Gradient boost (classification)

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

If we have the above dataset, and we want to predict if a person love to watch "Troll2" or not

Step 1: we start from a leaf (the "initial prediction" for all samples)

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

In gradient boosting for classification, the original leaf is the "log(odds)" of "Yes"

Likes Popcorn	Age	Favorite Color	Loves Troll 2	
Yes	12	Blue	Yes	
Yes	87	Green	Yes	
No	44	Blue	No	
Yes	19	Red	No	
No	32	Green	Yes	
No	14	Blue	Yes	

Step 1: we start from a leaf (the "initial prediction" for all samples)

In gradient boosting for classification, the original leaf is the "log(odds)" of "Yes"

- 4 people in the dataset "Loves Irolls -- True"

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

In gradient boosting for classification, the original leaf is the "log(odds)" of "Yes"

- 4 people in the dataset "Loves Trolls -- True"
- 2 people in the dataset "Loves Trolls == False"

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

In gradient boosting for classification, the original leaf is the "log(odds)" of "Yes"

- 4 people in the dataset "Loves Trolls -- True"
- 2 people in the dataset "Loves Trolls == False"

Then " $\log(\text{odds})$ " is $\log(4/2) = 0.69$

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

In gradient boosting for classification, the original leaf is the "log(odds)"

- 4 people in the dataset "Loves Trolls -- True"
- 2 people in the dataset "Loves Trolls == False"

Then " $\log(\text{odds})$ " is $\log(4/2) = 0.69$

So the first/original leaf is Log(odds) = 0.69

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

The probability is estimated by the "Logistic Function", which is

$$Probability_{Yes} = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$

Likes Popcorn	Age	Favorite Color	Loves Troll 2	
Yes	12	Blue	Yes	
Yes	87	Green	Yes	1
No	44	Blue	No	
Yes	19	Red	No	
No	32	Green	Yes	
No	14	Blue	Yes	/

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

The probability is estimated by the "Logistic Function", which is

$$Probability_{Yes} = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} = \frac{e^{0.69}}{1 + e^{0.69}} = 0.67$$

So we got 0.67 as the probability of "loving Troll2" (or "Yes")

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is

Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification
The probability is estimated by the "Logistic Function", which is

Probability_{Yes} =
$$\frac{e^{\log(odds)}}{1 + e^{\log(odds)}} = \frac{e^{0.69}}{1 + e^{0.69}} = 0.67$$

So we got 0.67 as the probability of "loving Troll2" (or "Yes")

Likes **Favorite** Loves Age Troll 2 **Popcorn** Color Yes 12 Blue Yes Yes 87 Green Yes 44 Blue No No 19 Red No Yes 32 Yes No Green No 14 Blue

Normally, the probability of "Yes" can be calculated as 4/6 = 0.67

The probability of "Yes" Step 1: we start from a leaf (the "initial prediction" for all samples)

(initial guess)

			2		
Likes Popcorn	Age	Favorite Color	Loves Troll 2	Prob (Yes)	9
Yes	12	Blue	Yes		
Yes	87	Green	Yes		
No	44	Blue	No	0.67	
Yes	19	Red	No	<u> </u>	
No	32	Green	Yes		
No	14	Blue	Yes		

Step 2: Create "probability" and use it to do the classification

The probability is estimated by the "Logistic Function", which is

$$Probability_{Yes} = \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} = \frac{e^{0.69}}{1 + e^{0.69}} = 0.67$$

So we got 0.67 as the probability of "loving Troll2" (or "Yes")

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Likes Popcorn	Age	Favorite Color	Loves Troll 2	
Yes	12	Blue	Yes	
Yes	87	Green	Yes	100%
No	44	Blue	No	
Yes	19	Red	No	
No	32	Green	Yes	
No	14	Blue	Yes	

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

For example for the first sample

People loves Troll2 from the observation, so the observation is $\frac{100\%}{100\%}$ And the predicted probability is $\frac{0.67}{100\%}$ Therefore the residual is:

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(c

Log(odds) = 0.69

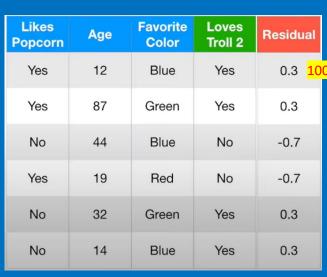
0.3 100% - 67% = 33% of 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

For example for the first sample

People loves Troll2 from the observation, so the observation is $\frac{100\%}{100\%}$ And the predicted probability is $\frac{0.67}{100\%}$ Therefore the residual is:

By doing this over all samples, we can have the Residuals column as the left



Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is

Log(odds) = 0.69

Likes Popcorn	Age	Favorite Color	Loves Troll 2	Residual
Yes	12	Blue	Yes	0.3 <mark>10</mark> 0
Yes	87	Green	Yes	0.3
No	44	Blue	No	-0.7
Yes	19	Red	No	-0.7
No	32	Green	Yes	0.3
No	14	Blue	Yes	0.3

Step 2: Create "probability" and use it to do the classification

^{%-67%=33%}t 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

For example for the first sample

People loves Troll2 from the observation, so the observation is $\frac{100\%}{100\%}$ And the predicted probability is $\frac{0.67}{100\%}$ Therefore the residual is:

By doing this over all samples, we can have the Residuals column as the left

The purpose of gradient boosting is to grow trees that gives the smallest "Residuals"

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is	Log(odds) = 0.69
----------------------------	------------------

Likes Popcorn	Age	Favorite Color	Loves Troll 2	Residual	
Yes	12	Blue	Yes	0.3	
Yes	87	Green	Yes	0.3	
No	44	Blue	No	-0.7	
Yes	19	Red	No	-0.7	
No	32	Green	Yes	0.3	

Blue

Yes

No

14

0.3

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual

Age

12

87

44

19

32

14

Favorite

Color

Blue

Green

Blue

Red

Green

Blue

Loves

Troll 2

Yes

Yes

No

No

Yes

Yes

Residual

0.3

0.3

-0.7

-0.7

0.3

0.3

Likes

Popcorn

Yes

Yes

No

Yes

No

No

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual



Following the regular tree building process, we can have the tree as left

Age

12

87

44

19

32

14

Favorite

Color

Blue

Green

Blue

Red

Green

Blue

Loves

Troll 2

Yes

Yes

No

No

Yes

Yes

Residual

0.3

0.3

-0.7

-0.7

0.3

0.3

Likes

Popcorn

Yes

Yes

No

Yes

No

No

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/	original or	leaf is	Lc
1116 11136	or ignitur	icai is	L

Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual



Following the regular tree building process, we can have the tree as left

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Age

12

87

44

19

32

14

Favorite

Color

Blue

Green

Blue

Red

Green

Blue

Loves

Troll 2

Yes

Yes

No

No

Yes

Yes

Residual

0.3

0.3

-0.7

-0.7

0.3

0.3

Likes

Popcorn

Yes

Yes

No

Yes

No

No

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/	original	leat is	(
1110 11134	or ignian	icai is	

Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual



Following the regular tree building process, we can have the tree as left

Step 5: Combine the ree and the first/original leaf to predict "log(odds)"

Let's look at the example for the 2nd and 3rd sample, which go through the three like above

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is	Log(odds) = 0.69
----------------------------	------------------

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual



Following the regular tree building process, we can have the tree as left

Sep 5: Combine the tree and the first/original leaf to predict "log(odds)"

Let's look at the example for the 2nd and 3rd sample, which go through the three like above

The "log(odds)" is calculated based on: $\sum_{i=1}^{n} \text{Residual}_{i}$ The "log(odds)" is calculated based on:

Likes **Favorite** Loves Age Residual Troll 2 Color **Popcorn** Yes 12 Blue Yes 0.3 87 Yes 0.3 Yes Green -0.7 No 44 Blue No 19 Yes Red No -0.732 Yes No Green 0.3 0.3 No 14 Blue Yes

However, here the output for the tree is "Residual", we need to convert it to "log(odds)"

Age

12

87

44

19

32

14

Favorite

Color

Blue

Green

Blue

Red

Green

Blue

Loves

Troll 2

Yes

Yes

No

No

Yes

Yes

Residual

0.3

0.3

-0.7

-0.7

0.3

0.3

Likes

Popcorn

Yes

Yes

No

Yes

No

No

Step 1: we start from a leaf (the "initial prediction" for all samples)

The tiret	ariginal	loat ic
The first/	והווצווט	ieai is
1110 11100	0.19.11a.	

Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual



Following the regular tree building process, we can have the tree as left

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Let's look at the example for the 2nd and 3rd sample, which go through the three like above

The prediction is calculated base 0.3 + (-0.7)

$$\frac{\sum \mathsf{Residual}_i}{\sum \left[\mathsf{Previous Probability}_i \times (1 - \mathsf{Previous Probability}_i)\right]}}$$

Age

12

87

44

19

32

14

Favorite

Color

Blue

Green

Blue

Red

Green

Blue

Loves

Troll 2

Yes

Yes

No

No

Yes

Yes

Residual

0.3

0.3

-0.7

-0.7

0.3

0.3

Likes

Popcorn

Yes

Yes

No

Yes

No

No

Step 1: we start from a leaf (the "initial prediction" for all samples)

	The first/original leaf is	Log(odds) = 0.6
--	----------------------------	-----------------

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual



Following the regular tree building process, we can have the tree as left

Step 5: Combist the tree and the first/original leaf to predict "log(odds)"

Let's loc $^{-1.0}$ ne example for the 2nd and 3rd sample, which go through the three like above

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is

Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

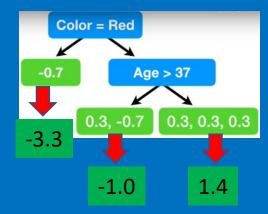
We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Similarly, we can postprocess the tree and get the log(odds) prediction as below





Age

12

87

44

19

32

14

Favorite

Color

Blue

Green

Blue

Red

Green

Blue

Loves

Troll 2

Yes

Yes

No

No

Yes

Yes

Likes

Popcorn

Yes

Yes

No

Yes

No

No

Step 1: we start from a leaf (the "initial prediction" for all samples)

ine	
Step 2	Residual
We	0.3
Step 3	0.3
otep s	-0.7

-0.7

0.3

0.3

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Color = Red



Step 1: we start from a leaf (the "initial prediction" for all samples)

Likes **Favorite** Loves Residual Age Troll 2 Yes 12 Blue Yes 0.3 Green Yes 0.3 Yes -0.7 No No 44 Blue 19 No -0.7 Yes Red 32 Yes 0.3 No Green

Blue

Yes

No

14

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residual

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"



So for the first sample, we can have the predicted Log(odds) as $0.69 + 0.8 \times 1.4 = 1.8$

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is

Log(odds) = 0.69

Likes Poposiii	Age	Favorite Color	Loves Troll 2	Residual
Yes	12	Blue	Yes	0.3
Yes	87	Green	Yes	0.3
No	44	Blue	No	-0.7
Yes	19	Red	No	-0.7
No	32	Green	Yes	0.3
No	14	Blue	Yes	0.3

reate "probability" and use it to do the classification 1.8 vve got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

So for the first sample, we can have the predicted Log(odds) as $0.69 + 0.8 \times 1.4 = 1.8$

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is

Log(odds) = 0.69

Likes	Age	Favorite Color	Loves Troll 2	Residual	
Yes	12	Blue	Yes	0.3	
Yes	87	Green	Yes	0.3	
No	44	Blue	No	-0.7	ŀ
Yes	19	Red	No	-0.7	ŀ
No	32	Green	Yes	0.3	
No	14	Blue	Yes	0.3	

reate "probability" and use it to do the classification over got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

So for the first sample, we can have the predicted Log(odds) as $0.69 + 0.8 \times 1.4 = 1.8$

Following this, we can predict "log(odds)" for all other samples

Step 1: we start from a leaf (the "initial prediction" for all samples)

og(odds) = reate "probability" and use it to do the classification

The first/original leaf is Log(odds) = 0.69

Likes	Age	Favorite Color	Loves Troll 2	Residua	
Yes	12	Blue	Yes	0.3	
Yes	87	Green	Yes	0.3	
No	44	Blue	No	-0.7	
Yes	19	Red	No	-0.7	
No	32	Green	Yes	0.3	
No	14	Blue	Yes	0.3	

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

So for the first sample, we can have the predicted Log(odds) as $0.69 + 0.8 \times 1.4 = 1.8$

Following this, we can predict "log(odds)" for all other samples

Step 6: Convert the predicted "log(odds)" to probability

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Likes	Age	Favorite Color	Loves Troll 2	Residual	
Yes	12	Blue	Yes	0.3	
Yes	87	Green	Yes	0.3	
No	44	Blue	No	-0.7	:
Yes	19	Red	No	-0.7	
No	32	Green	Yes	0.3	ı
No	14	Blue	Yes	0.3	

reate "probability" and use it to do the classification vive got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculat Residual

Step 4: Build the rst tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

So for the first san ale, we can have the predicted Log(odds) as $0.69 + 0.8 \times 1.4 = 1.8$

Following this, we capredict "log(odds)" for all other samples

Step 6: Convert the predicted "log(odds)" to probability

For example, for the first stapple,

We can convert log(odds) to probability as $\frac{e^{1.8}}{1+e^{1.8}} = 0.9$

Likes Popcorn	Age	Favorite Color	Loves Troll 2	First prediction	Update prediction 1
Yes	12	Blue	Yes		0.9
Yes	87	Green	Yes		
No	44	Blue	No	0.67	
Yes	19	Red	No		
No	32	Green	Yes		
No	14	Blue	Yes		

