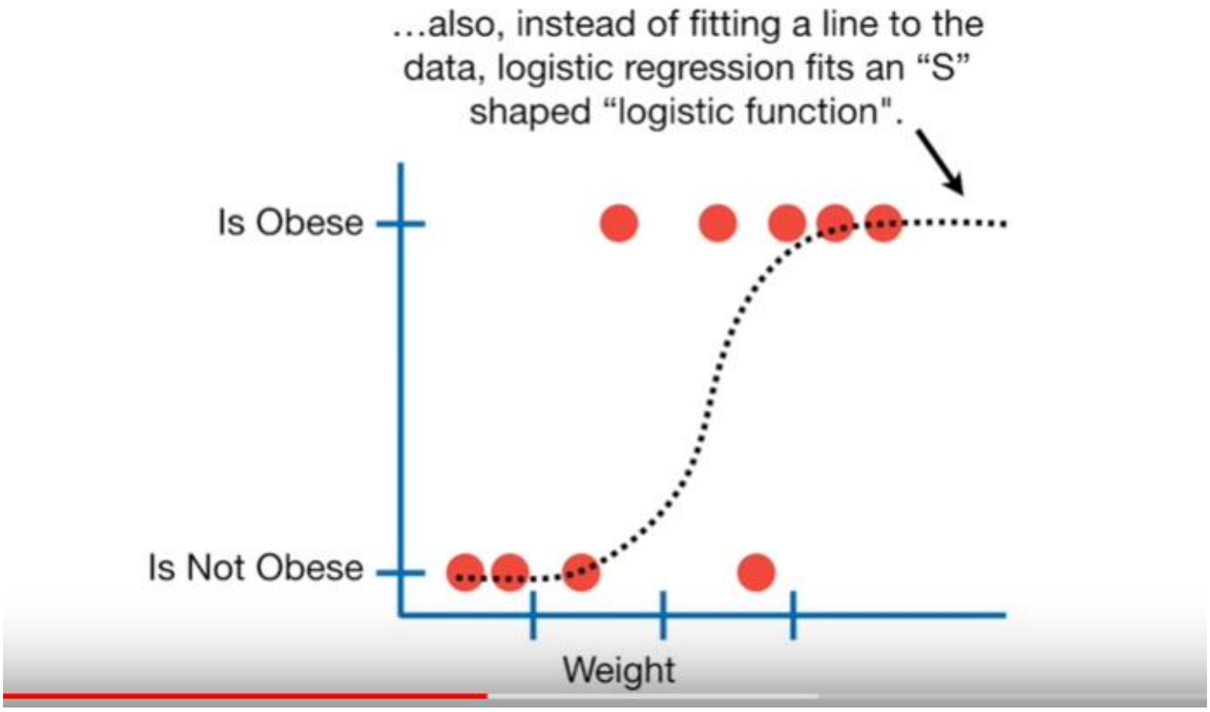
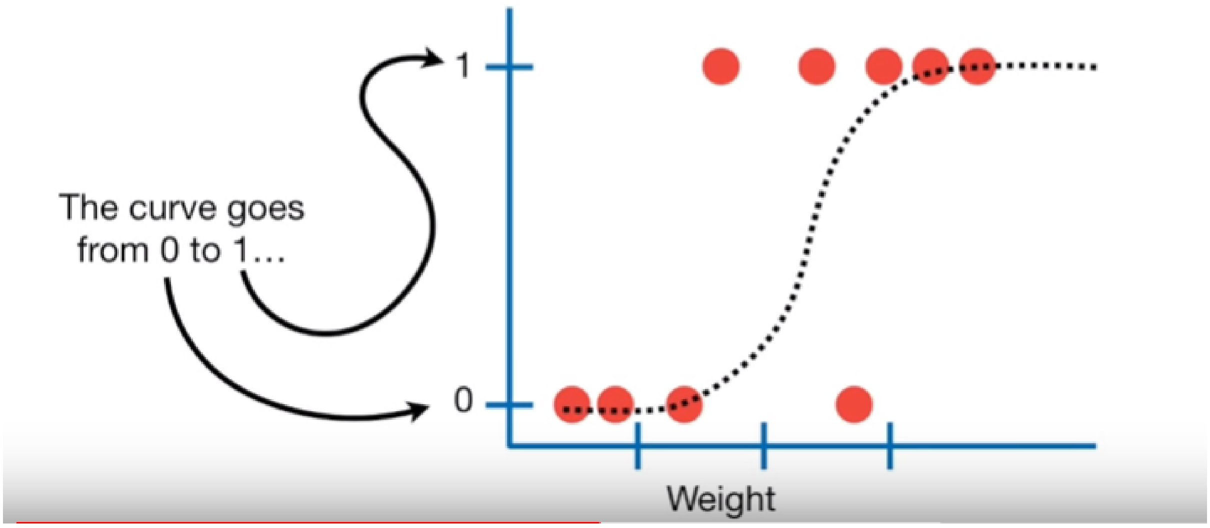
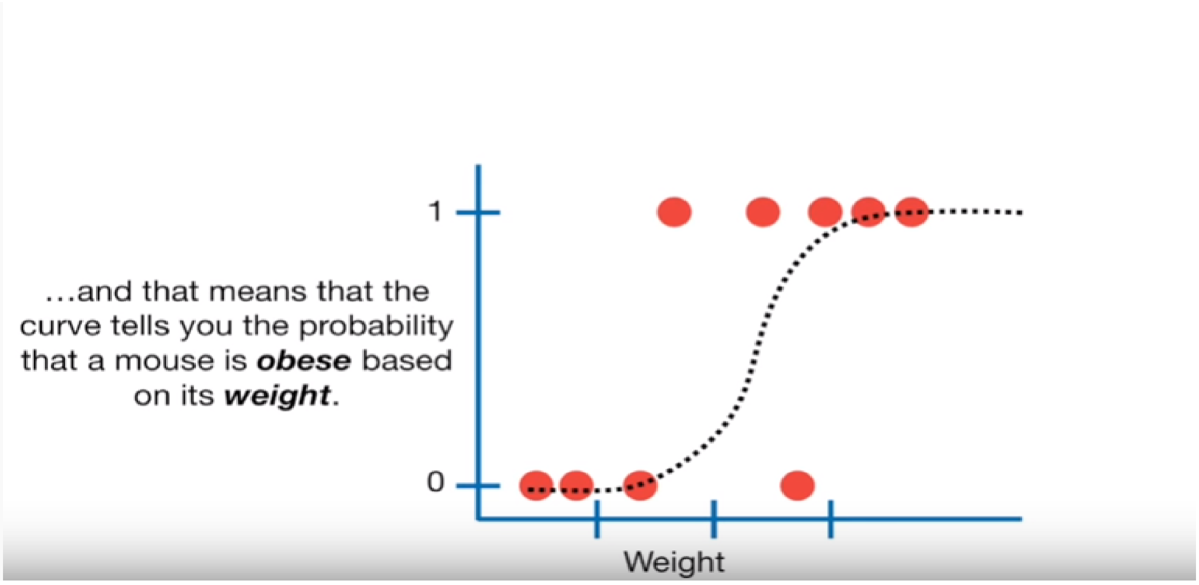
**Logistic Regression:**

Logistic regression: instead of fitting a line to data, Logistic regression fits an "S" shaped "logistic function".

