Gradient boost (regression)

Difference between Adaboost and Gradient boost

If we want to use this data to predict weight

Height (m)	Favorite Color	Gender	Weight (kg)
1.6	Blue	Male	88
1.6	Green	Female	76
etc…	etc…	etc…	etc…

Adaboost starts by building "stump" from the training data, and assign an initial weight → Calculate the error and the amount of say for the stump → Based on the error, update the weight and grow the next stump

Gradient starts by making a single "leaf" ~ which represents the initial guess for the weights for all the samples → Calculate the error and the for this fixed size tree → Based on the error, update the next tree

Each tree/stump will do a better job than the previous stump and we do this continuously until we get a satisfied result

Then gradient boost start growing tree (larger than the a stump but not a full tree, e.g., maximum leaf less than 4)

How gradient boost works

Height (m)	Favorite 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

Assuming that if we have the above dataset

How gradient boost works

Height (m)	Favorite 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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

How gradient boost works

Height (m)	Favorite 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

71.2

Step 1: calculate the original leaf

The first leave usually has the value of average of target variable (e.g., “weight” in this example), which is 71.2 in this example

Assuming that if we have the above dataset

How gradient boost works

Height (m)	Favorite 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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

In other words, if we stop now, all the predictions (does not matter the input) will have the prediction of 71.2

How gradient boost works

Height (m)	Favorite 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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

How gradient boost works

Height (m)	Favorite 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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

The error is the of ”prediction - observation”:

- The prediction is always 71.2
- The observation from the training data is 88, 76, …, 57

How gradient boost works

Height (m)	Favorite 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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

The error is the of ”prediction - observation”:

- The prediction is always 71.2
- The observation from the training data is 88, 76, …, 57

Err = 88 - 71.2

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

88 - 71.2

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

The error is the of ”prediction - observation”:

- The prediction is always 71.2
- The observation from the training data is 88, 76, …, 57

Each of the error can form a new column called “pseudo residual” or “residual”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

The error is the of ”prediction - observation”:

- The prediction is always 71.2
- The observation from the training data is 88, 76, …, 57

Each of the error can form a new column called “pseudo residual” or “residual”

The purpose of Gradient boost is to produce a tree predicting the smallest “Residual”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

The error is the of ”prediction - observation”:

- The prediction is always 71.2
- The observation from the training data is 88, 76, …, 57

Each of the error can form a new column called “pseudo residual” or “residual”

The purpose of Gradient boost is to produce a tree predicting the smallest “Residual”

Now we will create a tree using “Height”, “Color”, “Gender” to predict “Residual”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

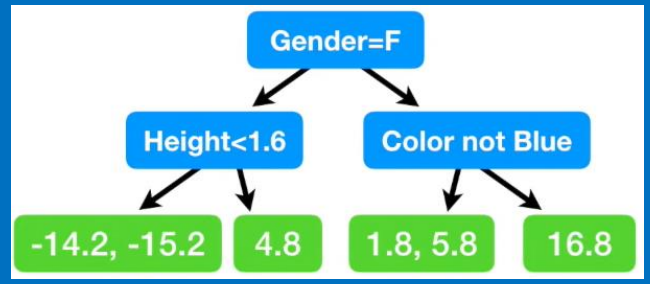
Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”

Following the regular process of creating a decision tree (we only allow 4 leaves in this example), we can have the tree like:



How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

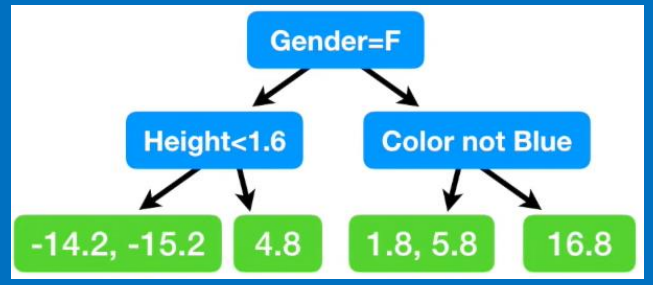
Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”

