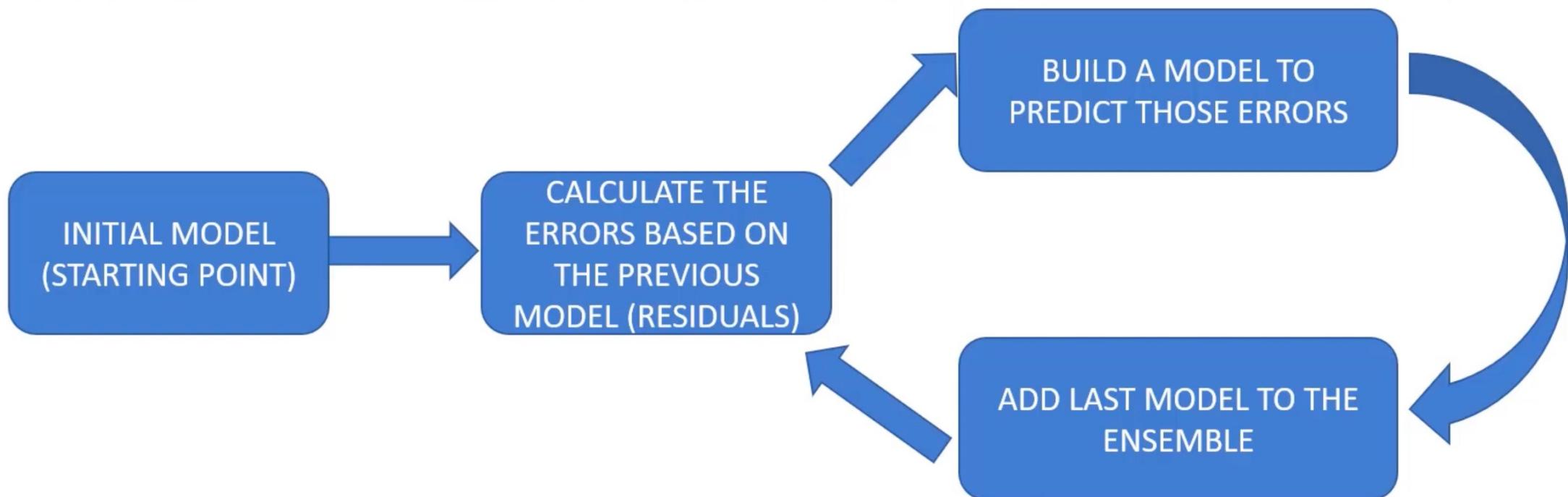


XG BOOSTING



- XGBoost repeatedly builds new models and combine them into an ensemble model
- Initially build the first model and calculate the error for each observation in the dataset
- Then you build a new model to predict those residuals (errors)
- Then you add prediction from this model to the ensemble of models
- XGboost is superior compared to gradient boosting algorithm since it offers a good balance between bias and variance (Gradient boosting only optimized for the variance so tend to overfit training data while Xgboost offers regularization terms that can improve model generalization).



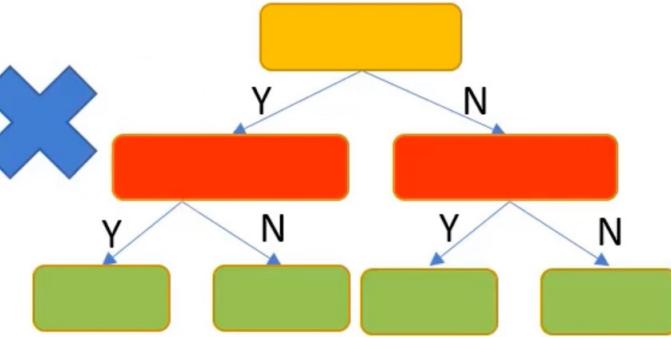
INITIAL GUESS
(AVERAGE)