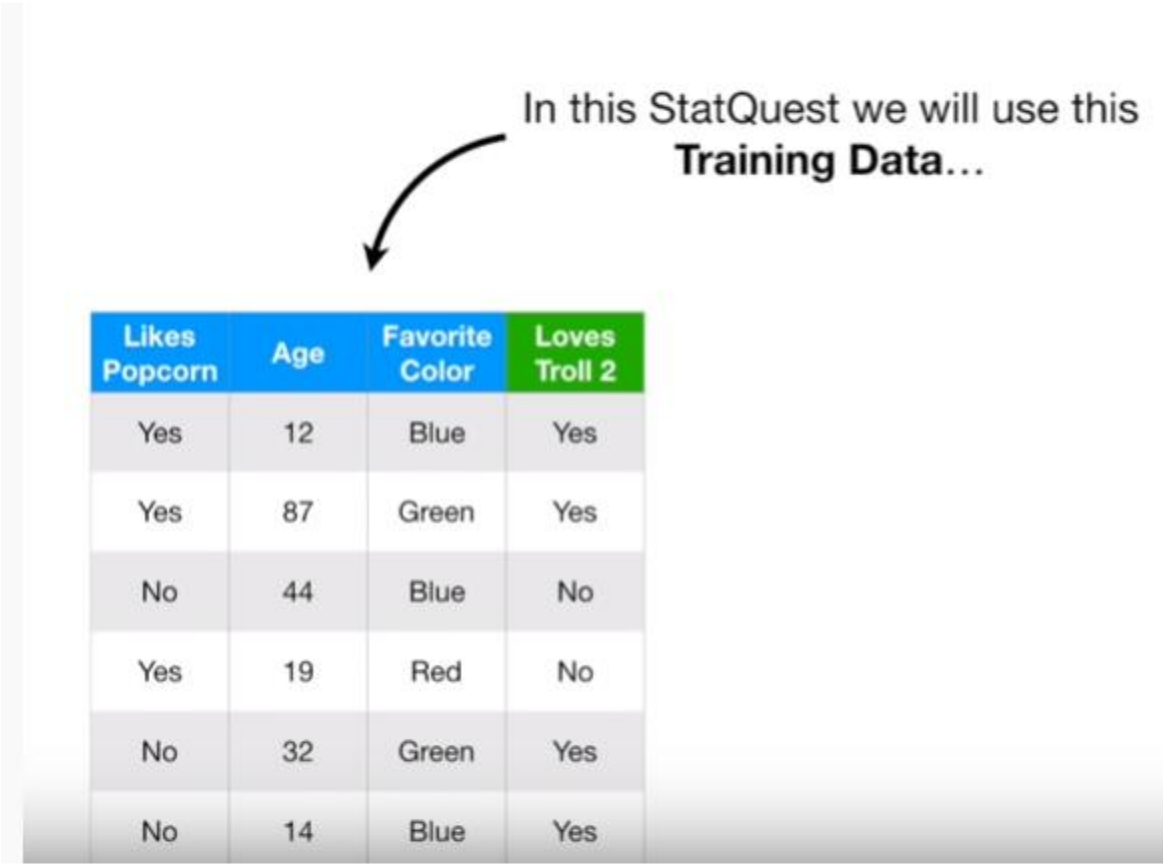




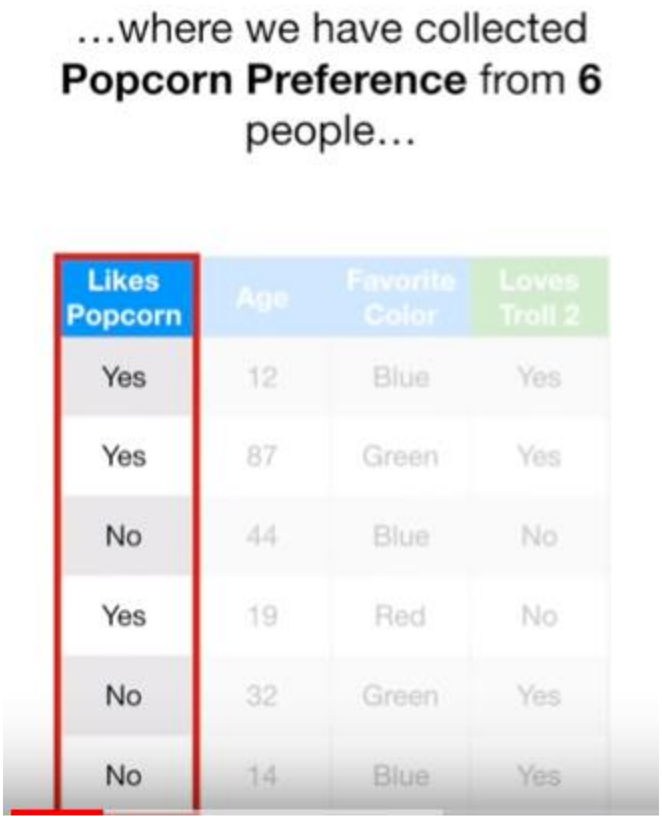

## How Gradient Boosting Classification Works

We are going to use following data:




| 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 |

Where we have collected **Popcorn preferences** from **6 People**:

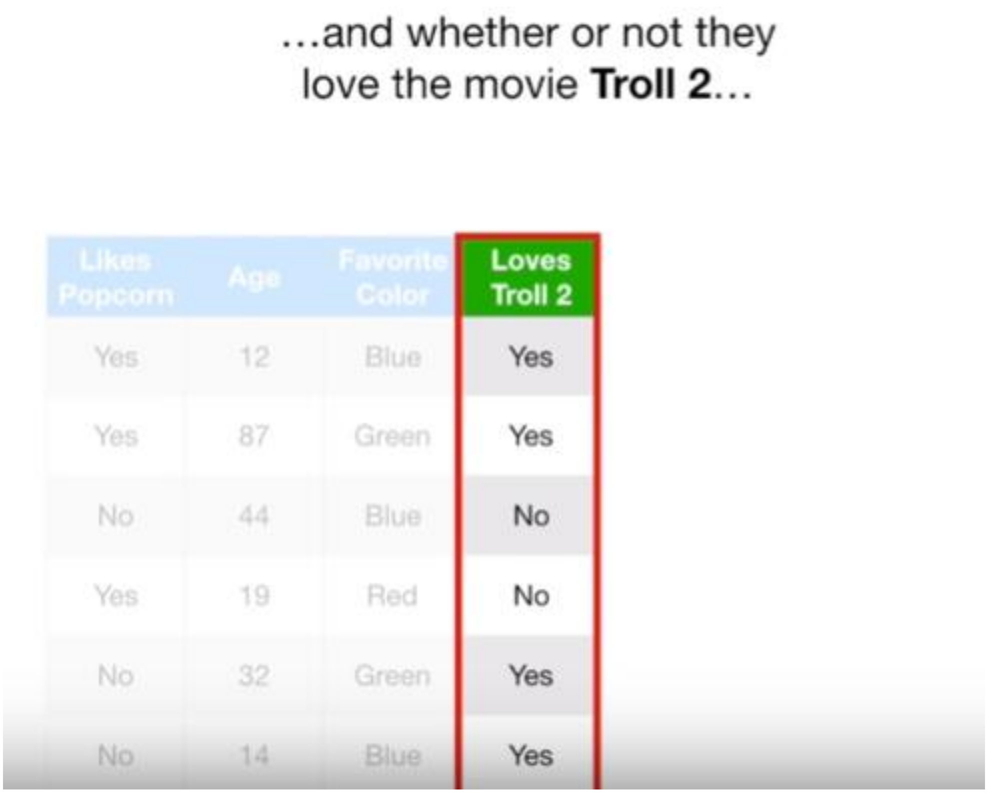



Their **Age**:

Their Favourite **Color**:

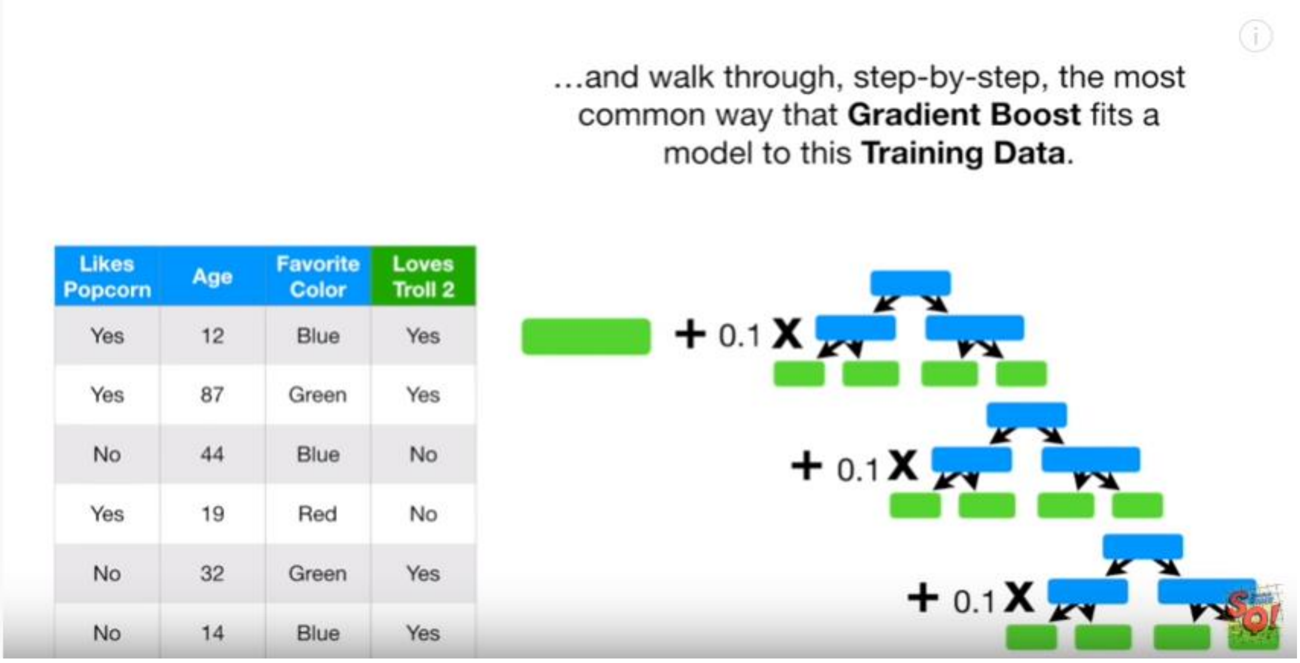

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 |
|---|---|---|---|
| Yes | 12 | Blue | Yes |
| Yes | 87 | Green | Yes |
| No | 44 | Blue | No |
| Yes | 19 | Red | No |
| No | 32 | Green | Yes |
| No | 14 | Blue | Yes |

