Just like in **Part 1** of this series, we start with a leaf that represents an initial **Prediction** for every individual.

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

When we use **Gradient Boost for Classification**, the initial **Prediction** for every individual is the **log(odds)**.

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

I like to think of the **log(odds)** as the **Logistic Regression** equivalent of the average.

So let's calculate the overall **log(odds)** that someone **Loves Troll 2**.

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

Since **4** people in the **Training Dataset Love Troll 2**…

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

…and **2** people do not…

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

…then the **log(odds)** that someone **Loves Troll 2** is…

$$\log\left(\frac{4}{2}\right) = 0.7$$

log(4/2) = 0.7

...which we will put into our initial leaf.

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

$\log(\frac{4}{2}) = 0.7$

log(4/2) = 0.7

So this is the **Initial Prediction**.

How do we use it for **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

log(4/2) = 0.7

Just like with **Logistic Regression**, the easiest way to use the **log(odds)** for **Classification** is to convert it to a **Probability**…

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

…and we do that with a **Logistic Function**.

$$\textbf{Probability of Loving Troll 2} = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$

log(4/2) = 0.7

…and we get **0.7** as the **Probability of Loving Troll 2**.

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

$$\textbf{Probability of Loving Troll 2} = \frac{e^{\log(4/2)}}{1 + e^{\log(4/2)}} = \boxed{0.7}$$

log(4/2) = 0.7

Probability of **Loving Troll 2** = 0.7

NOTE: These two numbers, the **log(4/2)** and the **Probability** are the same only because I'm rounding. If I allowed **4** digits passed the decimal place…

$$\log\left(\frac{4}{2}\right) = 0.6931$$

$$\frac{e^{\log(4/2)}}{1 + e^{\log(4/2)}} = 0.6667$$

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

log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

Since the **Probability** of **Loving Troll 2** is greater than **0.5**, we can **Classify** everyone in the **Training Datase**t as someone who **Loves Troll 2**.

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

NOTE: While **0.5** is a very common threshold for making **Classification** decisions based on **Probability**, we could have just as easily used a different value.

For more details, check out the **StatQuest: ROC and AUC, Clearly Explained**.

log(4/2) = 0.7

Probability of **Loving Troll 2** = 0.7

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

Now, **Classifying** everyone in the **Training Dataset** as someone who **Loves Troll 2** is pretty lame, because two of the people do not love the movie.

log(4/2) = 0.7

Probability of **Loving Troll 2** = 0.7

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

We can measure how bad the initial **Prediction** is by calculating **Pseudo Residuals**, the difference between the **Observed** and the **Predicted** values.

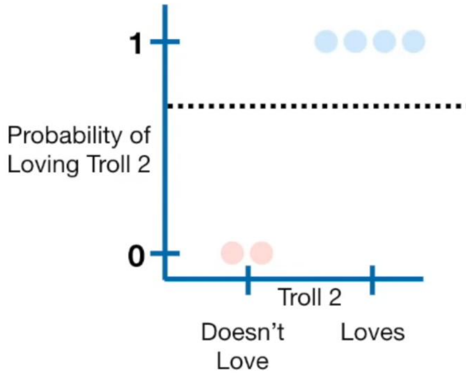
Residual = (**Observed** - **Predicted**)

log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

So for this sample…

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	$\boxed{\text{Yes}}$
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes



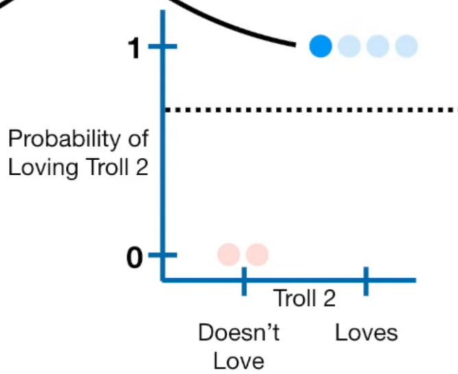
log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

