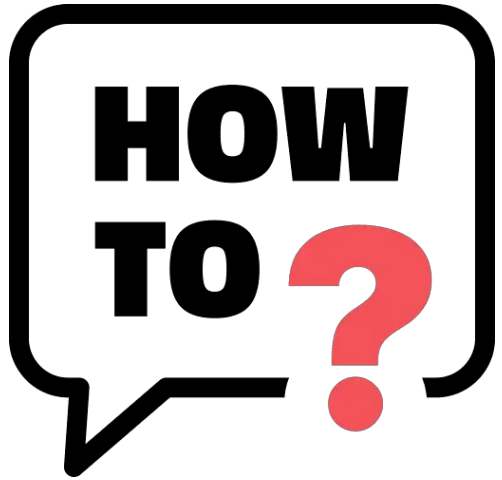
Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	New Weight
Yes	Yes	205	Yes	0.08
No	Yes	180	Yes	0.08
Yes	No	210	Yes	0.08
Yes	Yes	167	Yes	0.25
No	Yes	156	No	0.08
No	Yes	125	No	0.08
Yes	No	168	No	0.08
Yes	Yes	172	No	0.25
Sum				~1.0

AdaBoost: FOREST OF STUMPS



New Dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Normal Weight
Yes	Yes	205	Yes	0.08
No	Yes	180	Yes	0.08
Yes	No	210	Yes	0.08
Yes	Yes	167	Yes	0.25
No	Yes	156	No	0.08
No	Yes	125	No	0.08
Yes	No	168	No	0.08
Yes	Yes	172	No	0.25
Sum				~1.0



Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Normal Weight
Yes	Yes	205	Yes	0.08
No	Yes	180	Yes	0.08
Yes	No	210	Yes	0.08
Yes	Yes	167	Yes	0.25
No	Yes	156	No	0.08
No	Yes	125	No	0.08
Yes	No	168	No	0.08
Yes	Yes	172	No	0.25
Sum				~1.0

New Dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Normal Weight
Yes	Yes	205	Yes	0.08
No	Yes	180	Yes	0.08
Yes	No	210	Yes	0.08
Yes	Yes	167	Yes	0.25
No	Yes	156	No	0.08
No	Yes	125	No	0.08
Yes	No	168	No	0.08
Yes	Yes	172	No	0.25
Sum				~1.0

Random [0,1] to select samples

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Normal Weight
Yes	Yes	205	Yes	0.08
No	Yes	180	Yes	0.08
Yes	No	210	Yes	0.08
Yes	Yes	167	Yes	0.25
No	Yes	156	No	0.08
No	Yes	125	No	0.08
Yes	No	168	No	0.08
Yes	Yes	172	No	0.25
Sum				~1.0

New Dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Normal Weight	Range
Yes	Yes	205	Yes	0.08	[0-0.08]
No	Yes	180	Yes	0.08	(0.08-0.16]
Yes	No	210	Yes	0.08	(0.16-0.24]
Yes	Yes	167	Yes	0.25	(0.24-0.495]
No	Yes	156	No	0.08	(0.495-0.575]
No	Yes	125	No	0.08	(0.575-0.655]
Yes	No	168	No	0.08	(0.655-0.735]
Yes	Yes	172	No	0.25	(0.735-1.0]
Sum				~1.0	

New Dataset

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease	Normal Weight
Yes	Yes	205	Yes	0.08
No	Yes	180	Yes	0.08
Yes	No	210	Yes	0.08
Yes	Yes	167	Yes	0.25
No	Yes	156	No	0.08
No	Yes	125	No	0.08
Yes	No	168	No	0.08
Yes	Yes	172	No	0.25
Sum				~1.0

Old dataset

Random [0,1] to select samples

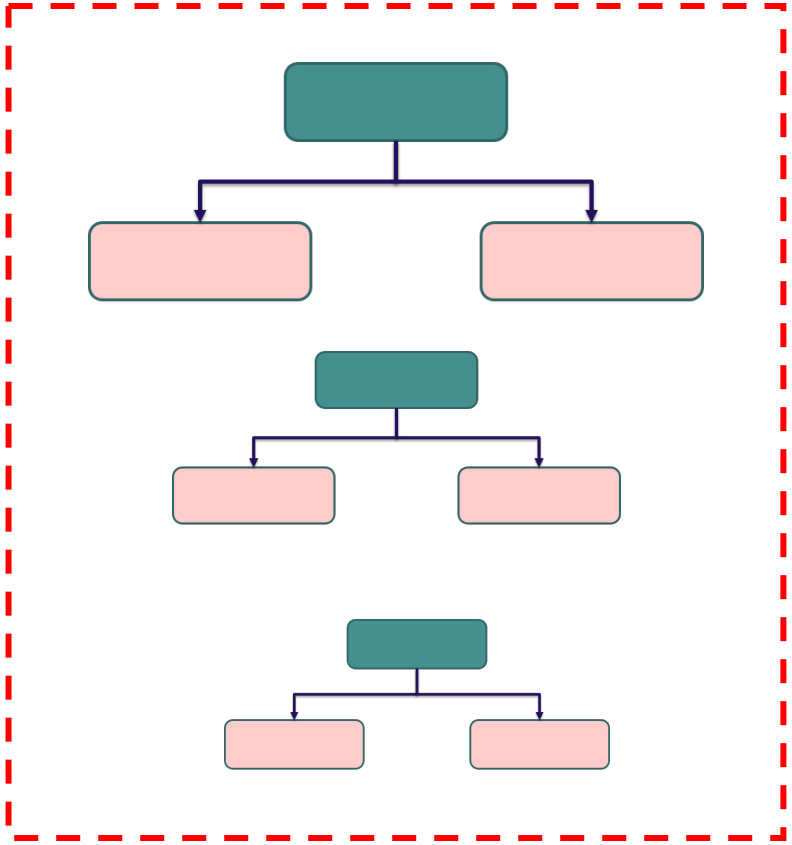
CONTINUE TO BUILD THE NEXT STUMP

Chest Pain	Blocked Arteries	Patient Weight	Heart Diease
Yes	Yes	167	Yes
No	Yes	180	Yes
Yes	No	210	Yes
Yes	Yes	167	Yes
No	Yes	156	No
No	Yes	125	No
Yes	Yes	172	No
Yes	Yes	172	No