And whether or not they love movie **Troll 2:**

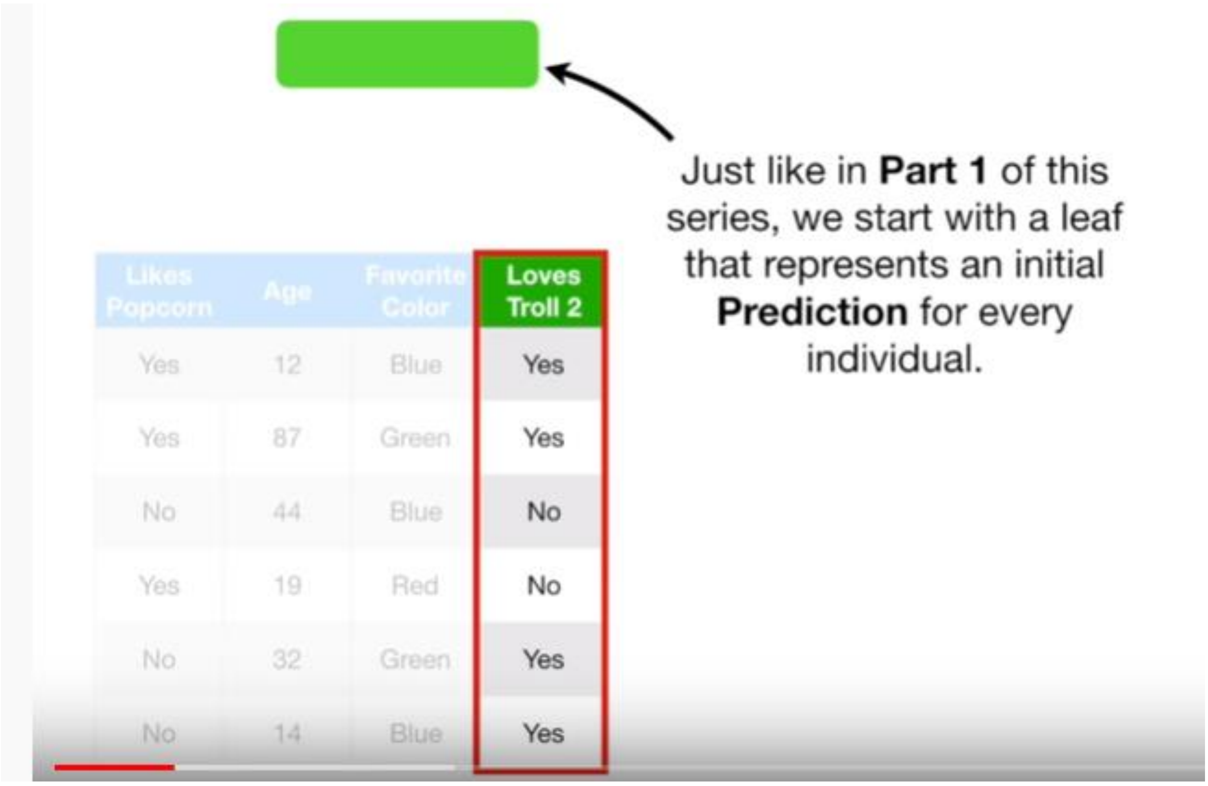



| Likes Popcorn | Age | Favorite Color | Loves Troll 2 |
|---|---|---|---|
| Yes | 12 | Blue | Yes |
| Yes | 87 | Green | Yes |
| No | 44 | Blue | No |
| Yes | 19 | Red | No |
| No | 32 | Green | Yes |
| No | 14 | Blue | Yes |

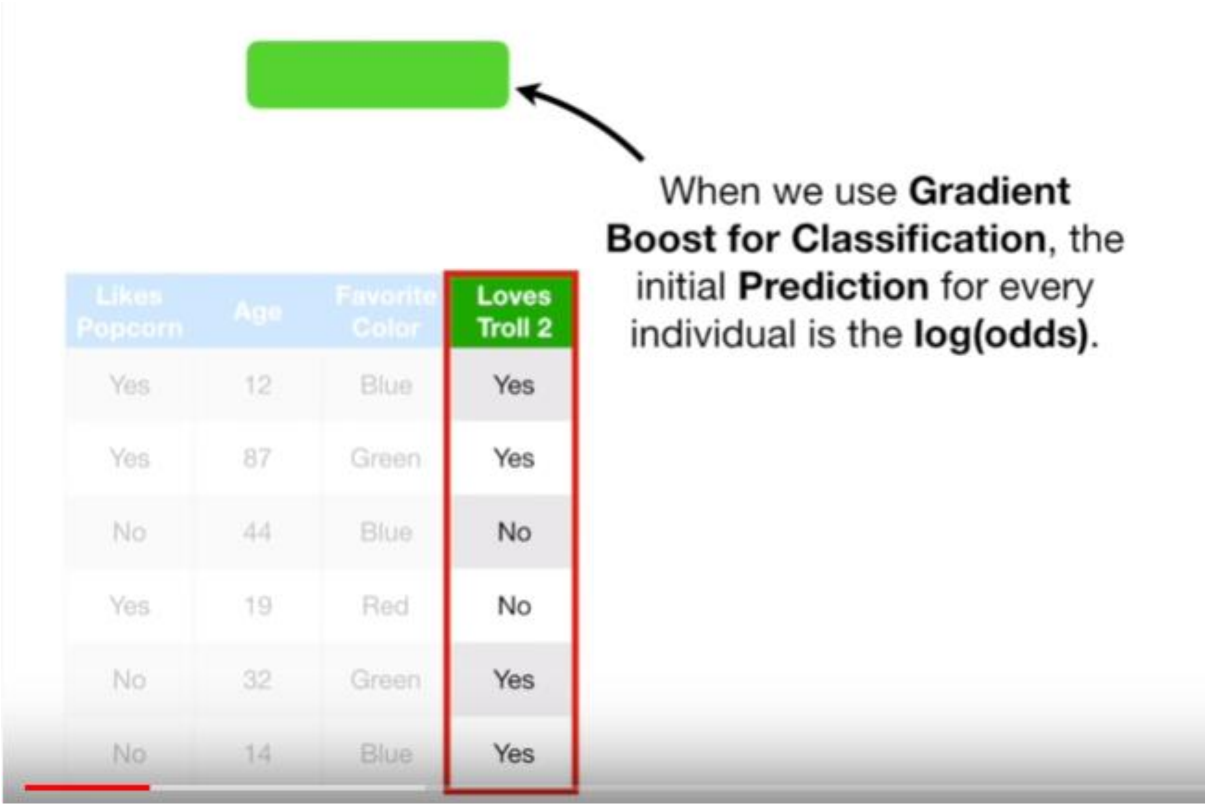
We will walk-through Step by Step to build a Model:

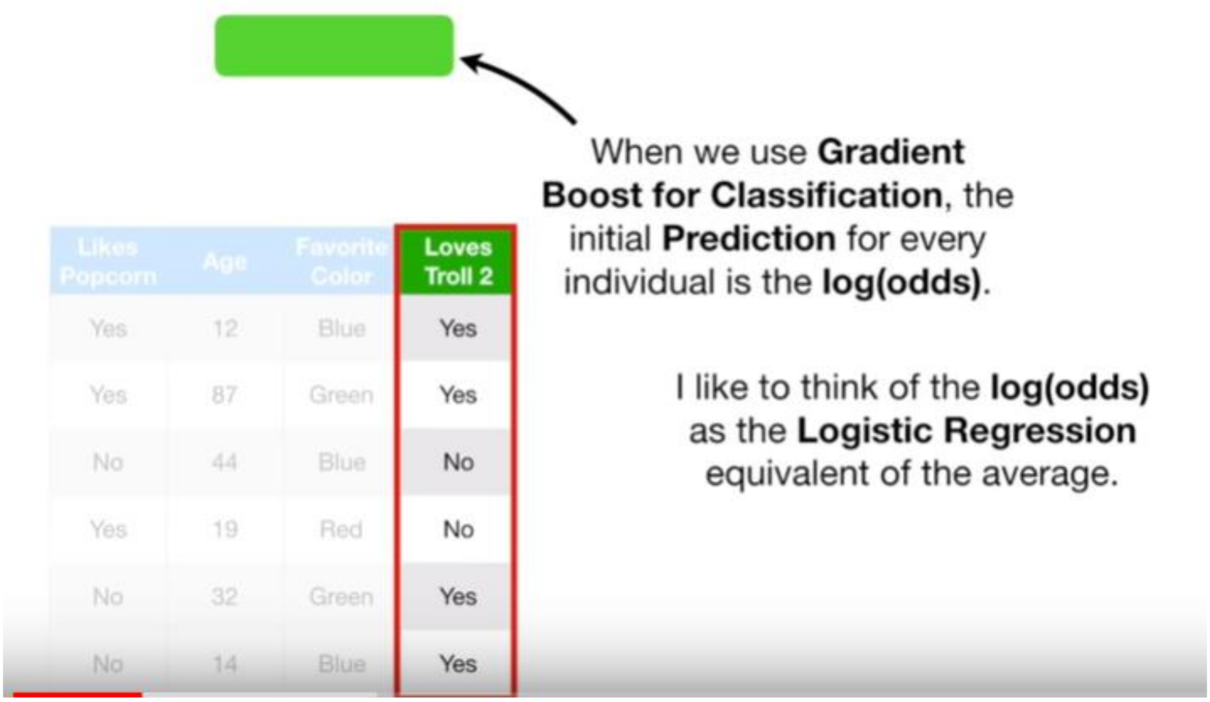



We will start with a **Leaf** that represent Initial Prediction for every individuals:




Step 1: We will initialize an Initial Constant model (Fo(x)) which consist of constant value for every individual. The initial Prediction for each individual will be **Log (Odds)**