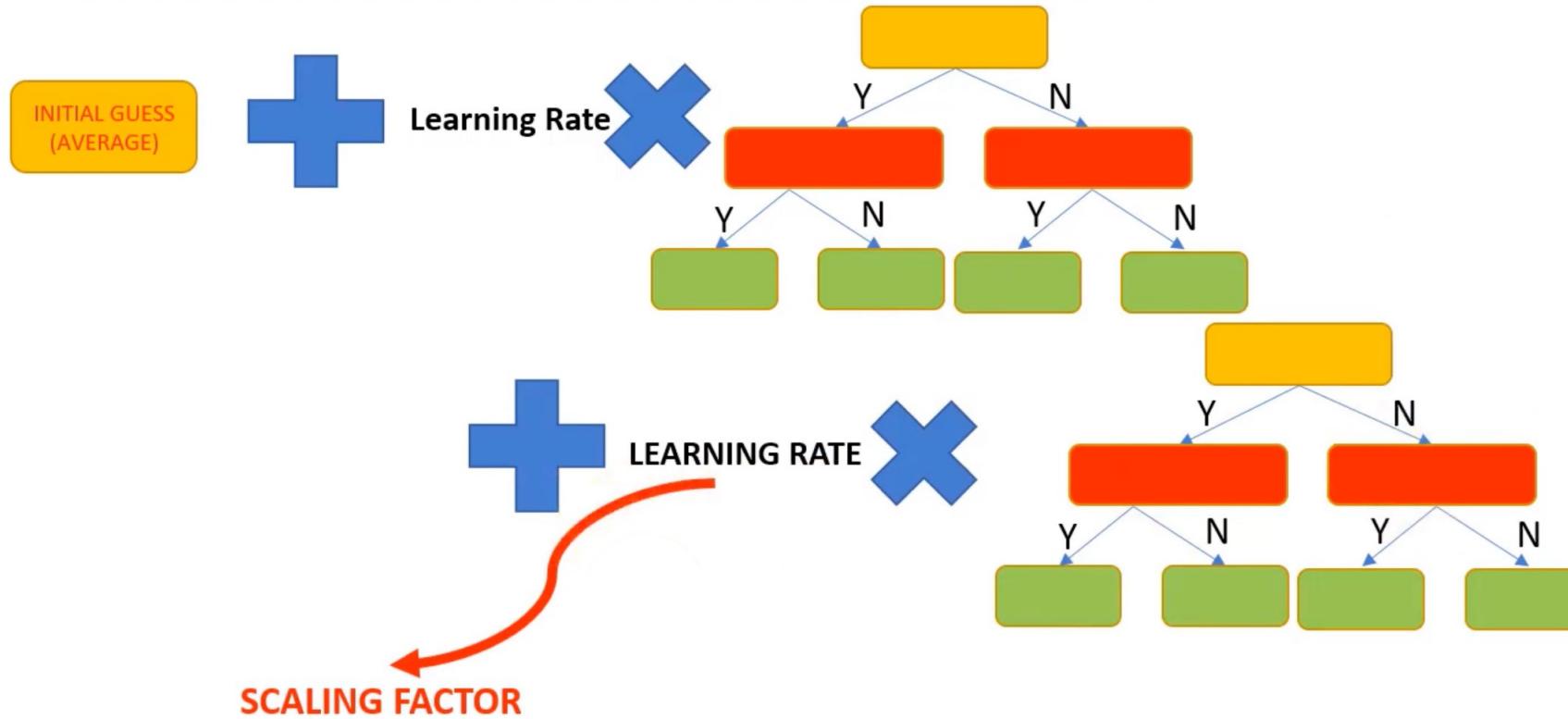


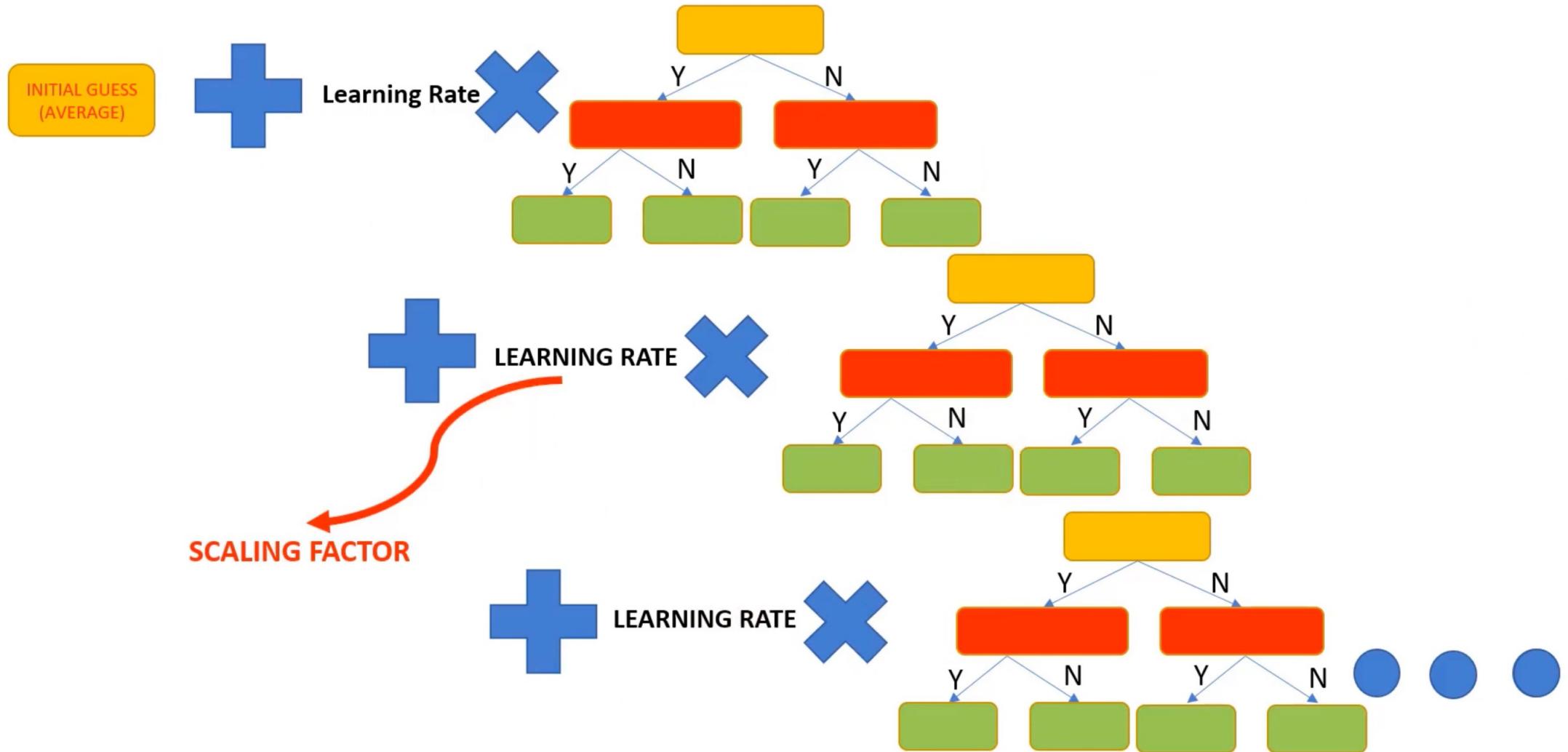
INITIAL GUESS
(AVERAGE)



Learning Rate

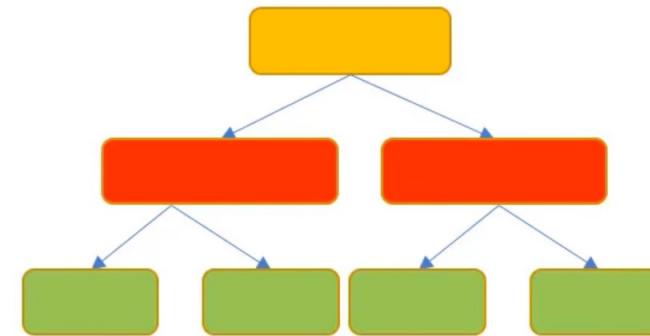






- Gradient boost works by building a tree based on the error (residuals) from the previous tree.
- Gradient boost scales the trees and then adds the predictions from the new tree to the predictions from previous trees.