Following the regular process of creating a decision tree (we only allow 4 leaves in this example), we can have the tree like:



⬅ As a result, we get less leaves (4) than residuals (6)

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

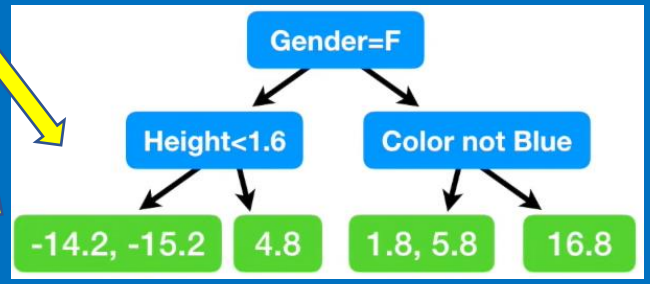
Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”

Following the regular process of creating a decision tree (we only allow 4 leaves in this example), we can have the tree like:



As a result, we get less leaves (4) than residuals (6)

For example, we have two samples end up at the same leaf

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

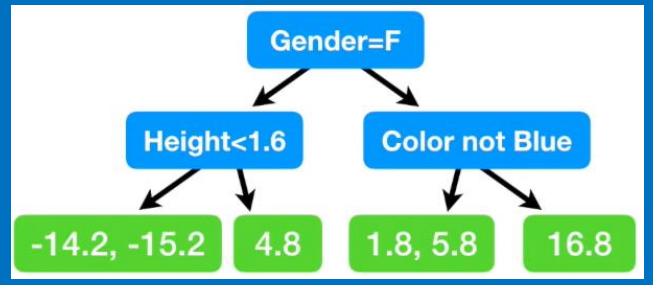
Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”

Following the regular process of creating a decision tree (we only allow 4 leaves in this example), we can have the tree like:



⬅ As a result, we get less leaves (4) than residuals (6)

All we need to do is just to replace these residuals with the “mean”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

Assuming that if we have the above dataset

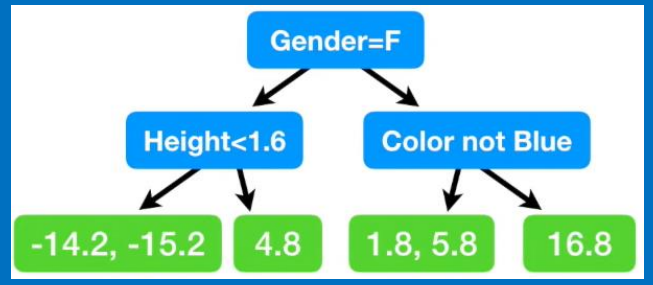
Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”

Following the regular process of creating a decision tree (we only allow 4 leaves in this example), we can have the tree like:



As a result, we get less leaves (4) than residuals (6)

All we need to do is just to replace these residuals with the mean

-14.7

3.8

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

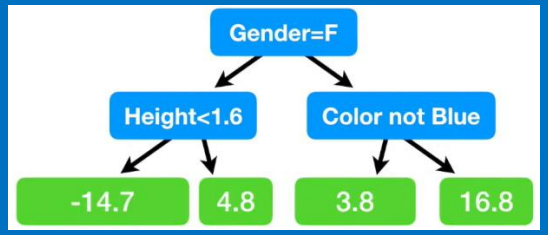
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

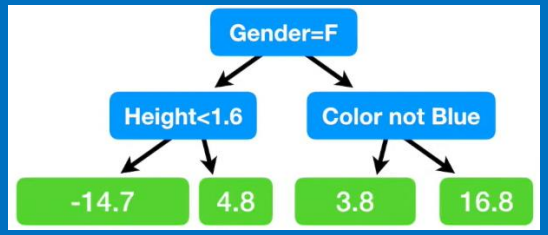
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