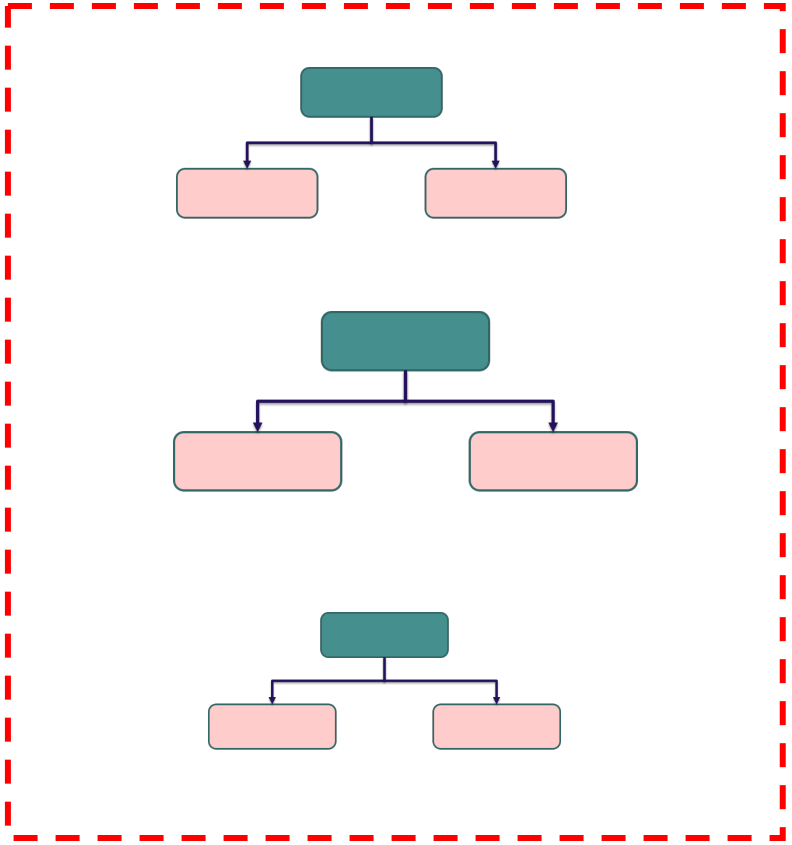
New dataset

Ý tưởng: các sample bị phân loại sai, sẽ được được nhiều hơn vào dataset mới

How to Classify The Final Result



These stumps for predicting heart disease



These stumps for predicting no heart disease

How to Classify The Final Result

Algorithm 1 *AdaBoost (Freund & Schapire 1997)*

1. *Initialize the observation weights* $w_i = 1/n,\ i = 1, 2, \ldots, n.$
2. *For* $m = 1$ *to* M*:*
   (a) *Fit a classifier* $T^{(m)}(\boldsymbol{x})$ *to the training data using weights* w_i.
   (b) *Compute*
   $$err^{(m)} = \sum_{i=1}^{n} w_i \mathbb{I}\left(c_i \neq T^{(m)}(\boldsymbol{x}_i)\right) / \sum_{i=1}^{n} w_i.$$
   (c) *Compute*
   $$\alpha^{(m)} = \log \frac{1 - err^{(m)}}{err^{(m)}}.$$
   (d) *Set*
   $$w_i \leftarrow w_i \cdot \exp\left(\alpha^{(m)} \cdot \mathbb{I}\left(c_i \neq T^{(m)}(\boldsymbol{x}_i)\right)\right),\ i = 1, 2, \ldots, n.$$
   (e) *Re-normalize* w_i.
3. *Output*
   $$C(\boldsymbol{x}) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \cdot \mathbb{I}(T^{(m)}(\boldsymbol{x}) = k).$$

https://hastie.su.domains/Papers/samme.pdf

How to Classify The Final Result

SAMME — Stagewise Additive Modeling using a Multi-class Exponential loss function

Algorithm 2 *SAMME*

1. *Initialize the observation weights* $w_i = 1/n,\ i = 1, 2, \ldots, n.$
2. *For* $m = 1$ *to* M*:*
   (a) *Fit a classifier* $T^{(m)}(\boldsymbol{x})$ *to the training data using weights* w_i.
   (b) *Compute*

$$err^{(m)} = \sum_{i=1}^{n} w_i \mathbb{I}\left(c_i \neq T^{(m)}(\boldsymbol{x}_i)\right) / \sum_{i=1}^{n} w_i.$$

   (c) *Compute*

$$\alpha^{(m)} = \log \frac{1 - err^{(m)}}{err^{(m)}} + \log(K - 1). \tag{1}$$

   (d) *Set*

$$w_i \leftarrow w_i \cdot \exp\left(\alpha^{(m)} \cdot \mathbb{I}\left(c_i \neq T^{(m)}(\boldsymbol{x}_i)\right)\right),\ i = 1, \ldots, n.$$

   (e) *Re-normalize* w_i.
3. *Output*

$$C(\boldsymbol{x}) = \arg\max_k \sum_{m=1}^{M} \alpha^{(m)} \cdot \mathbb{I}(T^{(m)}(\boldsymbol{x}) = k).$$

https://hastie.su.domains/Papers/samme.pdf

Behind The Scene

The hypothesis function f(x) is defined as:

$$f(x) = \sum_{m=1}^{M} \alpha_m * G_m(x)$$

M trees Label: {-1,1} N Samples

Exponential loss function:

$$Err(f) = \sum_{i=1}^{N} e^{-y_i * f(x_i)}$$

Supposing that:

$$f_m(x) = f_{m-1}(x) + \alpha_m * G_m(x)$$

Expand the error function:

$$Err(f) = \sum_{i=1}^{N} e^{-y_i f_{m-1}(x_i)} * e^{-y_i \alpha_m G_m(x_i)}$$

$$=> Err(\alpha_m, G_m) = \sum_{i=1}^{N} e^{-y_i f_{m-1}(x_i)} * e^{-y_i \alpha_m G_m(x_i)}$$

$w_i = e^{-y_i f_{m-1}(x_i)}$

not dependent on α_m and Gm

Behind The Scene

Expand the error function:

$$Err(f) = \sum_{i=1}^{N} e^{-y_i f_{m-1}(x_i)} * e^{-y_i \alpha_m G_m(x_i)}$$

$$=> Err(\alpha_m, G_m) = \sum_{i=1}^{N} e^{-y_i f_{m-1}(x_i)} * e^{-y_i \alpha_m G_m(x_i)}$$

M trees

N Samples

Label: {-1,1}

$$w_i = e^{-y_i f_{m-1}(x_i)}$$

$$Err = e^{-\alpha_m} \sum_{y_i = G_m(x_i)} w_i + e^{\alpha_m} \sum_{y_i \neq G_m(x_i)} w_i$$

It's easy to show that the expression $-y_i\alpha_m G_m(x_i)$ is $-\alpha_m$ if $y_i = G_m(x_i)$ and is α_m if $y_i != G_m(x_i)$.

$$Err = e^{-\alpha_m} \sum_{y_i = G_m(x_i)} w_i + e^{\alpha_m} \sum_{y_i \neq G_m(x_i)} w_i$$

Total weight T_w as:

$$T_w = \sum w_i$$

Error weight E_w :

$$E_w = \sum_{y_i \neq G_m(x_i)} w_i$$

$$Err = e^{-\alpha_m}(T_w - E_w) + e^{\alpha_m} E_w$$

$$\frac{dErr}{d\alpha_m} = -\alpha_m e^{-\alpha_m}(T_w - E_w) + \alpha_m e^{\alpha_m} E_w$$

Behind The Scene

$$Err = e^{-\alpha_m}(T_w - E_w) + e^{\alpha_m} E_w$$

$$\frac{dErr}{d\alpha_m} = -\alpha_m e^{-\alpha_m}(T_w - E_w) + \alpha_m e^{\alpha_m} E_w$$

Taking log on both sides we have

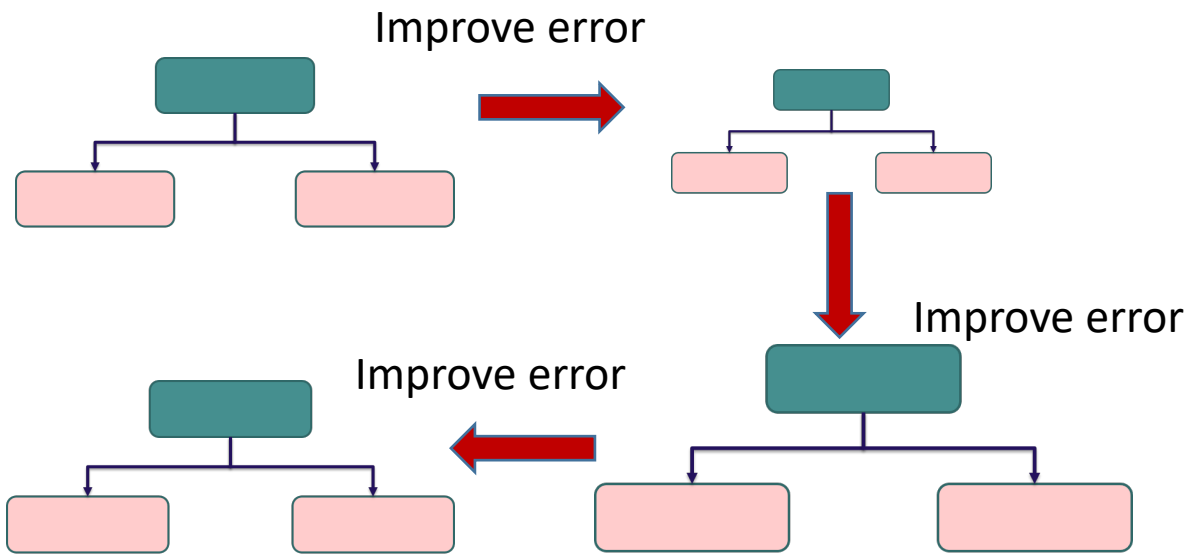
$$\begin{aligned}
&-\alpha_m e^{-\alpha_m}(T_w - E_w) + \alpha_m e^{\alpha_m} E_w = 0 \\
&=> e^{\alpha_m} E_w - e^{-\alpha_m}(T_w - E_w) = 0 \\
&=> e^{\alpha_m} E_w = e^{-\alpha_m}(T_w - E_w) \\
&=> e^{2\alpha_m} = (Tw - Ew)/Ew \\
&=> e^{2\alpha_m} = \frac{1 - (E_w/T_w)}{(E_w/T_w)}
\end{aligned}$$