…we plug in **1** for the **Observed** value…

Residual = (**1** - **Predicted**)

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	$\boxed{\text{Yes}}$
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes



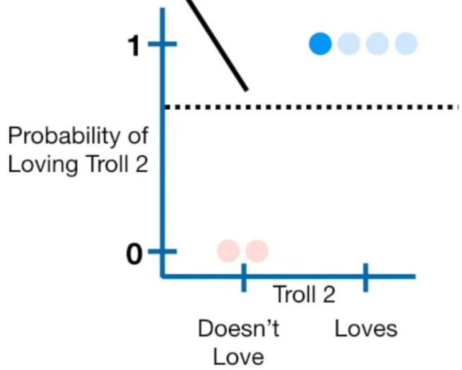
log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

…and **0.7** for the **Predicted** value…

Residual = (**1** - **0.7**)

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



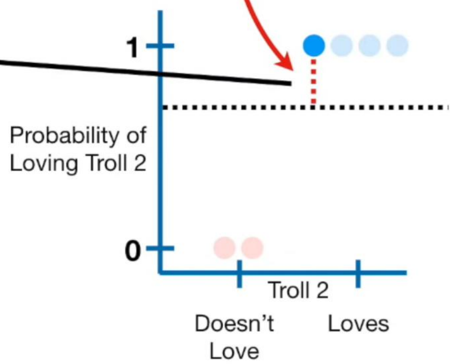
log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

…and we save the **Residual** in a new column.

Residual = (**1** - **0.7**) = 0.3

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



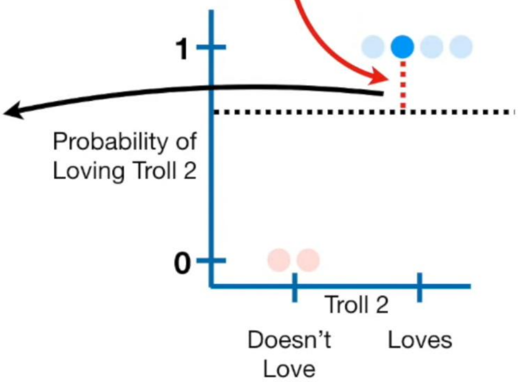
log(4/2) = 0.7

Probability of **Loving Troll 2** = 0.7

Then we calculate the rest of the **Residuals**…

Residual = (**1** - **0.7**) = 0.3

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	
Yes	19	Red	No	
No	32	Green	Yes	
No	14	Blue	Yes	



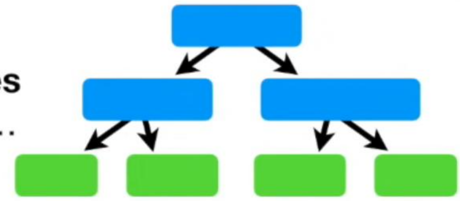
log(4/2) = 0.7

Probability of **Loving Troll 2** = 0.7

Hooray! We've calculated the **Residuals** for the leaf's initial **Prediction**.

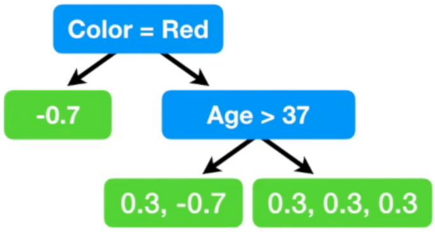
Likes Popcorn	Age	Favorite Color	Loves Troll 2	Residual
Yes	12	Blue	Yes	0.3
Yes	87	Green	Yes	0.3
No	44	Blue	No	-0.7
Yes	19	Red	No	-0.7
No	32	Green	Yes	0.3
No	14	Blue	Yes	0.3

Now we will build a **Tree**, using **Likes Popcorn**, **Age** and **Favorite Color**…