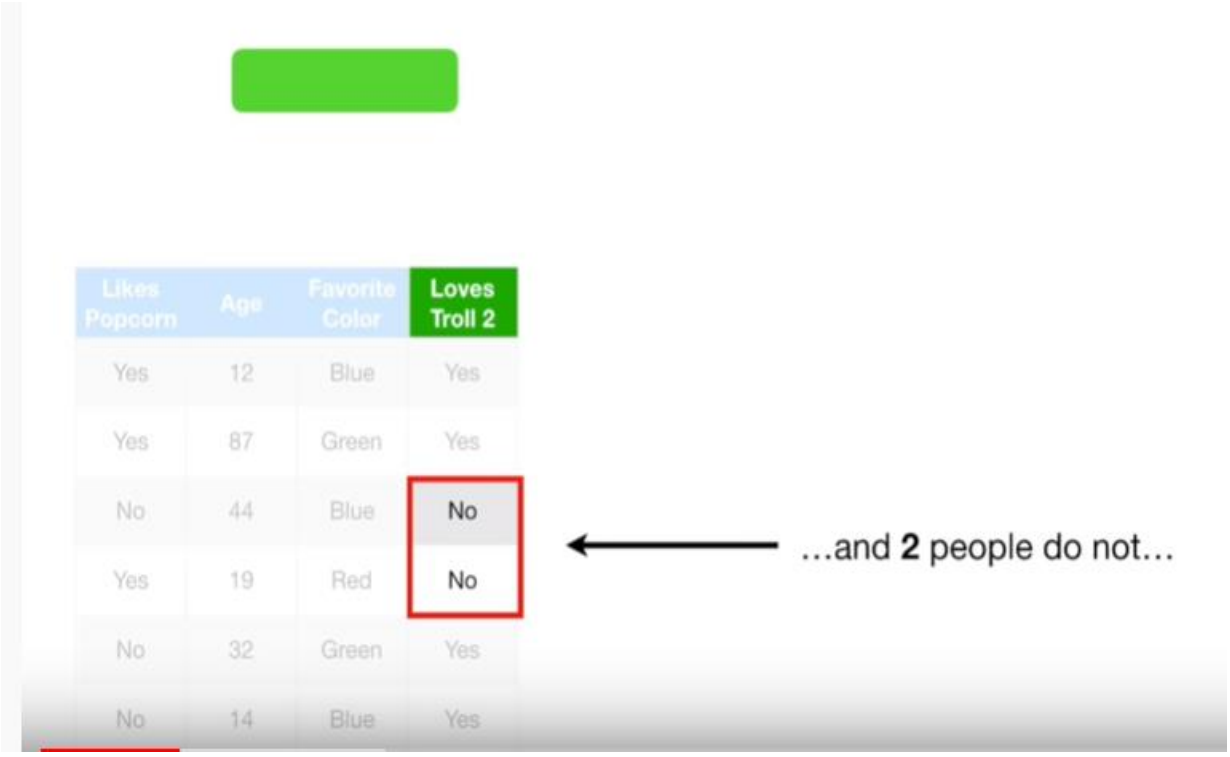


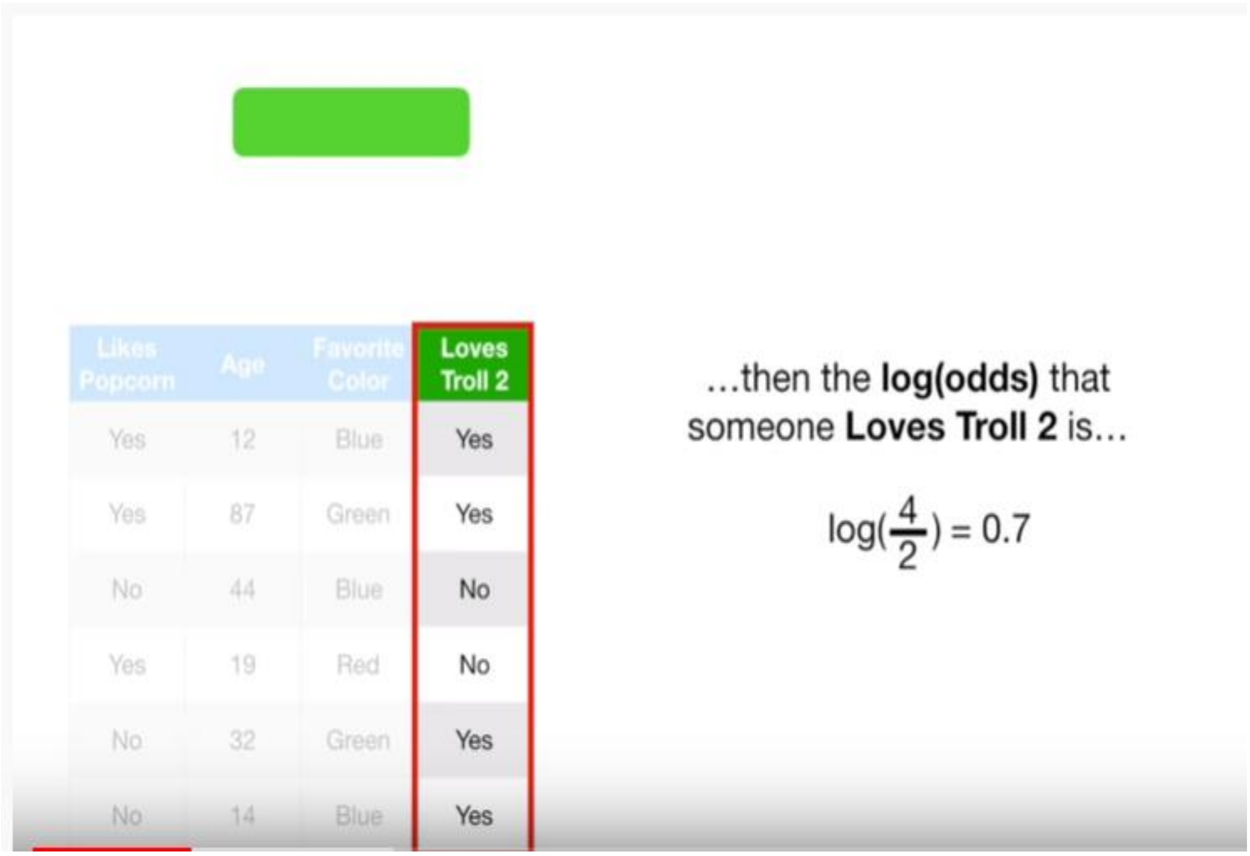




Let’s calculate the **Overall Log (Odds)** that someone loves **Troll 2**

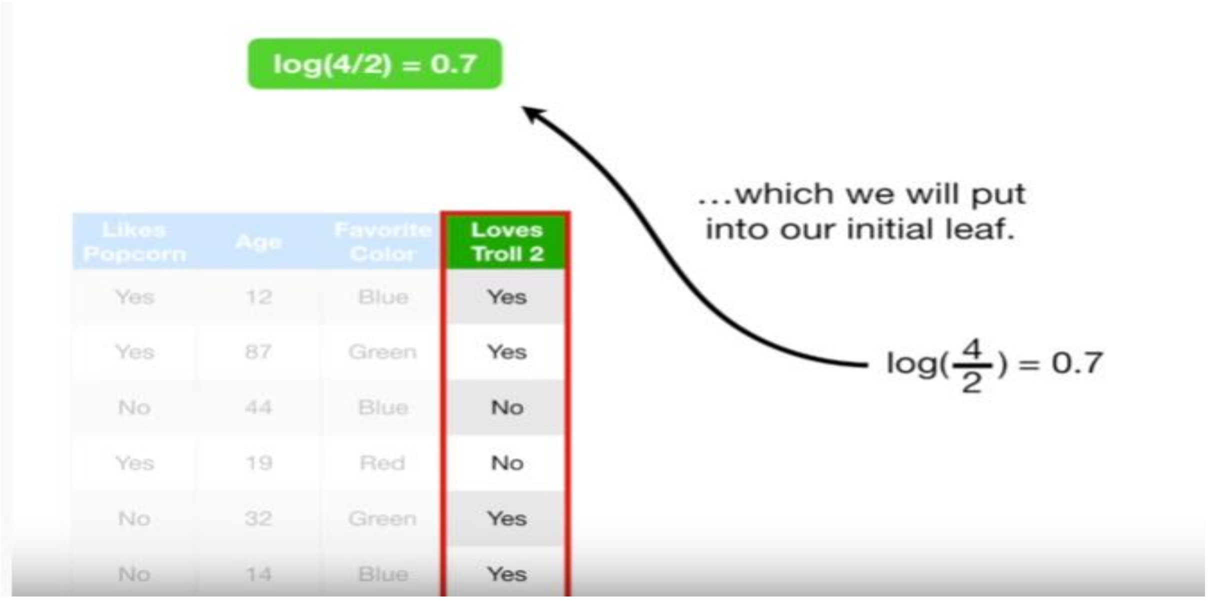

Logistic regression model always calculate/predict **Log (odds)/Probability** with respect to **default class**. Here default class is **Yes** (Love Troll 2). 4 People Love Troll 2 and 2 People doesn't love Troll 2.


So, Log (odds) = p/1-p, p is probability of loving Troll 2 and 1-p is the probability of not loving Troll 2.




Log (p/1-p) = log (4/2) = 0.7





| Likes_Popcorn | Age | Favorite_Color | Loves Troll 2 | Initial_Prediction[Log(odds)=log(4/2=0.7)] |
|---|---|---|---|---|
| Yes | 12 | Blue | **Yes** | **0.7** |
| Yes | 87 | Green | **Yes** | **0.7** |
| No | 44 | Blue | **No** | **0.7** |
| Yes | 19 | Red | **No** | **0.7** |
| No | 32 | Green | **Yes** | **0.7** |
| No | 14 | Blue | **Yes** | **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**...