$$\alpha_m = \frac{1}{2} log \frac{1 - err_m}{err_m}$$

$$err_m = \frac{E_w}{T_w} = \frac{\sum_{y_i \neq G_m(x_i)} w_i}{\sum w_i}$$

Summary

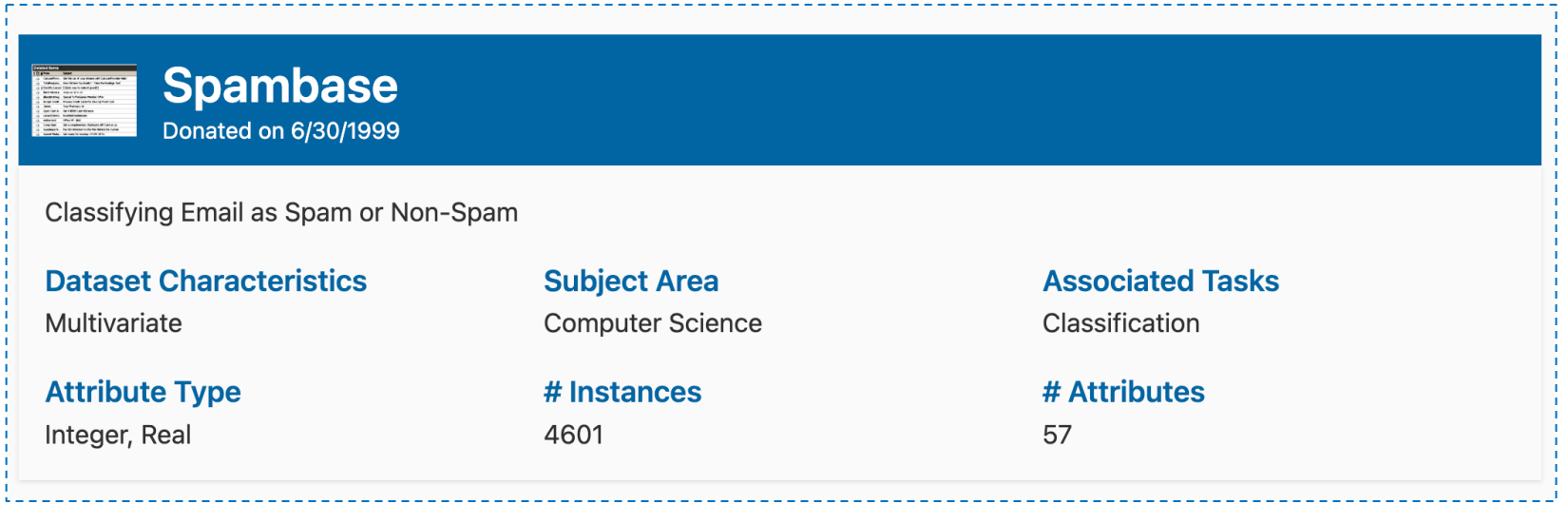
Height	Favorite Color	Gender	Weight
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57



Adaboost builds a stump based on the the error made by previous stumps

Example: Spam Classification

- Classifying Email as Spam or Non-Spam



Spambase

Donated on 6/30/1999

Classifying Email as Spam or Non-Spam

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Computer Science	Classification
Attribute Type	**# Instances**	**# Attributes**
Integer, Real	4601	57

http://archive.ics.uci.edu/dataset/94/spambase

Example: Spam Classification

1. Initialize the observation weights $w_i = 1/N,\ i = 1, 2, \ldots, N$.
2. For $m = 1$ to M:
   (a) Fit a classifier $G_m(x)$ to the training data using weights w_i.
   (b) Compute
   $$\text{err}_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i}.$$
   (c) Compute $\alpha_m = \log((1 - \text{err}_m)/\text{err}_m)$.
   (d) Set $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))],\ i = 1, 2, \ldots, N$.
3. Output $G(x) = \text{sign}\left[\sum_{m=1}^{M} \alpha_m G_m(x)\right]$.

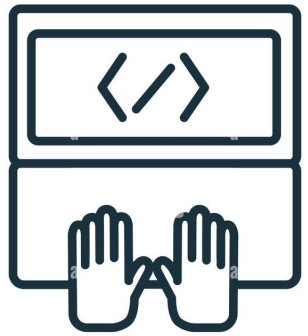
Example: Spam Classification

Our implementation

```
# Fit model
ab = AIVNAdaBoost()
ab.fit(X_train, y_train, M = 50)

# Predict on test set
y_pred = ab.predict(X_test)
print('The accuracy_score of the model is:', round(accuracy_score(y_test, y_pred), 4))
```

```
The accuracy_score of the model is: 0.9392
```



CUSTOM CODE

```
from sklearn.ensemble import AdaBoostClassifier

ab_sk = AdaBoostClassifier(n_estimators = 50)
ab_sk.fit(X_train, y_train)
y_pred_sk = ab_sk.predict(X_test)
print('The accuracy_score of the model is:', round(accuracy_score(y_test, y_pred_sk), 4))
```

```
The accuracy_score of the model is: 0.9435
```

Using Sklearn library

Outline

- Boosting Technique
- AdaBoost and Its Application
- Gradient Boost and Its Application

Gradient Boost For Regression

Height	Favorite Color	Gender	Weight
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Input **Output**

Tree-based Gradient Boost

- Step 1: Build 1st tree
  - Calculate the average of weights

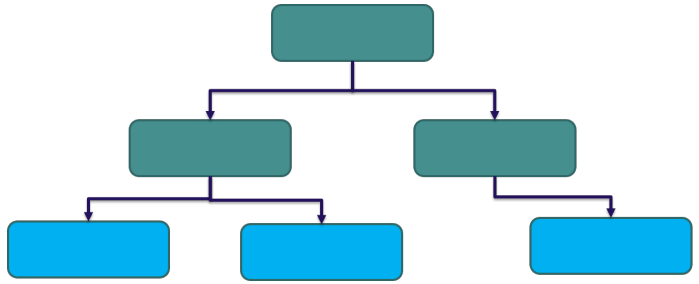
Height	Favorite Color	Gender	Weight
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Node of 1st Tree Average of Weights: 71.17

Tree-based Gradient Boost

- Step 2:
  - Build 2nd tree

Average of weights: 71.17

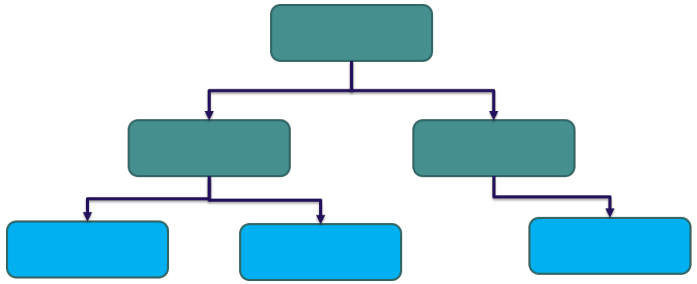


Height	Favorite Color	Gender	Weight
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