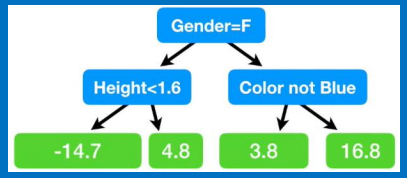
Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 +



How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	[illegible]	4.8
1.5	Blue	Female	56	[illegible]
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

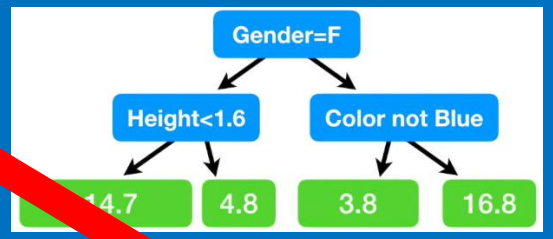
Assuming that if we have the above dataset

Step 1: calculate the original leaf

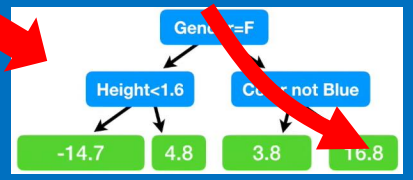
First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 +  = 71.2 + 16.8 = 88

Taking the first sample as an example, we can have the above prediction

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	[illegible]	4.8
1.5	Blue	Female	56	[illegible]
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

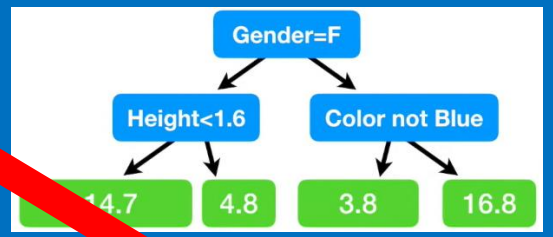
Assuming that if we have the above dataset

Step 1: calculate the original leaf

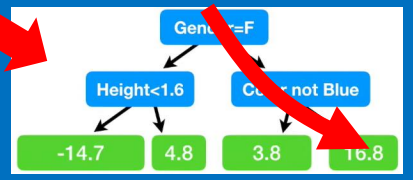
First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 +  = 71.2 + 16.8 = 88

Taking the first sample as an example, we can have the above prediction

So the prediction is exactly the same to the actual value

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	[illegible]	4.8
1.5	Blue	Female	56	[illegible]
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

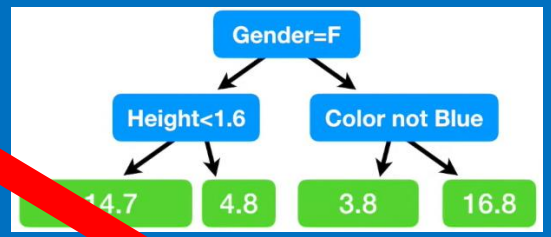
Assuming that if we have the above dataset

Step 1: calculate the original leaf

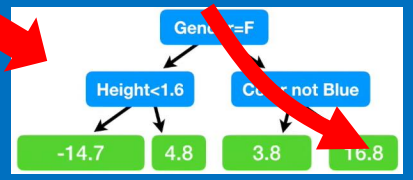
First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + 

= 71.2 + 16.8 = 88

Taking the first sample as an example, we can have the above prediction

So the prediction is exactly the same to the actual value

This is not too good ~ the model fits too well usually means overfitting (low bias but large variance)

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	[illegible]	4.8
1.5	Blue	Female	56	[illegible] 5.2
1.8	Red	Male	73	1.8
1.5	Green	Male	77	5.8
1.4	Blue	Female	57	-14.2

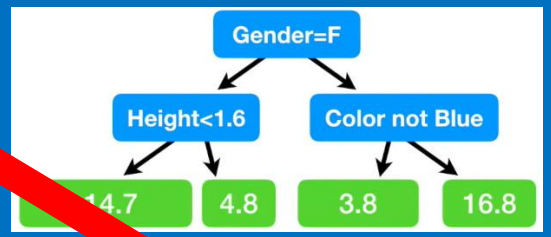
Assuming that if we have the above dataset