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


| 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 |

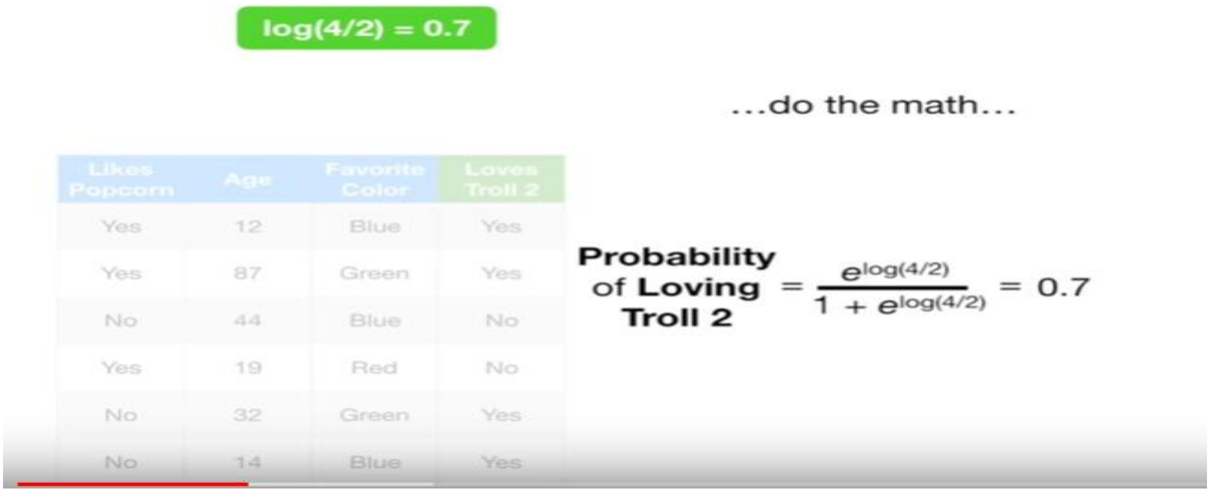

$$\textbf{Probability of Loving Troll 2} = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$

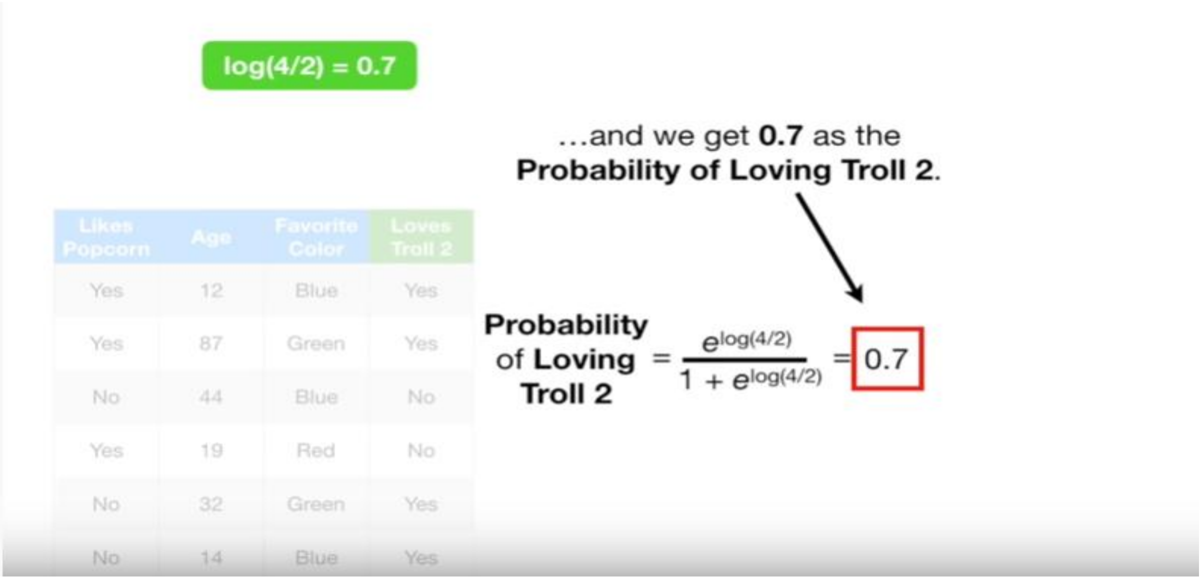
$\log(4/2) = 0.7$

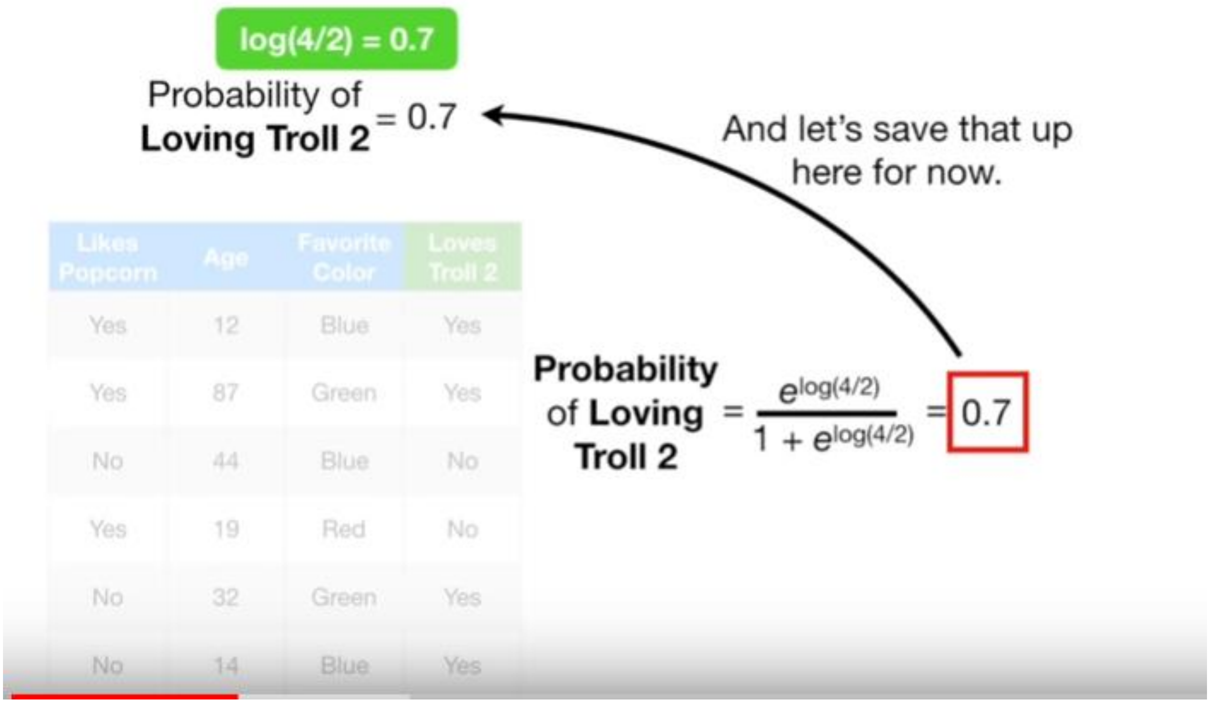
So we plug the **log(odds)** into the **Logistic Function**...


| 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 |

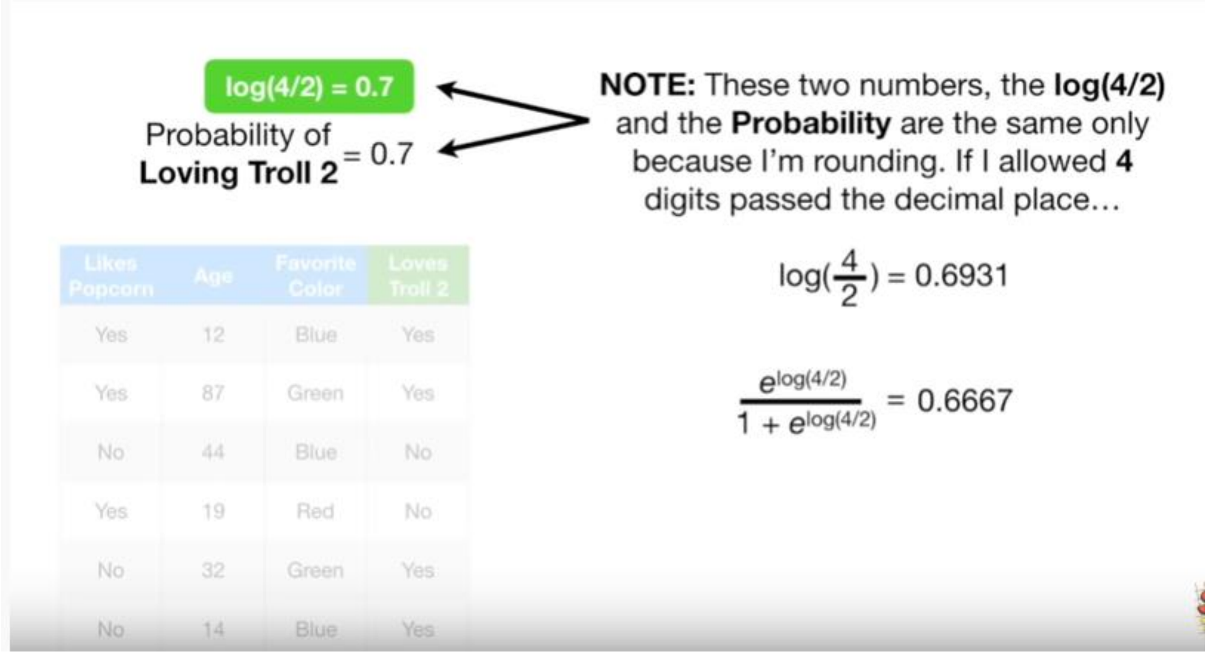

$$\textbf{Probability of Loving Troll 2} = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$$