Tree-based Gradient Boost

- Step 2:
  - Build 2nd tree

Average of weights: 71.2

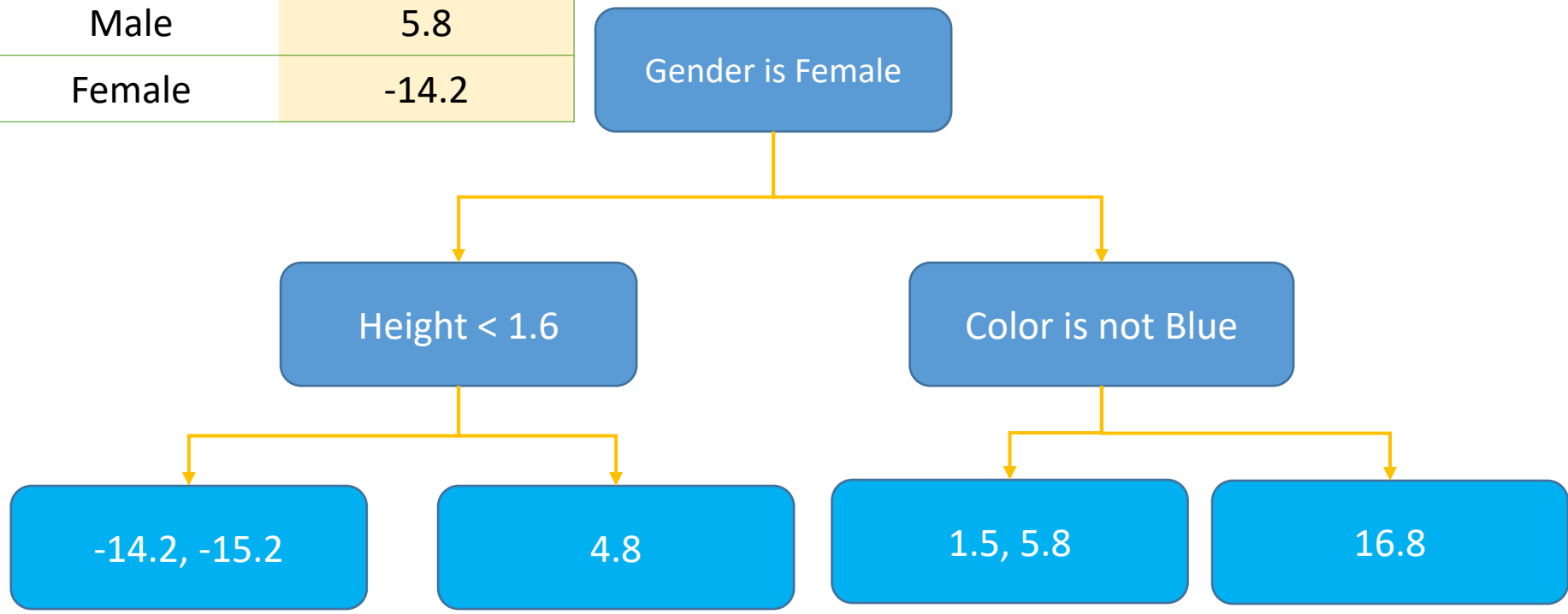


Height	Favorite Color	Gender	Weight	Residual Error
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	1.8
1.5	Blue	Female	56	-15.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

Tree-based Gradient Boost

Height	Favorite Color	Gender	Residual Error
1.6	Blue	Male	16.8
1.6	Green	Female	1.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2

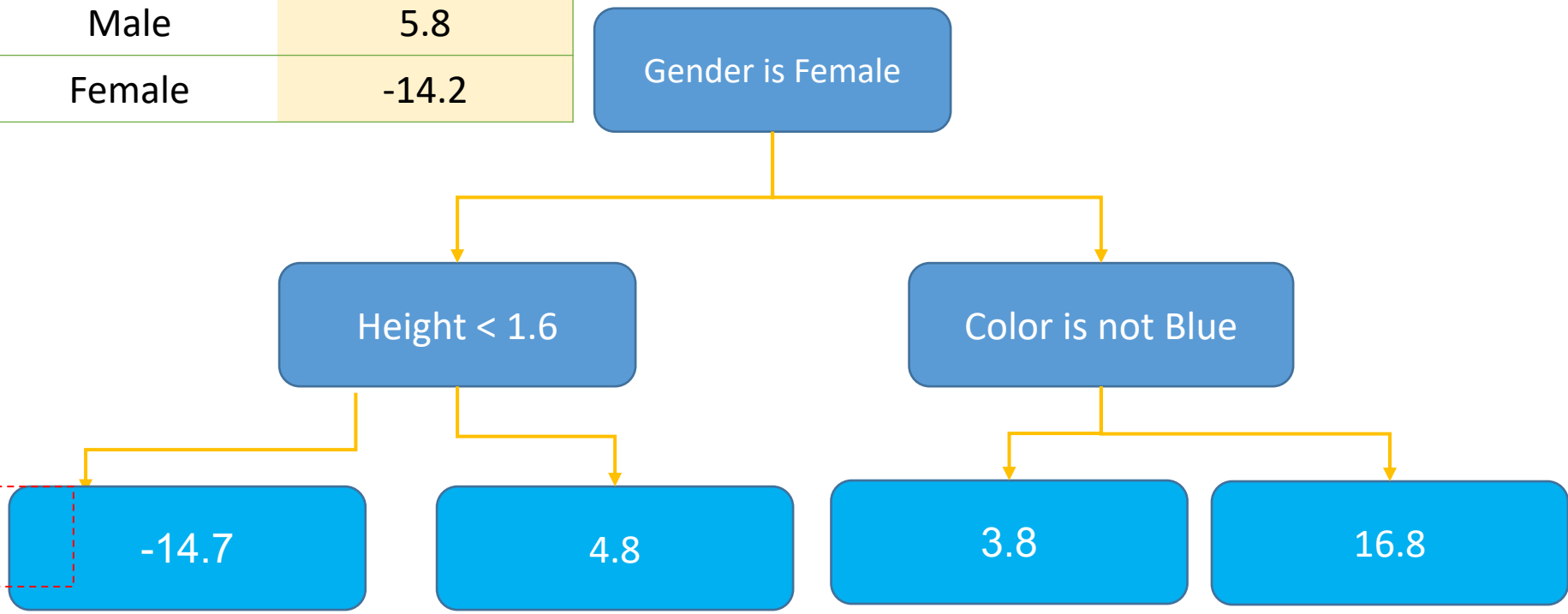
Tại sao lại xây dựng cây dự đoán Residual Error



Tree-based Gradient Boost

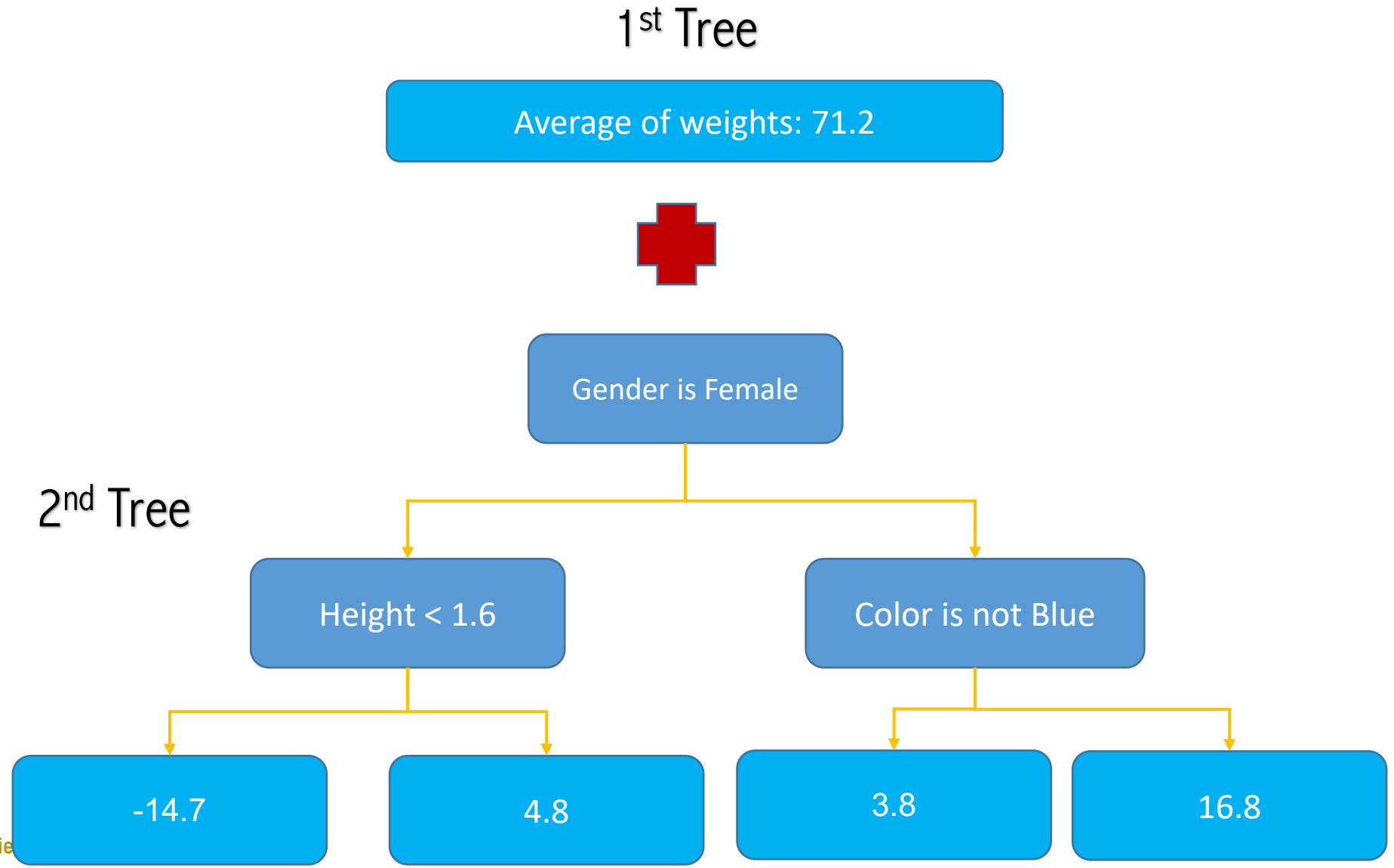
Height	Favorite Color	Gender	Residual Error
1.6	Blue	Male	16.8
1.6	Green	Female	1.8
1.5	Blue	Female	-15.2
1.8	Red	Male	1.8
1.5	Green	Male	5.8
1.4	Blue	Female	-14.2