Step 1: calculate the original leaf

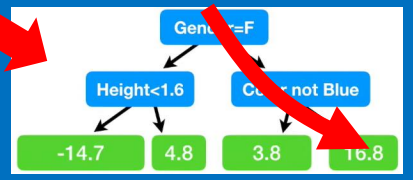
First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + 

= 71.2 + 16.8 = 88

Taking the first sample as an example, we can have the above prediction

So the prediction is exactly the same to the actual value

This is not too good ~ the model fits too well usually means overfitting (low bias but large variance)

Gradient Boost deals with this issue by introducing the concept of “learning rate”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

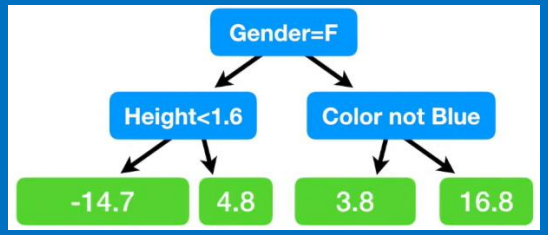
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

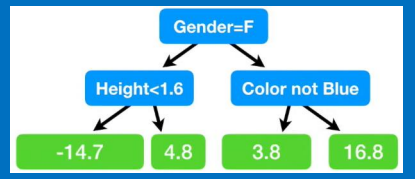
Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + Learning rate X



To scale the contribution from the tree (the rate is between 0 and 1)

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

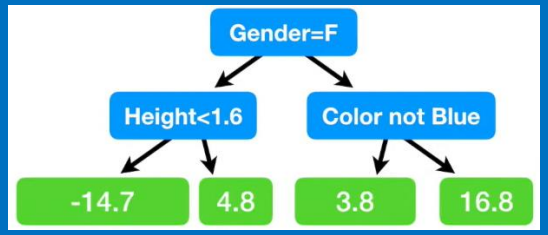
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

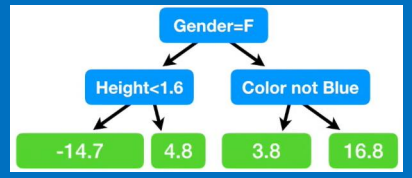
Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + 0.1 X



In this case, we assume the rate = 0.1

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

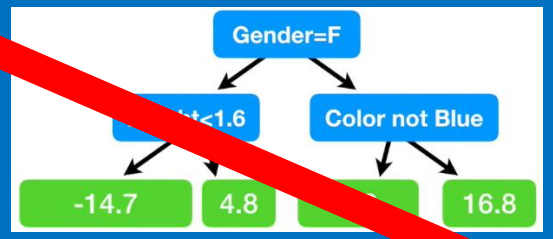
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + 0.1 X

= 71.2 + 0.1*16.8 = 72.9

In this case, we assume the rate = 0.1

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual
1.6	Blue	Male	88	16.8
1.6	Green	Female	76	4.8
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

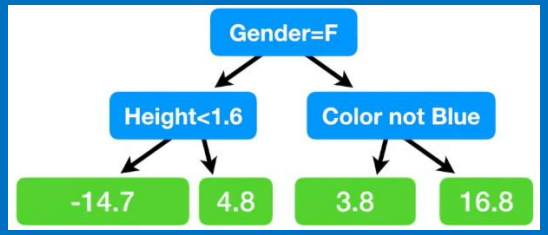
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

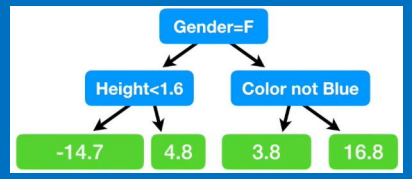
Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + 0.1 X