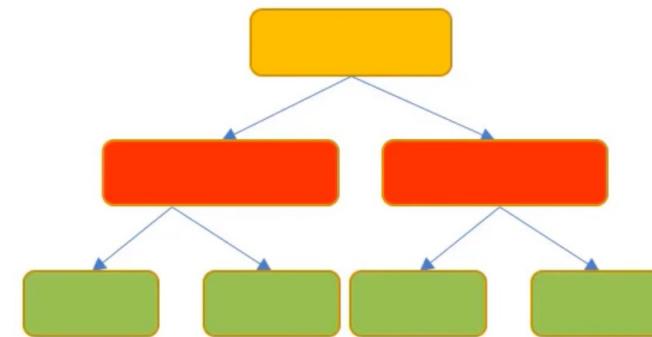
Height	Color	Gender	Weight (Kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57



- Gradient boost works by building a tree based on the error (residuals) from the previous tree.
- Gradient boost scales the trees and then adds the predictions from the new tree to the predictions from previous trees.

Height	Color	Gender	Weight (Kg)
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 FEATURES

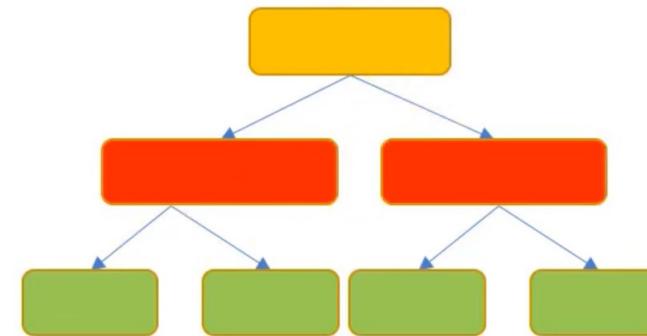


- Gradient boost works by building a tree based on the error (residuals) from the previous tree.
- Gradient boost scales the trees and then adds the predictions from the new tree to the predictions from previous trees.

Height	Color	Gender	Weight (Kg)
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 FEATURES

VARIABLE TO BE
PREDICTED



- Let's assume that the initial model predictions (starting point) is the average weight is 71.2
- Gradient boost builds a tree based on the error from the first tree.
- The tree is built by assuming that the features (heights, color, and gender) predicts the residuals (new column that we have just created).

71.2

INITIAL STARTING
POINT (PREDICTIONS)



- Let's assume that the initial model predictions (starting point) is the average weight is 71.2
- Gradient boost builds a tree based on the error from the first tree.
- The tree is built by assuming that the features (heights, color, and gender) predicts the residuals (new column that we have just created).

71.2

**INITIAL STARTING
POINT (PREDICTIONS)**

Height	Color	Gender	Weight (Kg)
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 FEATURES



- Let's assume that the initial model predictions (starting point) is the average weight is 71.2
- Gradient boost builds a tree based on the error from the first tree.
- The tree is built by assuming that the features (heights, color, and gender) predicts the residuals (new column that we have just created).

71.2

**INITIAL STARTING
POINT (PREDICTIONS)**

Height	Color	Gender	Weight (Kg)
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

Error = True – predicted
$88 - 71.2 = 16.8$
$76 - 71.2 = 4.8$
$56 - 71.2 = -15.2$
$73 - 71.2 = 1.8$
$77 - 71.2 = 5.8$
$57 - 71.2 = -14.2$

INPUT FEATURES

**ERRORS
(RESIDUALS)**

- Let's assume that the initial model predictions (starting point) is the average weight is 71.2
- Gradient boost builds a tree based on the error from the first tree.
- The tree is built by assuming that the features (heights, color, and gender) predicts the residuals (new column that we have just created).