Tại sao lại xây dựng cây dự đoán Residual Error



Trung bình residual

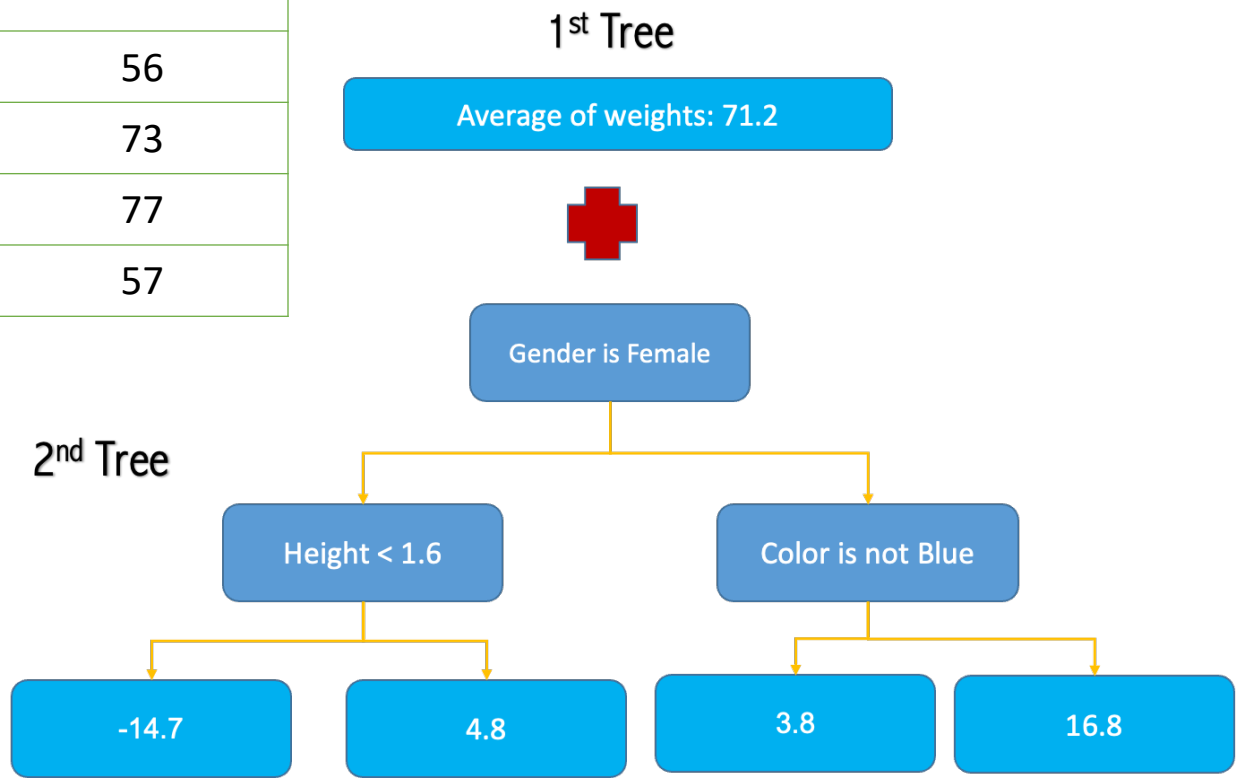
Tree-based Gradient Boost



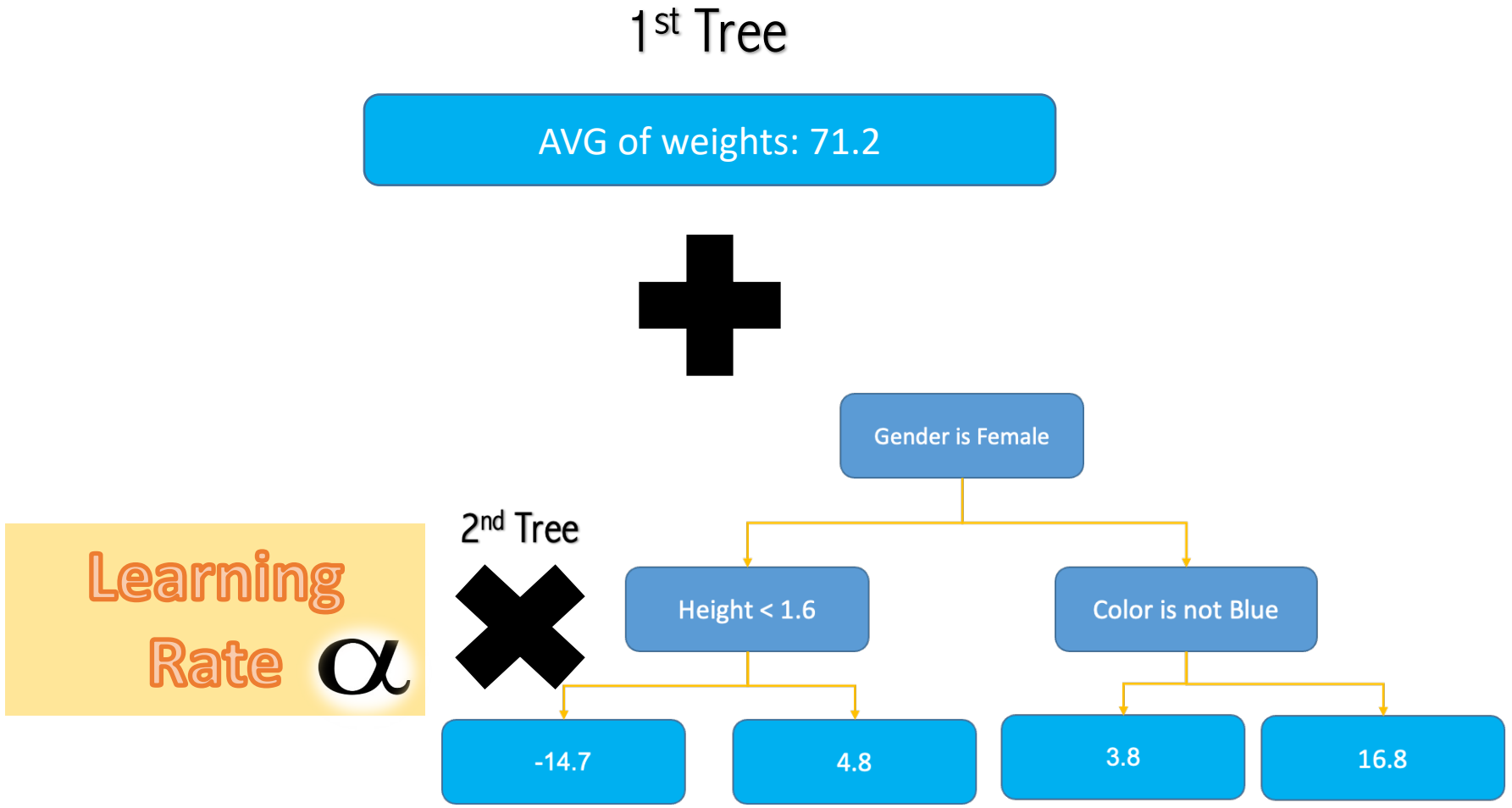
Prediction

Height	Favorite Color	Gender	Weight	Prediction
1.6	Blue	Male	88	88
1.6	Green	Female	76	76
1.5	Blue	Female	56	56
1.8	Red	Male	73	73
1.5	Green	Male	77	77
1.4	Blue	Female	57	57