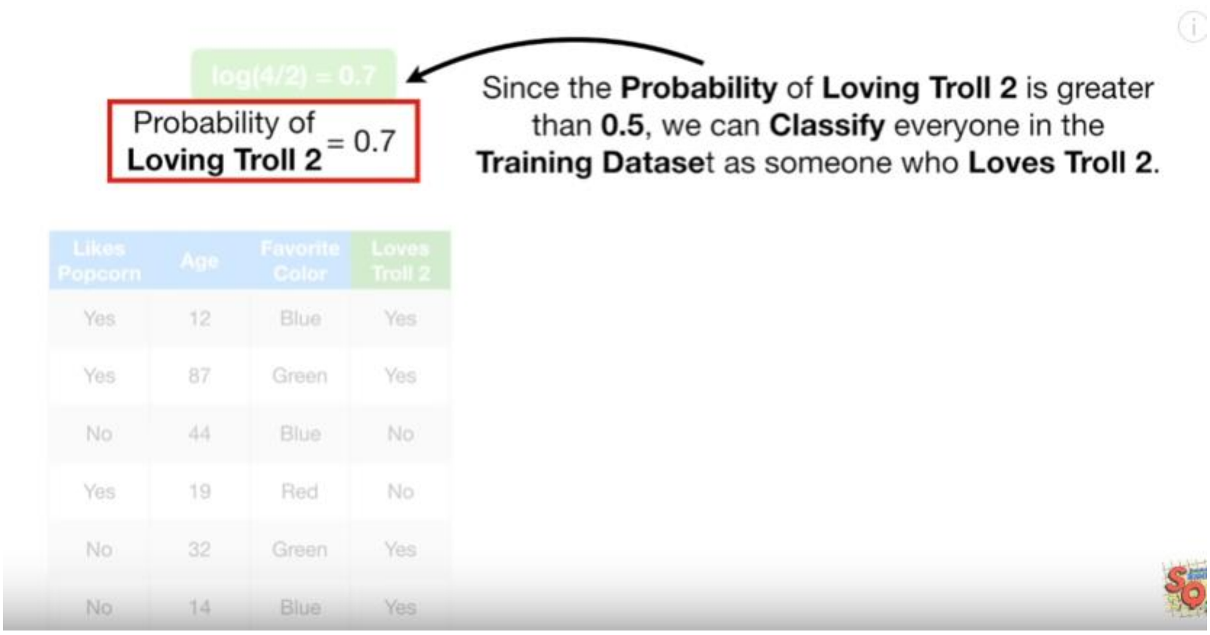

log(4/2) = 0.7
...do the math...
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Probability of Loving Troll 2 = $\frac{e^{\log(4/2)}}{1 + e^{\log(4/2)}}$ = 0.7


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


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
And let's save that up here for now.
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
Probability of Loving Troll 2 = $\frac{e^{\log(4/2)}}{1 + e^{\log(4/2)}}$ = 0.7

**And Probability equals to 0.6667 after taking 4digits after decimal**:




| Likes_Popcorn | Age | Favorite_Color | Loves Troll 2 | Initial_prediction[Log(odds)=log(4/2=0.7)] | Probability |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.7 | 0.7 |
| Yes | 87 | Green | Yes | 0.7 | 0.7 |
| No | 44 | Blue | No | 0.7 | 0.7 |
| Yes | 19 | Red | No | 0.7 | 0.7 |
| No | 32 | Green | Yes | 0.7 | 0.7 |
| No | 14 | Blue | Yes | 0.7 | 0.7 |

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.