= 71.2 + 0.1*16.8 = 72.9

In this case, we assume the rate = 0.1

Compared to the actual data for the first sample, the prediction (72.9) is better than the original leaf (71.2)

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	
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	

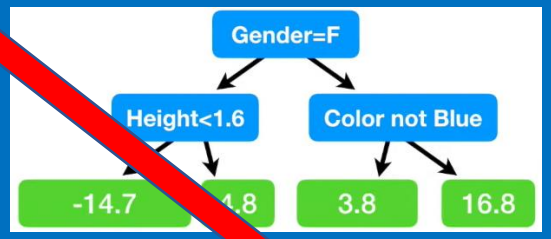
Assuming that if we have the above dataset

Step 1: calculate the original leaf

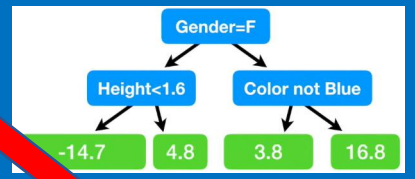
First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

71.2 + 0.1 ×  = 72.9

In this case, we assume the rate = 0.1

The new residual will be: 88 – 72.9 = 15.1

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	4.3
1.5	Blue	Female	56	-15.2	-13.7
1.8	Red	Male	73	1.8	1.4
1.5	Green	Male	77	5.8	5.4
1.4	Blue	Female	57	-14.2	-12.7

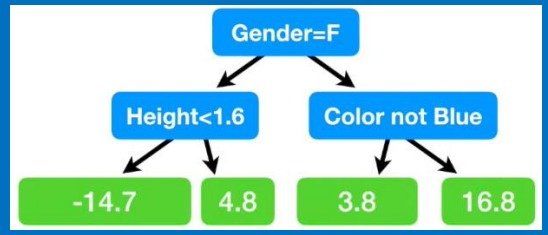
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

We go through all the samples then we can get a new set of residuals

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	4.3
1.5	Blue	Female	56	-15.2	-13.7
1.8	Red	Male	73	1.8	1.4
1.5	Green	Male	77	5.8	5.4
1.4	Blue	Female	57	-14.2	-12.7

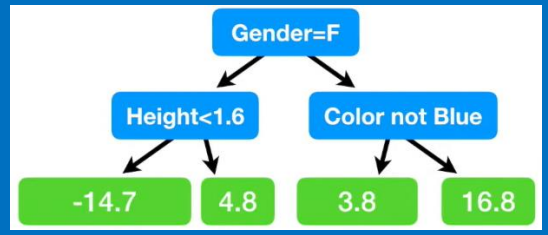
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

We go through all the samples then we can get a new set of residuals

Apparently after the first tree, the residual/error is getting smaller ~ so we are making a small step towards the right direction

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	4.3
1.5	Blue	Female	56	-15.2	-13.7
1.8	Red	Male	73	1.8	1.4
1.5	Green	Male	77	5.8	5.4
1.4	Blue	Female	57	-14.2	-12.7

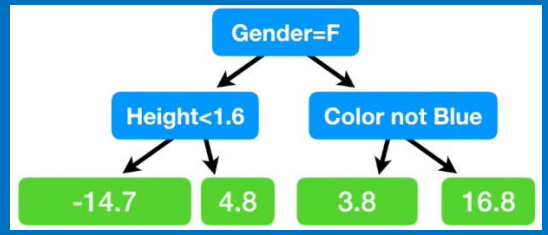
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

Step 5: create a new tree with the new “Residual”

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	4.3
1.5	Blue	Female	56	-15.2	-13.7
1.8	Red	Male	73	1.8	1.4
1.5	Green	Male	77	5.8	5.4
1.4	Blue	Female	57	-14.2	-12.7