OVER FITTING



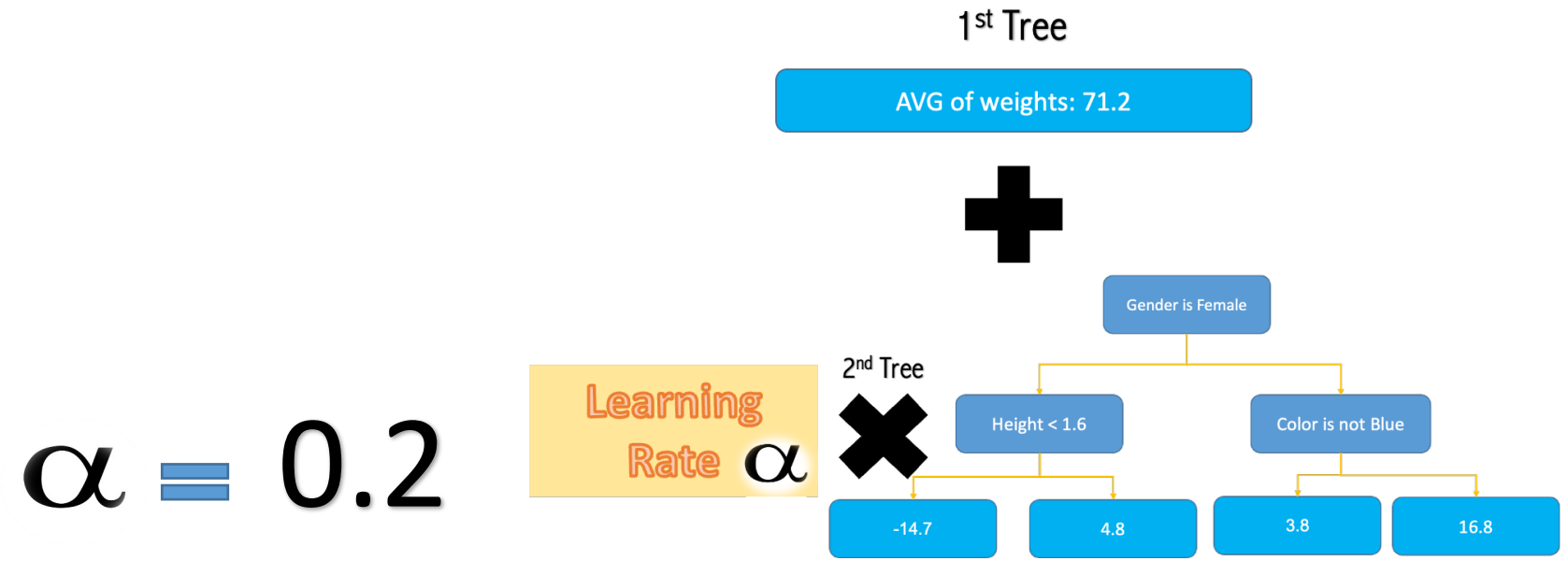
Prediction


1st Tree
AVG of weights: 71.2
Learning Rate α
2nd Tree
Gender is Female
Height < 1.6
Color is not Blue
-14.7
4.8
3.8
16.8

Prediction

Height	Favorite Color	Gender	Weight	Prediction
1.6	Blue	Male	88	74.56

Prediction = 71.2 + 0.2 * 16.8 = 74.56

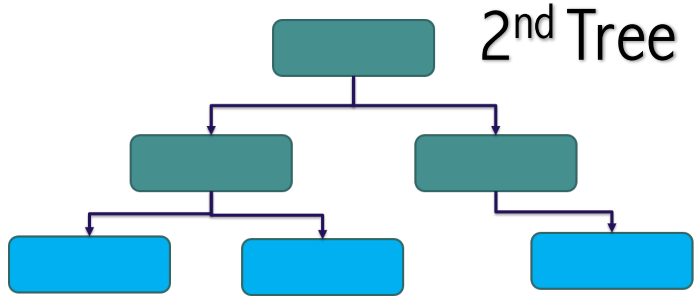


Tree-based Gradient Boost

- Step 3:
  - Build 3rd tree

1st Tree

Average of weights: 71.2



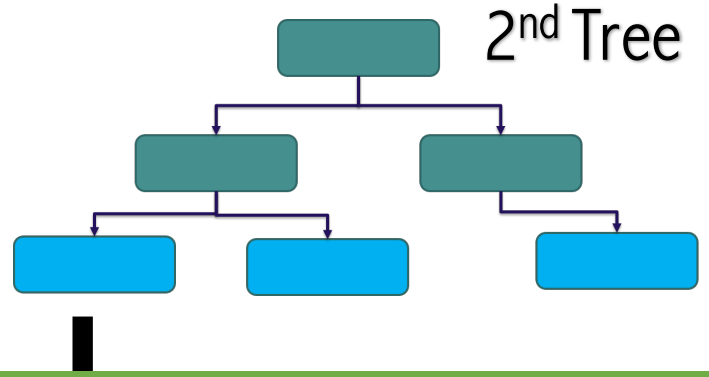
Height	Favorite Color	Gender	Weight	Predicted Weight	Residual Error
1.6	Blue	Male	88	74.56	12.44
1.6	Green	Female	76	…	…
1.5	Blue	Female	56	…	…
1.8	Red	Male	73	…	…
1.5	Green	Male	77	…	…
1.4	Blue	Female	57	…	…