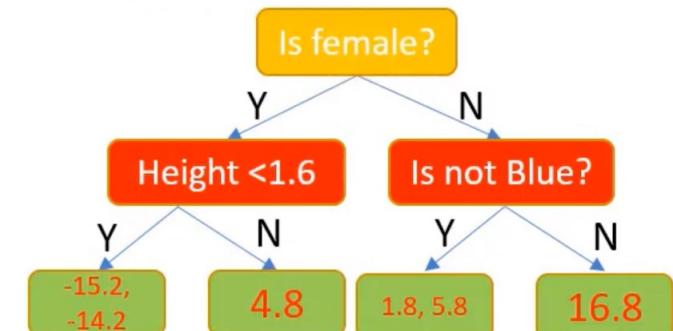
Height	Color	Gender	Weight (Kg)
1.6	Blue	Male	88
1.6	Green	Female	76
1.5	Blue	Female	56
1.8	Red	Male	73
1.5	Green	Male	77
1.4	Blue	Female	57

INPUT FEATURES

ERRORS
(RESIDUALS)

Error = True – predicted
$88-71.2=16.8$
$76-71.2=4.8$
$56-71.2=-15.2$
$73-71.2=1.8$
$77-71.2=5.8$
$57-71.2=-14.2$

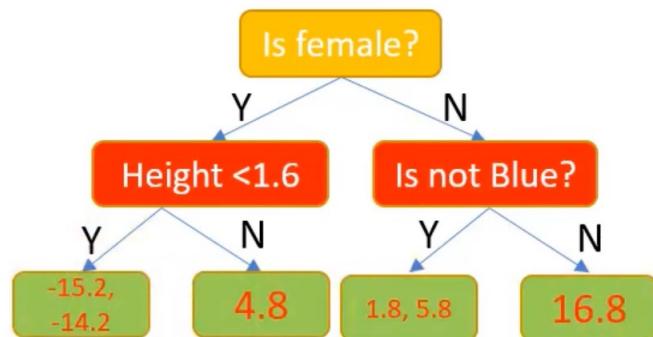
71.2
INITIAL STARTING
POINT (PREDICTIONS)



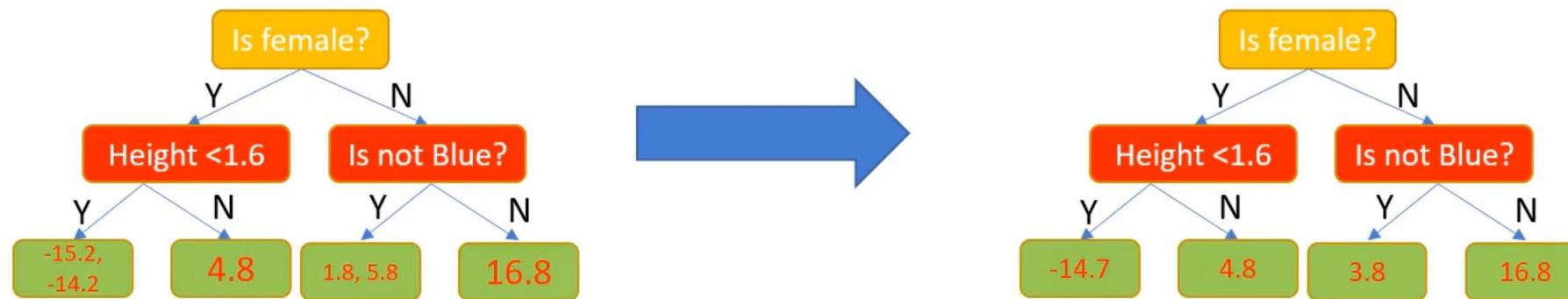
- Note that the number of leaves is restricted to 4 in this example for the sake of simplicity.
- Let's replace the values with the average a shown below.



- Note that the number of leaves is restricted to 4 in this example for the sake of simplicity.
- Let's replace the values with the average a shown below.



- Note that the number of leaves is restricted to 4 in this example for the sake of simplicity.
- Let's replace the values with the average a shown below.



$$Average_1 = (-15.2 - 14.2)/2 = -14.7$$

$$Average_2 = (1.8 + 5.8)/2 = 3.8$$



- Now that we have built a tree, let's combine the previous predictions with the new tree to generate new predictions!