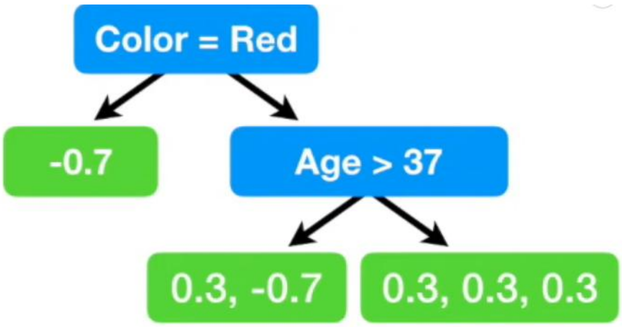
Likes Popcorn	Age	Favorite Color	Loves Troll 2	Residual
Yes	12	Blue	Yes	0.3
Yes	87	Green	Yes	0.3
No	44	Blue	No	-0.7
Yes	19	Red	No	-0.7
No	32	Green	Yes	0.3
No	14	Blue	Yes	0.3

…to **Predict** the **Residuals**.



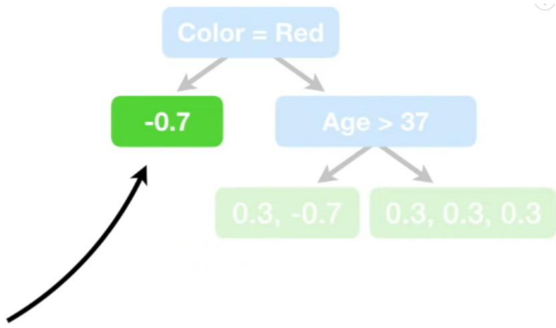
Likes Popcorn	Age	Favorite Color	Loves Troll 2	Residual
Yes	12	Blue	Yes	0.3
Yes	87	Green	Yes	0.3
No	44	Blue	No	-0.7
Yes	19	Red	No	-0.7
No	32	Green	Yes	0.3
No	14	Blue	Yes	0.3

NOTE: Just like when we used **Gradient Boost for Regression**, we are limiting the number of leaves that we will allow in the tree.



In this simple example, we are limiting the number of leaves to **3**.

In practice people often set the maximum number of leaves to be between **8** and **32**

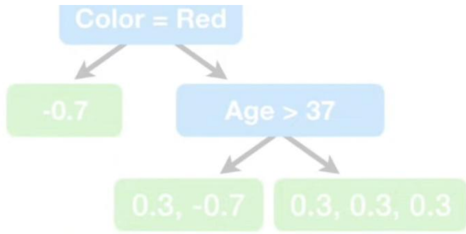


When we used **Gradient Boost** for **Regression**, a leaf with single **Residual** had an **Output Value** equal to that **Residual**.

In contrast, when we use **Gradient Boost** for **Classification**, the situation is a little more complex.

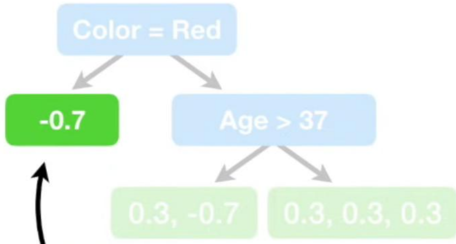
log(4/2) = 0.7

This is because the **Predictions** are in terms of the **log(odds)**…

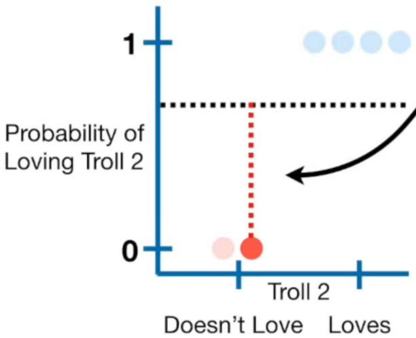


log(4/2) = 0.7

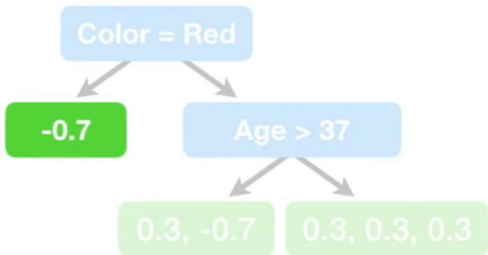
This is because the **Predictions** are in terms of the **log(odds)**…