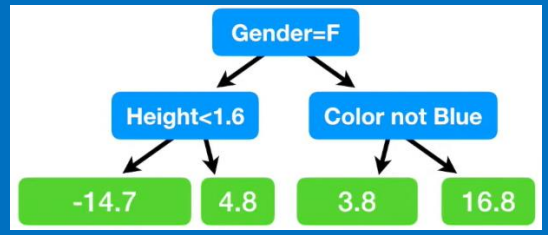
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

Step 5: create a new tree with the new “Residual”

……

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	4.3
1.5	Blue	Female	56	-15.2	-13.7
1.8	Red	Male	73	1.8	1.4
1.5	Green	Male	77	5.8	5.4
1.4	Blue	Female	57	-14.2	-12.7

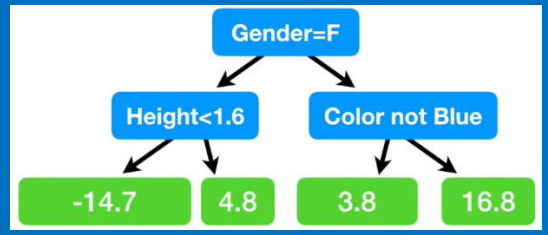
Assuming that if we have the above dataset

Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

Step 5: create a new tree with the new “Residual”

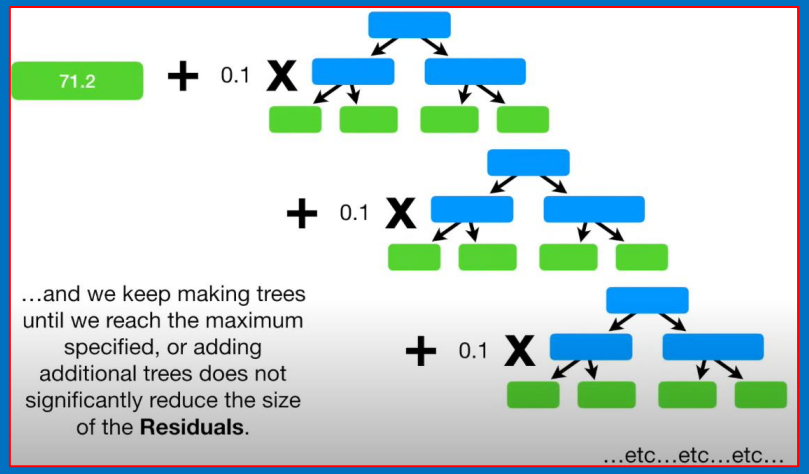
……

We do these again and again until the residual does not change much ….

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)	Residual	Residual
1.6	Blue	Male	88	16.8	15.1
1.6	Green	Female	76	4.8	4.3
1.5	Blue	Female	56	-15.2	-13.7
1.8	Red	Male	73	1.8	1.4
1.5	Green	Male	77	5.8	5.4
1.4	Blue	Female	57	-14.2	-12.7

Assuming that if we have the above dataset

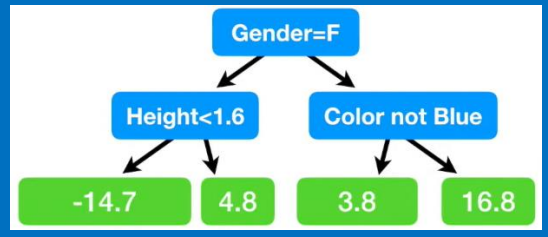


Step 1: calculate the original leaf

First leaf ➡ 71.2

Step 2: Create error of original leaf or “pseudo residual”

Step 3: Create a tree using “Height”, “Color”, “Gender” to predict “Residual”



Step 4: combine the original leaf with the 1st tree

Step 5: create a new tree with the new “Residual”

……

We do these again and again until the residual does not change much ….

How gradient boost works

Height (m)	Favorite Color	Gender	Weight (kg)
1.7	Green	Female	???

After we get the test dataset, we can just walk through the original leaf + trained scaled trees, and get the prediction, e.g.,

71.2 + 0.1xTree1 + 0.1xTree2 + …