71.2

AVERAGE WEIGHT
(INITIAL GUESS FROM
PREVIOUS STEP)

Height	Color	Gender	Weight (Kg)
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



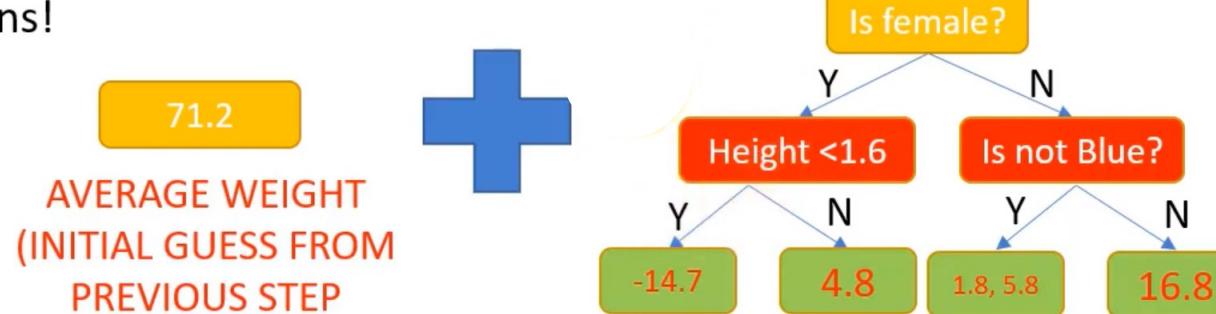
- We add a learning rate (range from 0 to 1) to overcome this issue.
- This parameter is used for scaling purposes by adjusting the newly added information from the new tree.
- Adding this tree and scaling it with the learning rate helps us get a little closer to the true values.
- By taking smaller steps, the model results in better predictions on the testing dataset (low variance).



Height	Color	Gender	Weight (Kg)
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



- Now that we have built a tree, let's combine the previous predictions with the new tree to generate new predictions!



Height	Color	Gender	Weight (Kg)
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

New predictions
88

$$\text{Predictions} = 71.2 + 16.8 = 88$$



- We add a learning rate (range from 0 to 1) to overcome this issue.
- This parameter is used for scaling purposes by adjusting the newly added information from the new tree.
- Adding this tree and scaling it with the learning rate helps us get a little closer to the true values.
- By taking smaller steps, the model results in better predictions on the testing dataset (low variance).

71.2



AVERAGE WEIGHT
(INITIAL GUESS FROM
PREVIOUS STEP)

Height	Color	Gender	Weight (Kg)
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



- We add a learning rate (range from 0 to 1) to overcome this issue.
- This parameter is used for scaling purposes by adjusting the newly added information from the new tree.
- Adding this tree and scaling it with the learning rate helps us get a little closer to the true values.
- By taking smaller steps, the model results in better predictions on the testing dataset (low variance).



Height	Color	Gender	Weight (Kg)
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

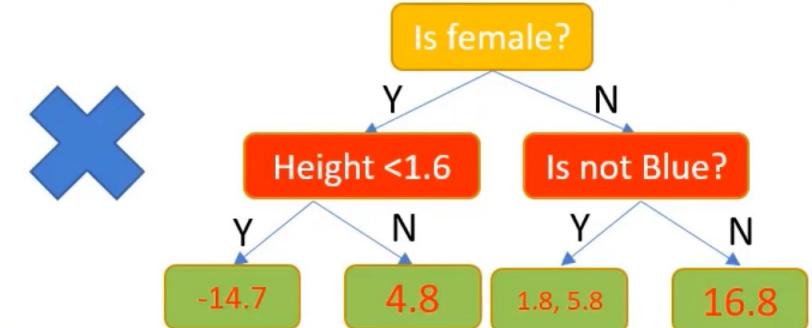


- We add a learning rate (range from 0 to 1) to overcome this issue.
- This parameter is used for scaling purposes by adjusting the newly added information from the new tree.
- Adding this tree and scaling it with the learning rate helps us get a little closer to the true values.
- By taking smaller steps, the model results in better predictions on the testing dataset (low variance).