…and this leaf is derived from a **Probability**…

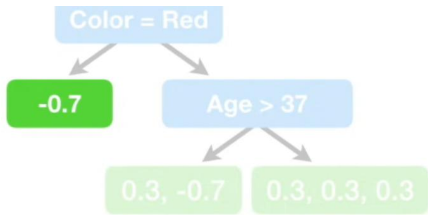


log(4/2) = 0.7



…so we can't just add them together to get a new **log(odds) Prediction** without some sort of transformation.

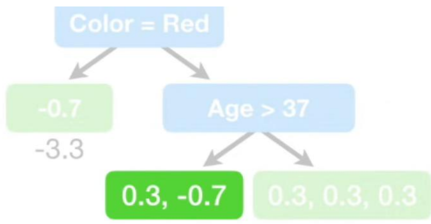
When we use **Gradient Boost** for **Classification**, the most common transformation is the following formula.



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

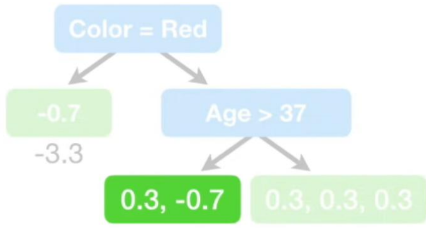
The numerator is the sum of the all of the **Residuals** in the leaf…

…and the denominator is the sum of the previously predicted probabilities for each **Residual** times **1** minus the same predicted probability.



$$\frac{0.3 + -0.7}{\sum \left[\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)\right]}$$

Since we have two **Residuals** in the leaf, we'll add them together in the numerator…



$$\frac{0.3 + -0.7}{\sum \left[\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)\right]}$$

…and in the denominator we just add up…

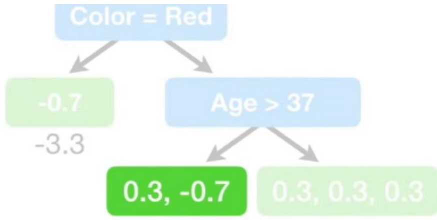
Previous Probability x (1 - Previous Probability)

…for each residual.

log(4/2) = 0.7

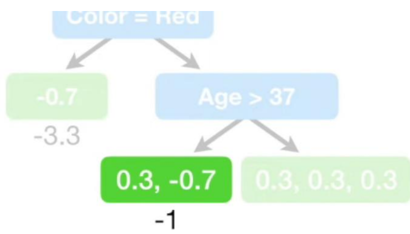
Initial Probability of **Loving Troll 2** = 0.7

$$\frac{0.3 + -0.7}{(0.7 \times (1 - 0.7)) + (0.7 \times (1 - 0.7))}$$



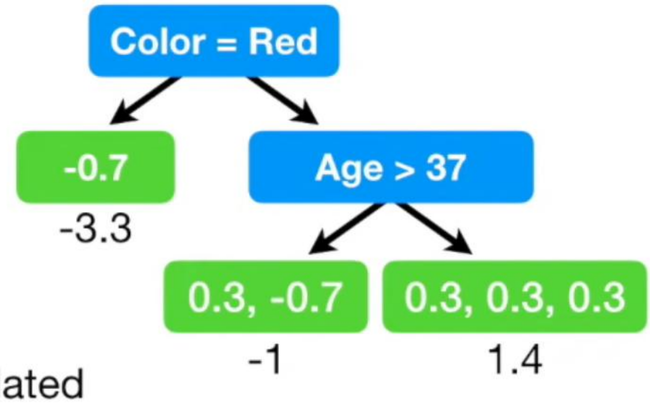
So we plug in the **Previous Probability** for each **Residual**.