Tree-based Gradient Boost

- Step 3:
  - Build 3rd tree

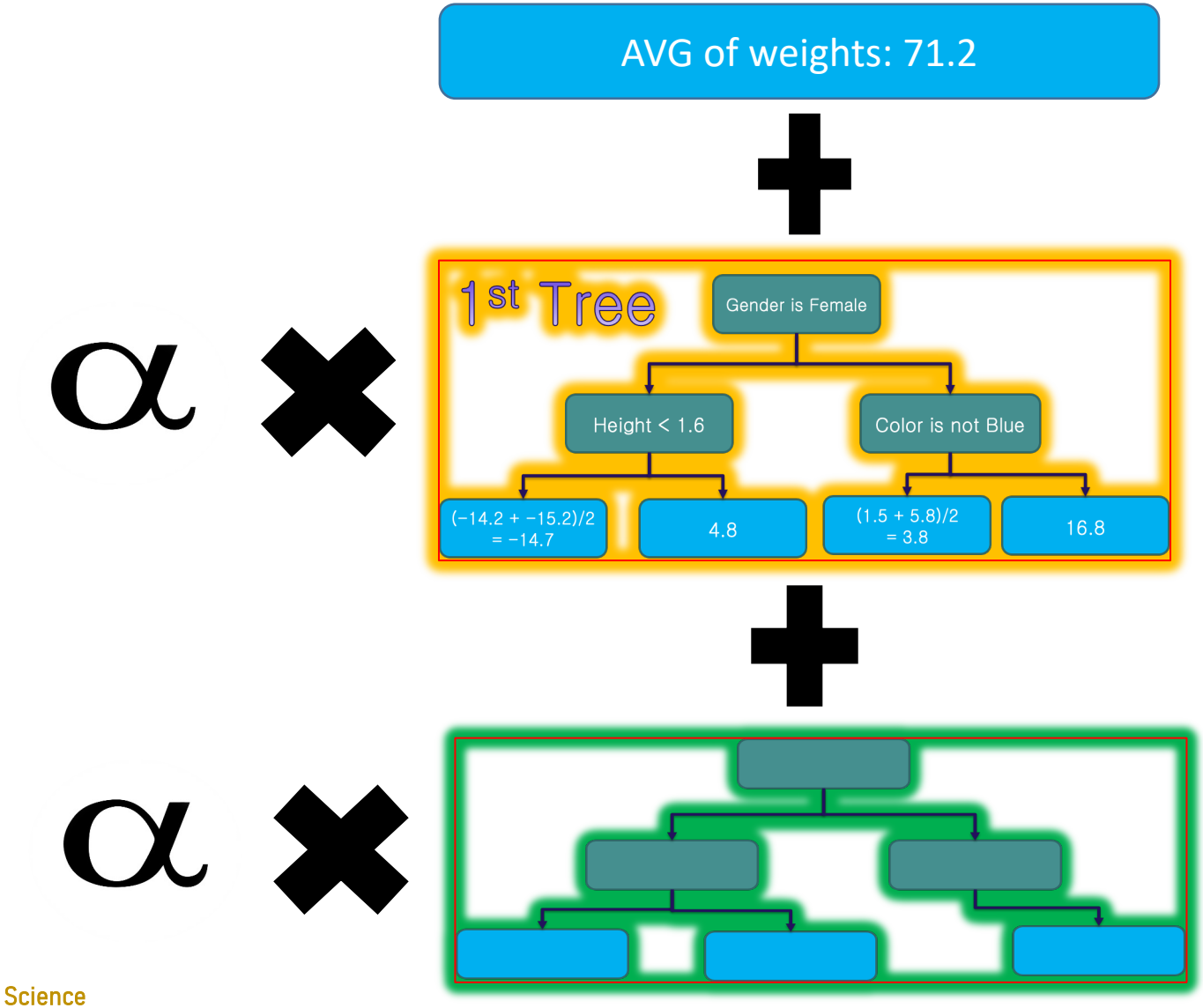
1st Tree

Average of weights: 71.2



Height	Favorite Color	Gender	First Tree Residual	Second Tree Residual	Third Tree Residual
1.6	Blue	Male	16.8	12.44	
1.6	Green	Female	1.8	…	
1.5	Blue	Female	-15.2	…	
1.8	Red	Male	1.8	…	
1.5	Green	Male	5.8	…	
1.4	Blue	Female	-14.2	…	

Prediction


AVG of weights: 71.2
α
1st Tree
Gender is Female
Height < 1.6
Color is not Blue
(−14.2 + −15.2)/2
= −14.7
4.8
(1.5 + 5.8)/2
= 3.8
16.8
α

Gradient Boost: Behind The Scenes

Loss Function:

n

Height	Favorite Color	Gender	Weight
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

- $\{(x_i, y_i)\}^n_1$

Loss function = $L(y_i, F(x)) = 1/2 * (\text{Output} - \text{Predicted})^2$

$$J = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y} - y)^2$$

$$\frac{dL}{dPredicted} = 2/2\ (\text{Output} - \text{Predicted}) * -1 = -\ (\text{Output} - \text{Predicted})$$

Tricky implementation here

Gradient Boost: Behind The Scenes

- Step 1:
  - Initialize a model with a constant value:
  - $F_{0(x)} = \underset{\delta}{argmin} \sum_{i=1}^{n} L(y, \delta)$

			y
Height	Favorite Color	Gender	Weight
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56

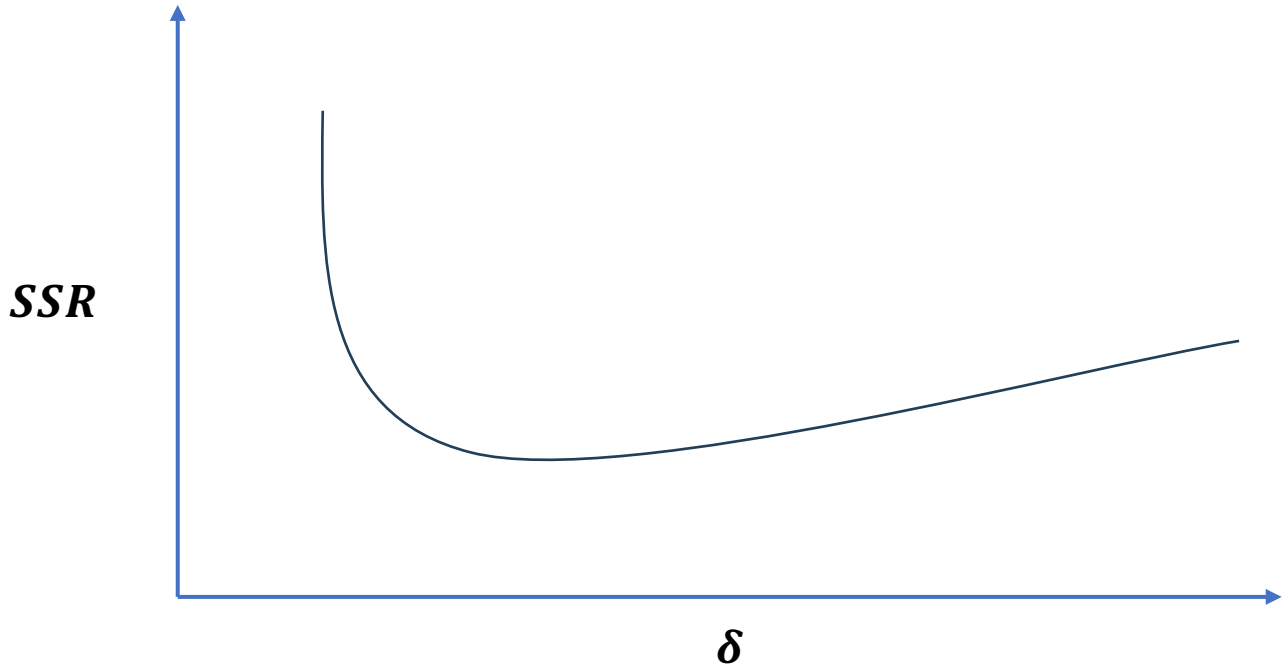
SSR = 1/2 {(88 – δ)^2 + (76 – δ)^2 + (56 - δ)^2}

$\frac{dSSR}{d\delta}$ = -(88 - δ) – (76 - δ) – (56 - δ) = 0

$\delta = \frac{88+76+56}{3}$ = 73.3 = average of all sample' weights

Gradient Boost: Behind The Scenes

- Step 1:
- Initialize a model with a constant value:
  - $\boldsymbol{F_{0(x)} = argmin \sum_{i=1}^{n} L(y, \delta)} = 73.3$



Gradient Boost: Behind The Scenes

- M is the number of trees
- n is the number of samples

Step 2: for $m = 1$ to M:

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

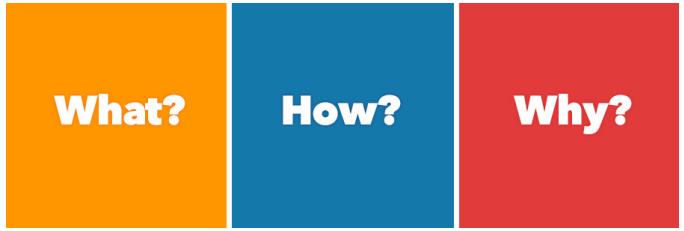
(B) Fit a regression tree to the r_{im} values and create terminal regions R_{jm}, for $j = 1\ldots J_m$

(C) For $j = 1\ldots J_m$ compute $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma)$

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

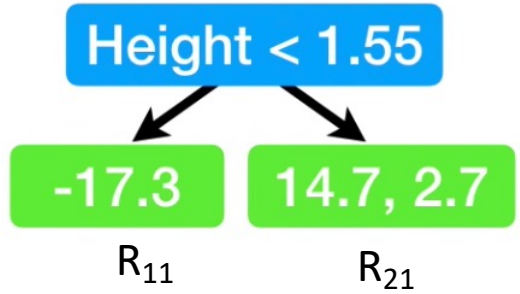
Gradient Boost: Behind The Scenes

(B) Fit a regression tree to the r_{im} values and create terminal regions R_{jm}, for $j = 1 \ldots J_m$