71.2
 + LEARNING RATE
 AVERAGE WEIGHT
 (INITIAL GUESS FROM
 PREVIOUS STEP)
 0.1

Height	Color	Gender	Weight (Kg)
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



New predictions	
72.9	

Predictions = $71.2 + 0.1 * 16.8 = 72.9$



- Now let's build another tree with the new residuals from the new predictions.

Height	Color	Gender	Weight (Kg)
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

Predictions = 71.2 +
 $0.1 * 16.8 = 72.9$

RECALL THAT THESE
ARE THE INITIAL
RESIDUALS

Initial Residuals
16.8
4.8
-15.2
1.8
5.8
-14.2

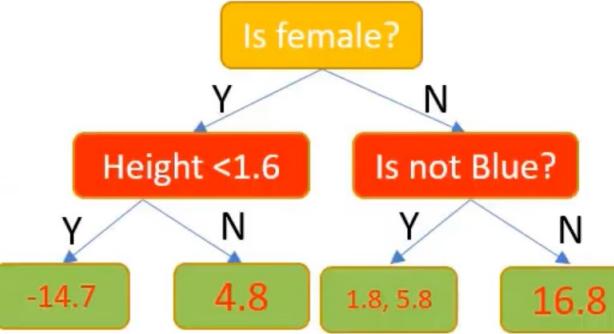
New Predictions	New Residuals
72.9	$88 - 72.9 = 15.1$
	4.3
	-13.7
	1.4
	5.4
	-12.7

RESIDUALS
HAVE GONE
DOWN!

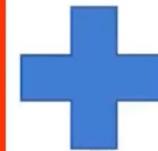
71.2
AVERAGE WEIGHT
(INITIAL GUESS FROM
PREVIOUS STEP)



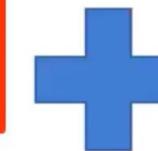
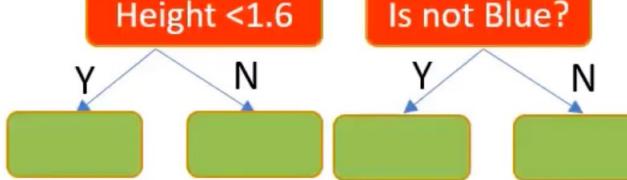
LEARNING RATE
0.1



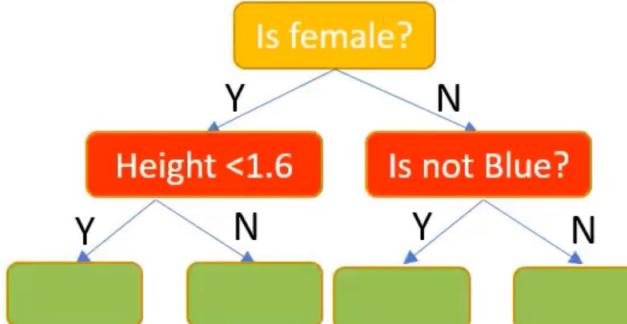
**NOW YOU CAN
MAKE NEW
PREDICTIONS BY
COMBINING ALL
THE SCALED
PREDICTIONS
FROM ALL TREES**



LEARNING RATE
0.1



LEARNING RATE
0.1



SOME REFERENCES

- <https://stats.stackexchange.com/questions/282459/xgboost-vs-python-sklearn-gradient-boosted-trees>
- <https://machinelearningmastery.com/xgboost-with-python/>
- <https://www.youtube.com/watch?v=OtD8wVaFm6E>
- <https://www.quora.com/What-is-the-difference-between-the-R-gbm-gradient-boosting-machine-and-xgboost-extreme-gradient-boosting/answer/Tianqi-Chen-1>
- <https://towardsdatascience.com/beginners-guide-to-xgboost-for-classification-problems-50f75aac5390>
- <https://arxiv.org/pdf/1603.02754.pdf>