NOTE: For now, the **Previous Probabilities** are the same for all of the **Residuals**, but this will change when we build the next tree.

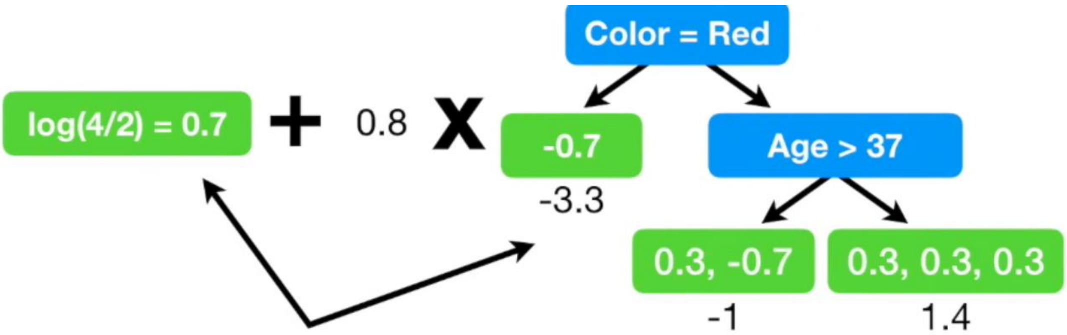


$$\frac{0.3 + -0.7}{2(0.7 \times (1 - 0.7))} = \boxed{-1}$$

…and the **Output Value** for this leaf is **-1**.



Hooray!!! We've calculated **Output Values** for all three leaves in the tree!



Now we are ready to update our **Predictions** by combining the initial leaf with the new tree.

NOTE: Just like before, the new tree is scaled by a **Learning Rate**.

This example uses a relatively large **Learning Rate** for illustrative purposes. However, **0.1** is more common.

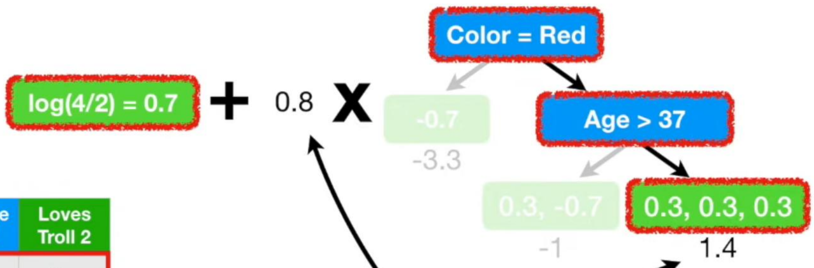
Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	12	Blue	Yes
Yes	87	Green	Yes
No	44	Blue	No
Yes	19	Red	No
No	32	Green	Yes
No	14	Blue	Yes

Now let's calculate the **log(odds) Prediction** for this person…

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

The **log(odds) Prediction** is the previous **Prediction**…

log(odds) Prediction = 0.7



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

…plus the **Output Value** from the tree scaled by the **Learning Rate**…

log(odds) Prediction = 0.7 + (0.8 × 1.4)

...and the new **log(odds) Prediction** = **1.8**.

$$\text{log(odds) Prediction} = 0.7 + (0.8 \times 1.4) = \boxed{1.8}$$

Now we convert the new **log(odds) Prediction** into a **Probability**...

$$\text{Probability} = \frac{e^{1.8}}{1 + e^{1.8}} = 0.9$$

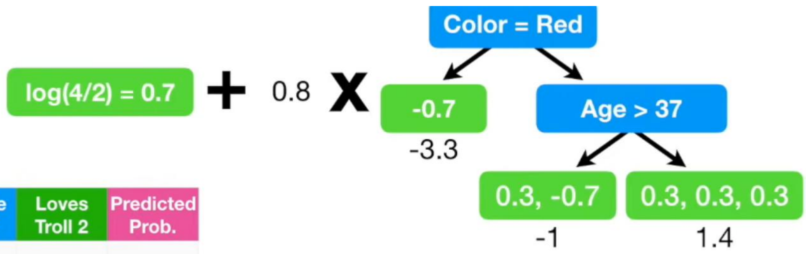
$$\text{log(odds) Prediction} = 0.7 + (0.8 \times 1.4) = 1.8$$

Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.
Yes	12	Blue	Yes	0.9
Yes	87	Green	Yes	
No	44	Blue	No	
Yes	19	Red	No	
No	32	Green	Yes	
No	14	Blue	Yes	

We save the new **Predicted Probability** here.

$$\text{Probability} = \frac{e^{1.8}}{1 + e^{1.8}} = \boxed{0.9}$$

$$\text{log(odds) Prediction} = 0.7 + (0.8 \times 1.4) = 1.8$$



Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.
Yes	12	Blue	Yes	0.9
Yes	87	Green	Yes	0.5
No	44	Blue	No	0.5
Yes	19	Red	No	0.1
No	32	Green	Yes	0.9
No	14	Blue	Yes	0.9

Then we calculate the **Predicted Probabilities** for the remaining people.

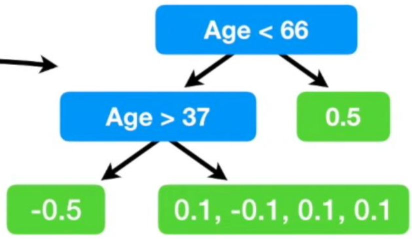
And now, just like before, we calculate the new **Residuals**…

Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Yes	12	Blue	Yes	0.9	
Yes	87	Green	Yes	0.5	
No	44	Blue	No	0.5	
Yes	19	Red	No	0.1	
No	32	Green	Yes	0.9	
No	14	Blue	Yes	0.9	

Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Yes	12	Blue	Yes	0.9	0.1
Yes	87	Green	Yes	0.5	0.5
No	44	Blue	No	0.5	-0.5
Yes	19	Red	No	0.1	-0.1
No	32	Green	Yes	0.9	0.1
No	14	Blue	Yes	0.9	0.1

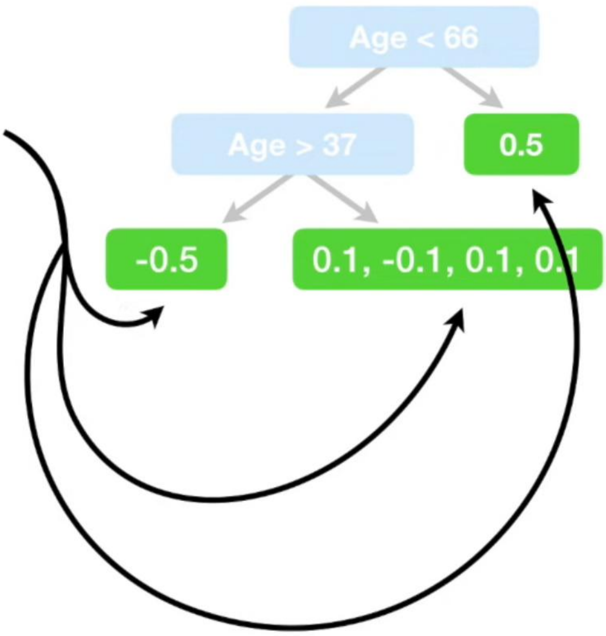
BAM!

Now that we have the **Residuals**, we can build a new tree…



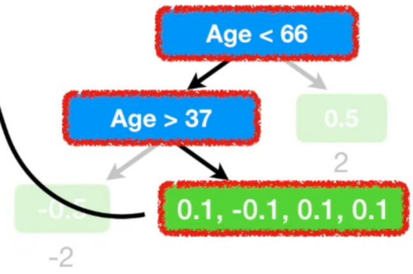
Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Yes	12	Blue	Yes	0.9	0.1
Yes	87	Green	Yes	0.5	0.5
No	44	Blue	No	0.5	-0.5
Yes	19	Red	No	0.1	-0.1
No	32	Green	Yes	0.9	0.1
No	14	Blue	Yes	0.9	0.1

…and then we need to calculate the **Output Values** for each leaf.



Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Blue	Yes	0.9	0.1
Green	Yes	0.5	0.5
Blue	No	0.5	-0.5
Red	No	0.1	-0.1
Green	Yes	0.9	0.1

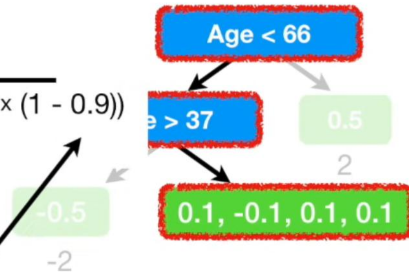
$$\frac{0.1 + -0.1 + 0.1 + 0.1}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$$



Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Yes	12	Blue	Yes	0.9	0.1
Yes	87	Green	Yes	0.5	0.5
No	44	Blue	No	0.5	-0.5
Yes	19	Red	No	0.1	-0.1
No	32	Green	Yes	0.9	0.1
No	14	Blue	Yes	0.9	0.1

So we plug the **Residuals** into the formula for the **Output Values**…

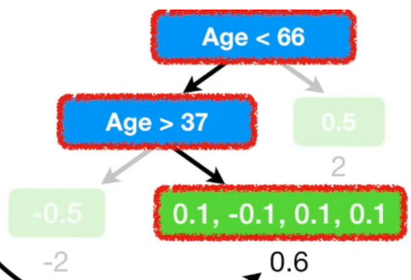
$$\frac{0.1 + -0.1 + 0.1 + 0.1}{(0.9 \times (1 - 0.9)) + (0.1 \times (1 - 0.1)) + (0.9 \times (1 - 0.9)) + (0.9 \times (1 - 0.9))}$$



Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Yes	12	Blue	Yes	0.9	0.1
Yes	87	Green	Yes	0.5	0.5
No	44	Blue	No	0.5	-0.5
Yes	19	Red	No	0.1	-0.1
No	32	Green	Yes	0.9	0.1
No	14	Blue	Yes	0.9	0.1

…and we plug in the **Predicted Probability** for each individual in the leaf…

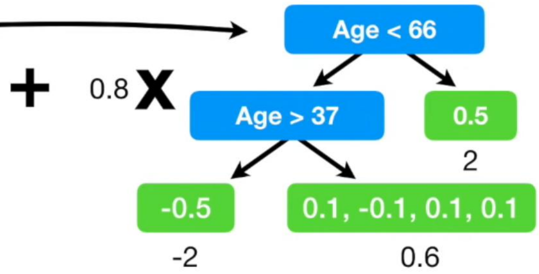
$$\frac{0.2}{0.09 + 0.09 + 0.09 + 0.09} = \boxed{0.6}$$

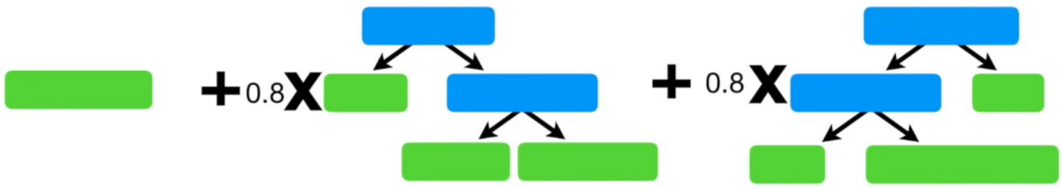


Likes Popcorn	Age	Favorite Color	Loves Troll 2	Predicted Prob.	Residual
Yes	12	Blue	Yes	0.9	0.1
Yes	87	Green	Yes	0.5	0.5
No	44	Blue	No	0.5	-0.5
Yes	19	Red	No	0.1	-0.1
No	32	Green	Yes	0.9	0.1
No	14	Blue	Yes	0.9	0.1

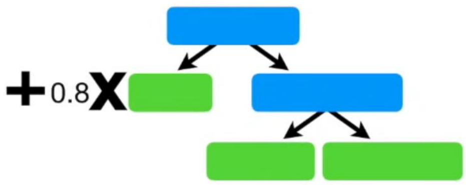
…and the **Output Value** for this leaf is **0.6**.