R is the residual error; m is the m-th tree; j is the j-th leaf

Height	Favorite Color	Gender	Weight	r_{i1}
1.6	Blue	Male	88	14.7
1.6	Green	Female	76	2.7
1.5	Blue	Female	56	-17.3



Step 1

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

(C) For $j = 1 \ldots J_m$ compute $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$

$$F_{m\_1}(x) = F_0(x) = 73.3$$

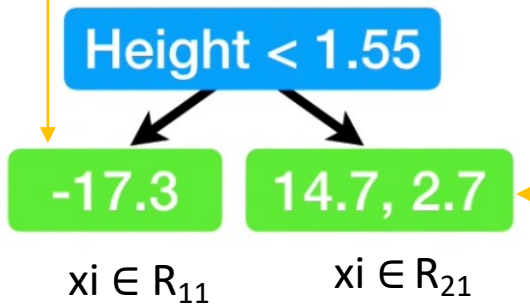
R is the residual error; m is the m-th tree; j is the i-th leaf

Height	Favorite Color	Gender	Weight	r_{i1}
1.6	Blue	Male	88	14.7
1.6	Green	Female	76	2.7
1.5	Blue	Female	56	-17.3

$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2} \{y3 - (F0(x3) + \gamma)\}^2\right]$$
$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2} \{56 - (73.3 + \gamma)\}^2\right]$$
$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2} \{-17.3 - \gamma\}^2\right]$$

γ_{11} = -17.3

γ_{21} = 8.7



Step 1

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

(C) For $j = 1 \ldots J_m$ compute $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$

$$F_{m\_1}(x) = F_0(x) = 73.3$$

R is the residual error; m is the m-th tree; j is the i-th leaf

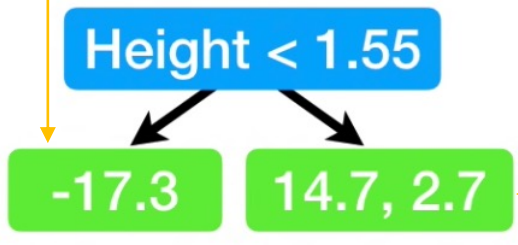
Height	Favorite Color	Gender	Weight	r_{i1}
1.6	Blue	Male	88	14.7
1.6	Green	Female	76	2.7
1.5	Blue	Female	56	-17.3

$$\gamma_{21} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2}\{y1 - (F0(x1) + \gamma)\}^2 + \frac{1}{2}\{y2 - (F0(x2) + \gamma)\}^2\right]$$

$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2}\{88 - (73.3 + \gamma)\}^2 + \frac{1}{2}\{76 - (73.33 + \gamma)\}^2\right]$$

$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \left[\frac{1}{2}\{14{,}7 - \gamma\}^2 + \frac{1}{2}\{2.7 - \gamma)\}^2\right]$$

$\gamma_{21} = 8.7$



xi ∈ R_{11} xi ∈ R_{21}

$\gamma_{11} = -17.3$ $\gamma_{21} = 8.7$

Giá trị trung bình của 2 samples

Step 2

Step 1

$$xi \in R_{11}$$

What? How? Why?

$$F_{m\_1}(x) = F_0(x) = 73.3$$

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

R is the residual error; m is the m-th tree; j is the i-th leaf

Height	Favorite Color	Gender	Weight	r_{i1}	F1(X)
1.6	Blue	Male	88	14.7	74.2
1.6	Green	Female	76	2.7	74.2
1.5	Blue	Female	56	-17.3	71.6

$xi \in R_{11}$ $xi \in R_{21}$

$\gamma_{11} = -17.3$ $\gamma_{21} = 8.7$

Repeat for the next m + 1 iteration

Step 2: for $m = 1$ to M:

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

(B) Fit a regression tree to the r_{im} values and create terminal regions R_{jm}, for $j = 1\ldots J_m$

(C) For $j = 1\ldots J_m$ compute $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma)$

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

Summary

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, a differentiable loss function $L(y, F(x))$, number of iterations M.

Algorithm:

1. Initialize model with a constant value:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma).$$

2. For m = 1 to M:
    1. Compute so-called *pseudo-residuals*:

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

    2. Fit a base learner (e.g. tree) $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $\{(x_i, r_{im})\}_{i=1}^n$.
    3. Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\right).$$

    4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. Output $F_M(x)$.

Sales Prediction

TV	Radio	Newspaper	Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	12
151.5	41.3	58.5	16.5
180.8	10.8	58.4	17.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2
8.6	2.1	1	4.8
199.8	2.6	21.2	15.6

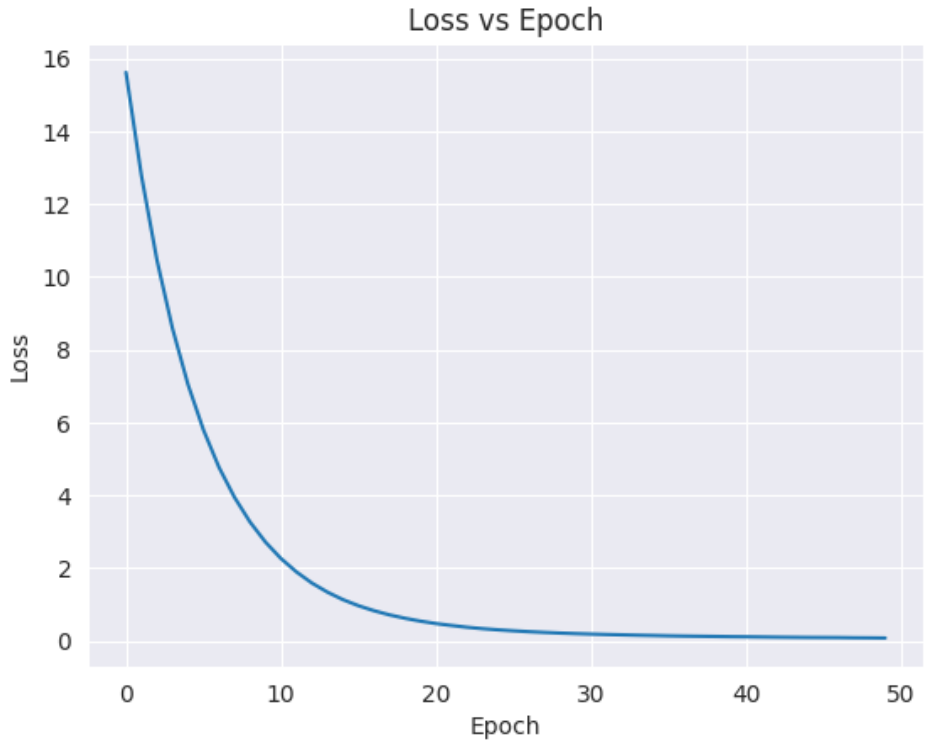
Sales Prediction

```python
import numpy as np
from sklearn.tree import DecisionTreeRegressor

class AIVNGradientBooster:

    def __init__(self, max_depth=8, min_samples_split=5, min_samples_leaf=5, max_features=3, lr=0.1, num_iter=50):
        self.max_depth = max_depth
        self.min_samples_split = min_samples_split
        self.min_samples_leaf = min_samples_leaf
        self.max_features = max_features
        self.lr = lr
        self.num_iter = num_iter
        self.y_mean = 0
```

```python
#READ DATA
data = pd.read_csv("/content/advertising.csv")
data.fillna(0,inplace=True)
#X,y
X = data.iloc[:,:-1].values
y = data.iloc[:,-1].values
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=100)
#scaling
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.fit_transform(X_test)
y_train = np.array(y_train).reshape(X_train.shape[0],1)
y_test = np.array(y_test).reshape(X_test.shape[0],1)
#TRAIN
G = AIVNGradientBooster()
models, losses, pred_0 = G.train(X_train,y_train)
```




```python
sns.set_style('darkgrid')
ax = sns.lineplot(x=range(50),y=losses)
ax.set(xlabel='Epoch',ylabel='Loss',title='Loss vs Epoch')
```


Thank you