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Step 6: Convert the predicted "log(odds)" to probability

For example, for the first sample,

We can convert log(odds) to probability as $\frac{e^{1.8}}{1+e^{1.8}}=0.9$

Likes Popcorn	Age	Favorite Color	Loves Troll 2	First prediction	Update prediction 1	
Yes	12	Blue	Yes		0.9	
Yes	87	Green	Yes		The updated	
No	44	Blue	No	0.67	prediction (0.9) is clearly better tha	n the
Yes	19	Red	No		original one (0.67)
No	32	Green	Yes			
No	14	Blue	Yes			

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Step 6: Convert the predicted "log(odds)" to probability

For example, for the first sample,

We can convert log(odds) to probability as $\frac{e^{1.8}}{1+e^{1.8}}=0.9$

Likes Popcorn	Age	Favorite Color	Loves Troll 2	First prediction	Update prediction 1
Yes	12	Blue	Yes		0.9
Yes	87	Green	Yes		0.5
No	44	Blue	No	0.67	0.5
Yes	19	Red	No		0.1
No	32	Green	Yes		0.9
No	14	Blue	Yes		0.9

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Step 6: Convert the predicted "log(odds)" to probability

For example, for the first sample,

We can convert log(odds) to probability as $\frac{e^{1.8}}{1+e^{1.8}}=0.9$

We do these for all the samples, and can get the updated predictions for all of them

Likes Popcorn	Age	Favorite Color	Loves Troll 2	First prediction	Update prediction 1
Yes	12	Blue	Yes		0.9
Yes	87	Green	Yes		0.5
No	44	Blue	No	0.67	0.5
Yes	19	Red	No		0.1
No	32	Green	Yes		0.9
No	14	Blue	Yes		0.9

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

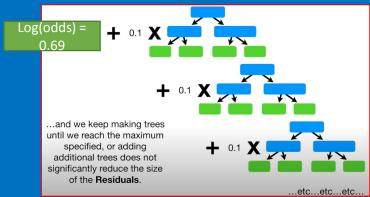
Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Step 6: Convert the predicted "log(odds)" to probability

Step 7: We then repeat Step 3-6 with the updated predictions, until the calculated residuals do not change much



Likes Popcorn	Age	Favorite Color	Loves Troll 2	First prediction	Update predictio	n 1
Yes	12	Blue	Yes		0.9	
Yes	87	Green	Yes		0.5	
No	44	Blue	No	0.67	0.5	
Yes	19	Red	No		0.1	
No	32	Green	Yes		0.9	
No	14	Blue	Yes		0.9	

In prediction, we just need to use the test data to go through all the trees and add them up, the results are predicted Log(odds), which can be converted to probability and indicate the classified category

Step 1: we start from a leaf (the "initial prediction" for all samples)

The first/original leaf is Log(odds) = 0.69

Step 2: Create "probability" and use it to do the classification

We got 0.67 as the predicted probability of "loving Troll2"

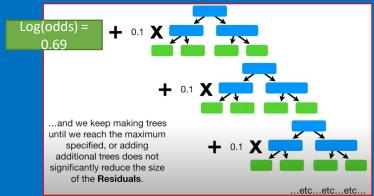
Step 3: calculate Residual

Step 4: Build the first tree and predict the Residuals

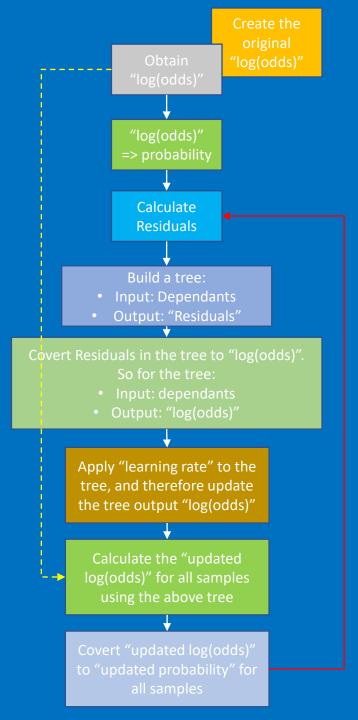
Step 5: Combine the tree and the first/original leaf to predict "log(odds)"

Step 6: Convert the predicted "log(odds)" to probability

Step 7: We then repeat Step 3-6 with the updated predictions, until the calculated residuals do not change much



Summary



For example