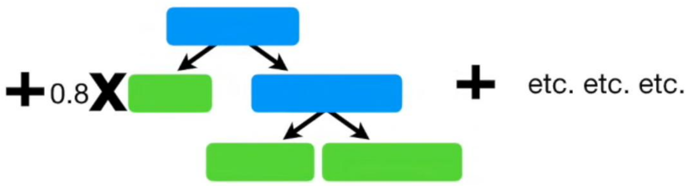
Then we built another tree based on the new **Residuals**, the difference between the **Observed** values and the values **Predicted** by the leaf ***and*** the first tree…

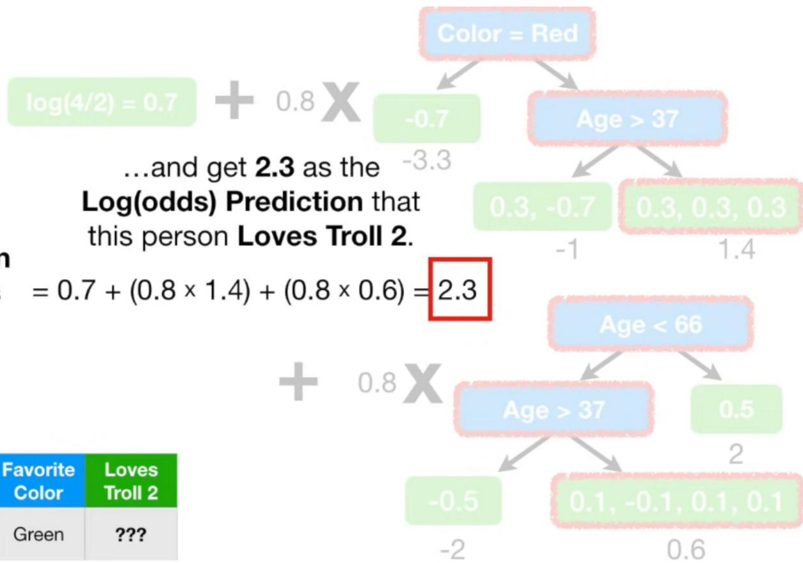




This process repeats until we have made the maximum number of trees specified, or the residuals get super small.







…and get **2.3** as the **Log(odds) Prediction** that this person **Loves Troll 2**.

Log(odds) Prediction that someone **Loves Troll 2**:

$$= 0.7 + (0.8 \times 1.4) + (0.8 \times 0.6) = \boxed{2.3}$$

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	25	Green	???

…and the **Predicted Probability** that this individual will **Love Troll 2** is **0.9**.

Log(odds) Prediction that someone **Loves Troll 2**: $= 0.7 + (0.8 \times 1.4) + (0.8 \times 0.6) = 2.3$

$$\text{Probability} = \frac{e^{2.3}}{1 + e^{2.3}} = \boxed{0.9}$$

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	25	Green	**???**

…we will **Classify** this person as someone who **Loves Troll 2**.

Log(odds) Prediction that someone **Loves Troll 2**: $= 0.7 + (0.8 \times 1.4) + (0.8 \times 0.6) = 2.3$

$$\text{Probability} = \frac{e^{2.3}}{1 + e^{2.3}} = 0.9$$

Likes Popcorn	Age	Favorite Color	Loves Troll 2
Yes	25	Green	**YES!!!**