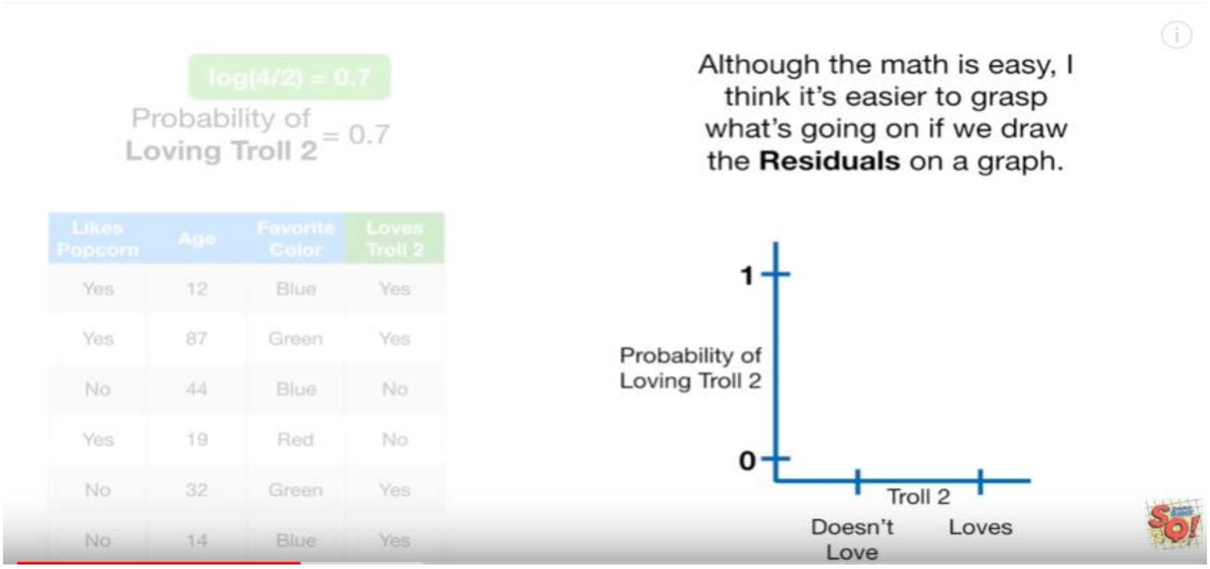
log(4/2) = 0.7

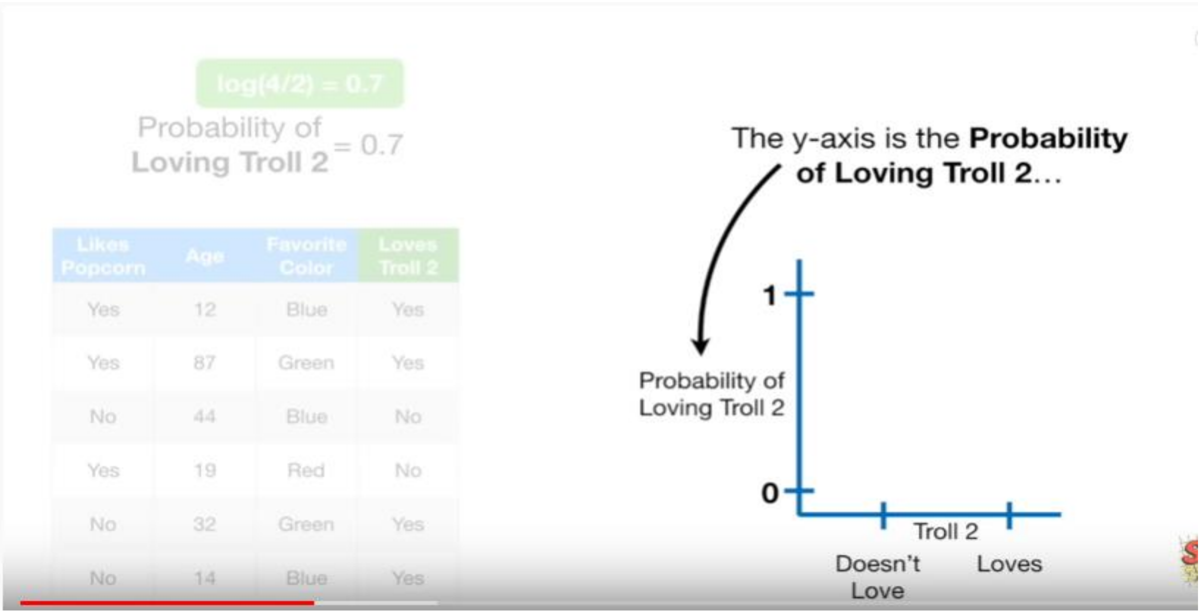
Probability of **Loving Troll 2** = 0.7

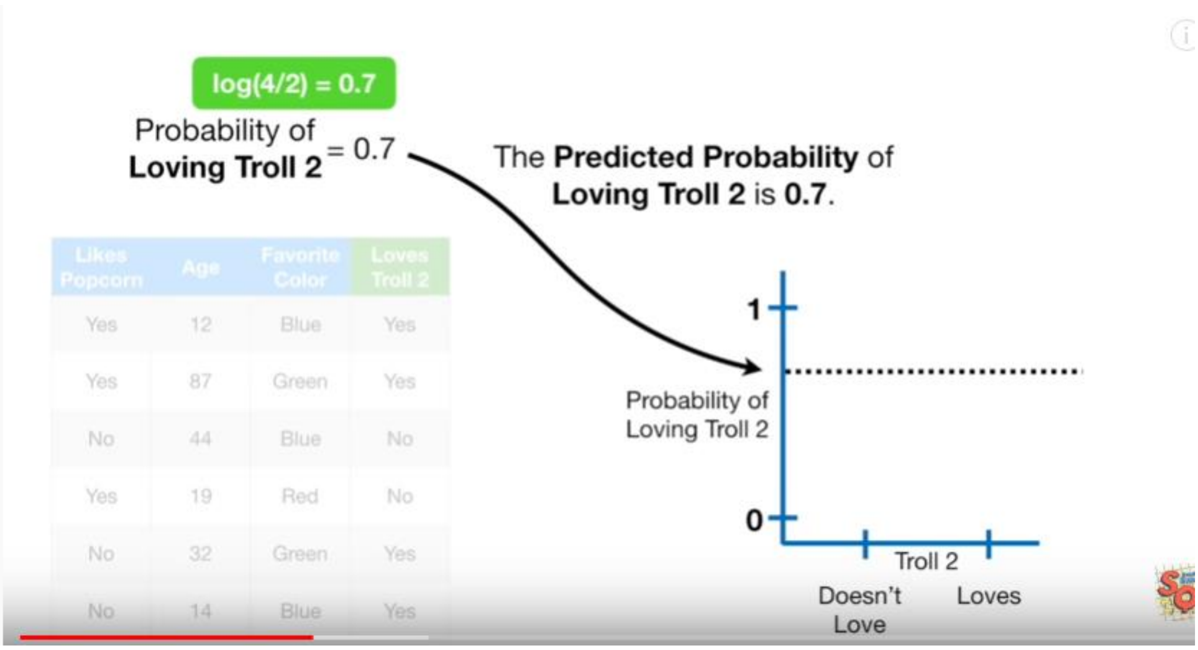
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 |
|---|---|---|---|
| Yes | 12 | Blue | Yes |
| Yes | 87 | Green | Yes |
| No | 44 | Blue | No |
| Yes | 19 | Red | No |
| No | 32 | Green | Yes |
| No | 14 | Blue | Yes |

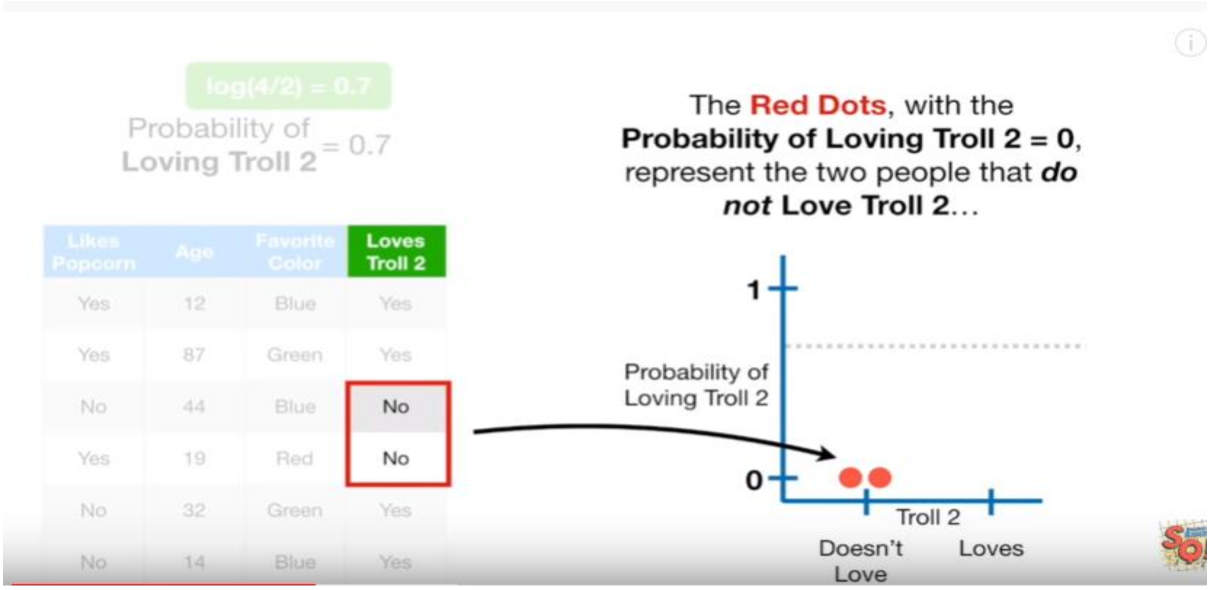
We can measure how bad the initial **Prediction** is by calculating **Pseudo Residuals**, the difference between the **Observed** and the **Predicted** values.

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

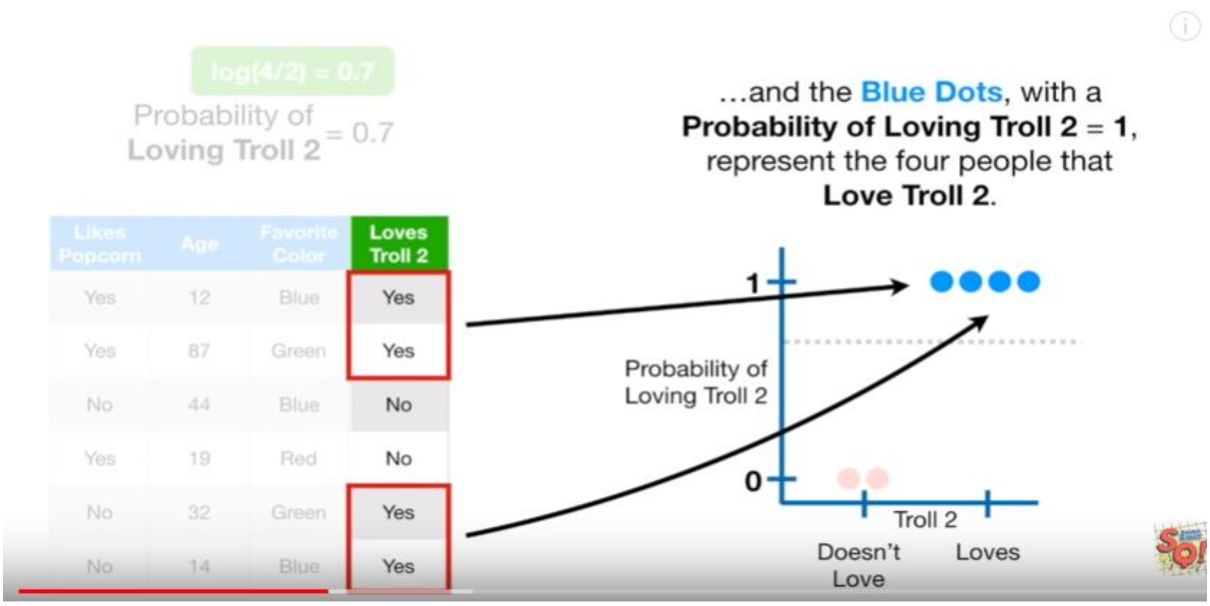

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Although the math is easy, I think it's easier to grasp what's going on if we draw the **Residuals** on a graph.
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves


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
The y-axis is the **Probability of Loving Troll 2**...
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves

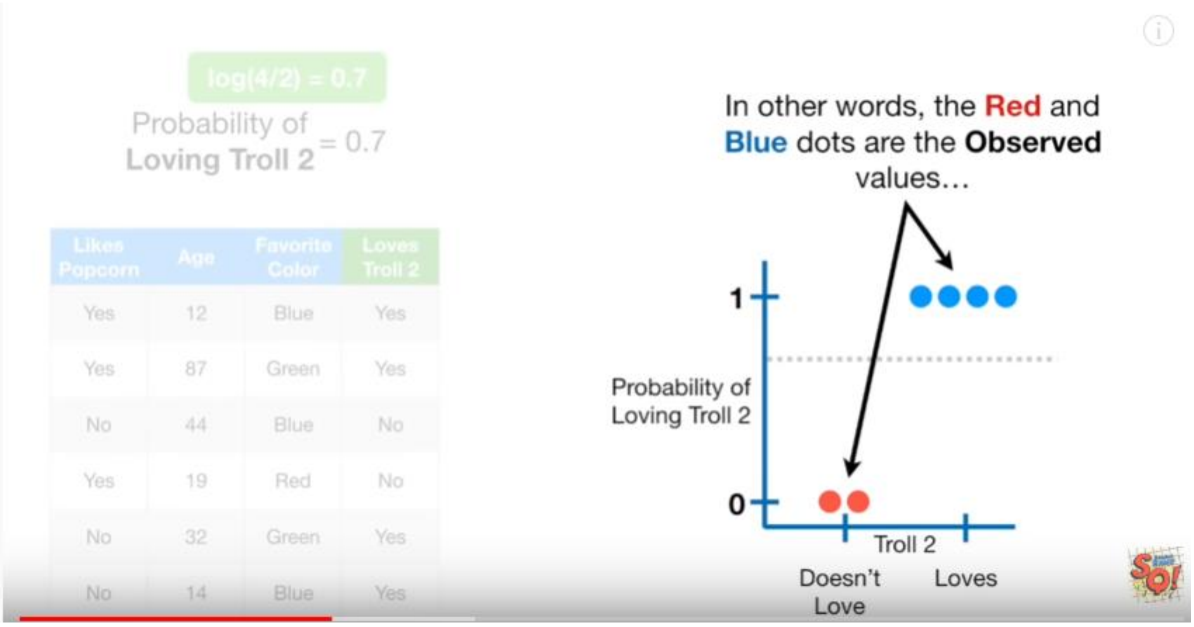

log(4/2) = 0.7
Probability of **Loving Troll 2** = 0.7
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
The **Predicted Probability** of **Loving Troll 2** is **0.7**.
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves



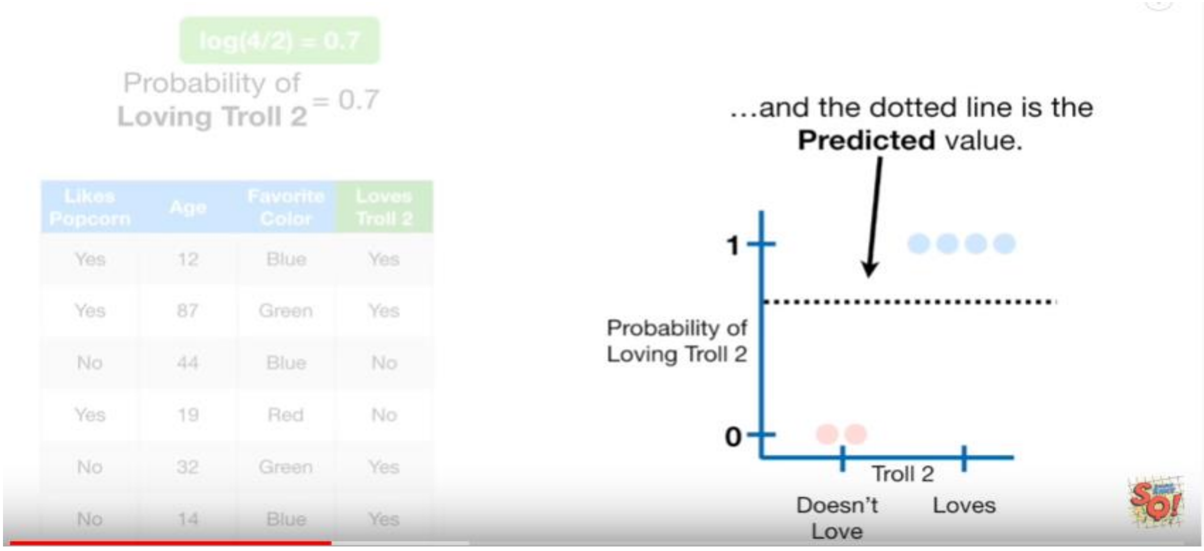
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
The Red Dots, with the Probability of Loving Troll 2 = 0, represent the two people that do not Love Troll 2…
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

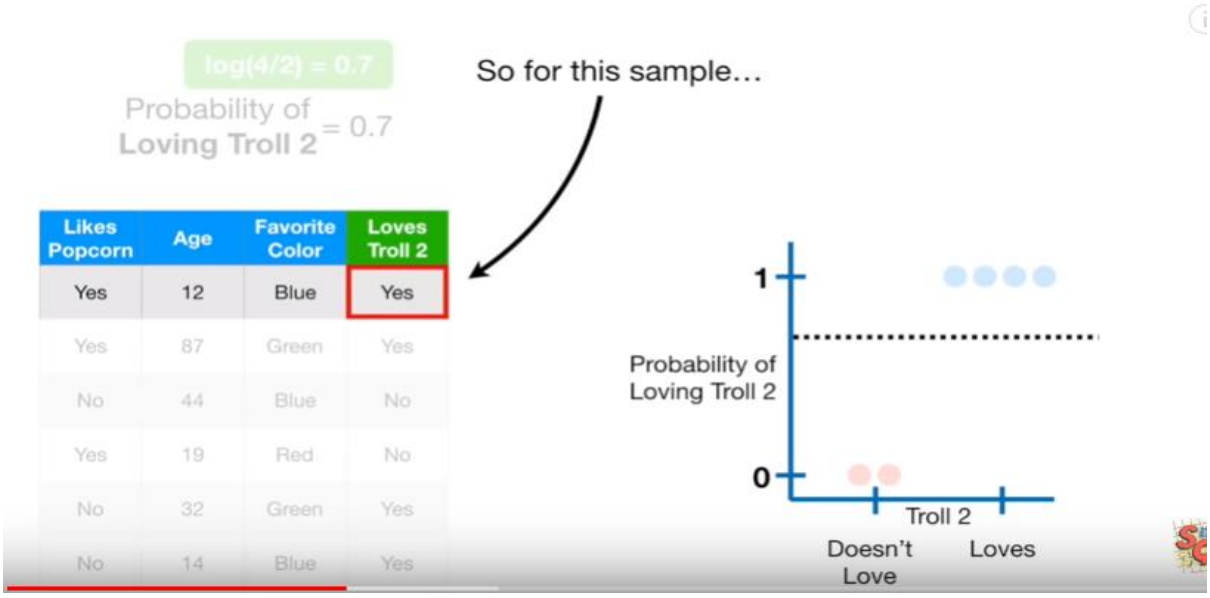



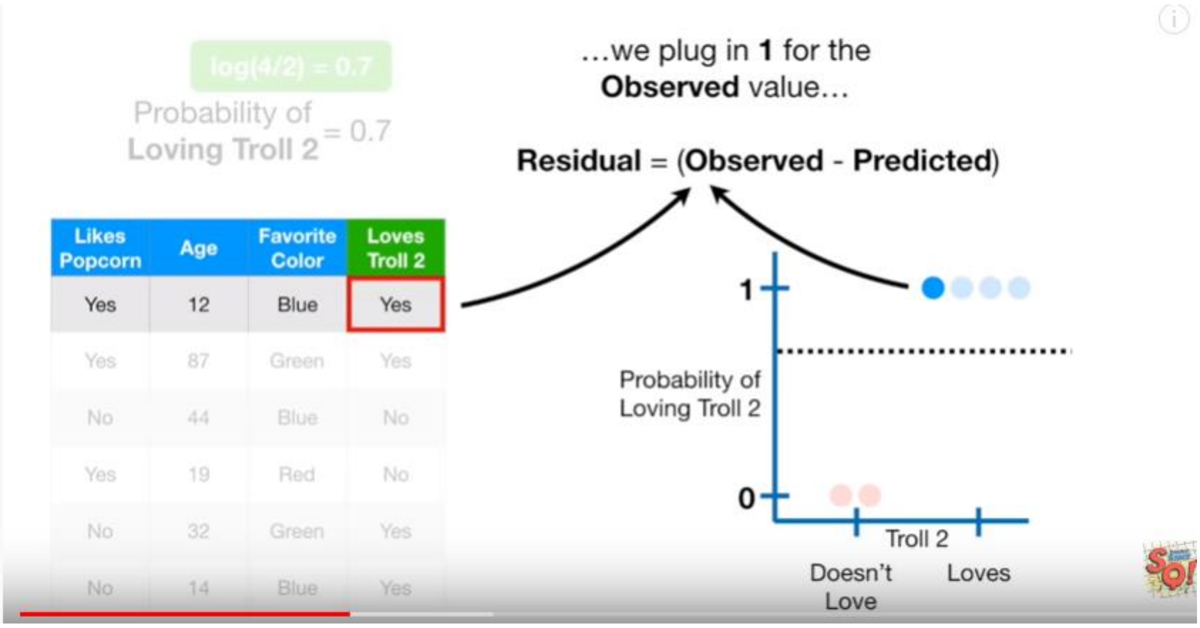
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
…and the Blue Dots, with a Probability of Loving Troll 2 = 1, represent the four people that Love Troll 2.
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

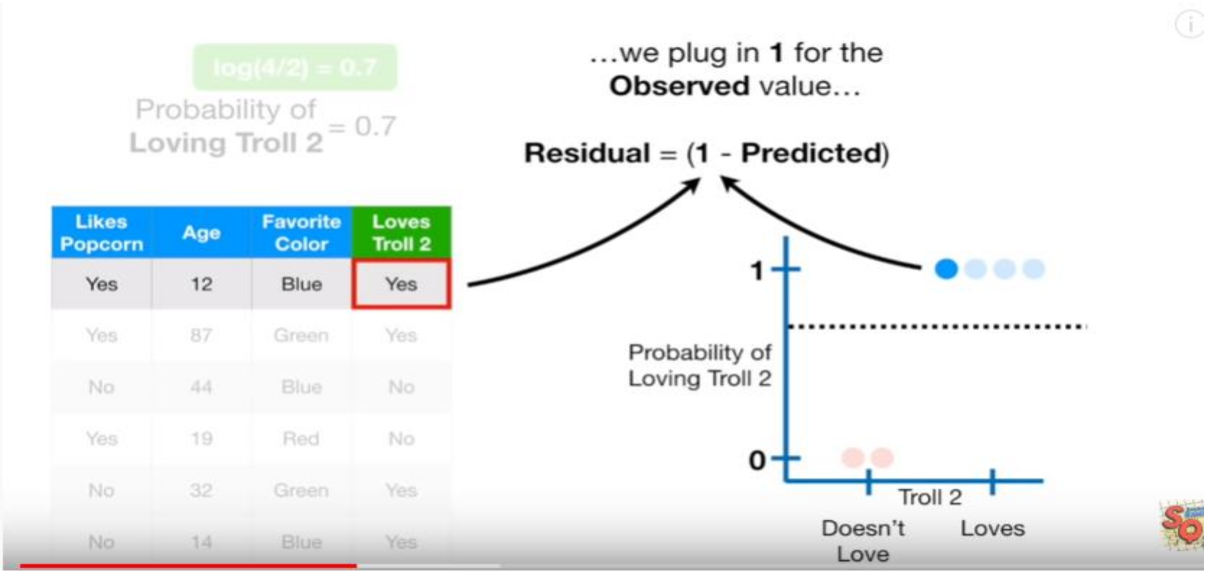



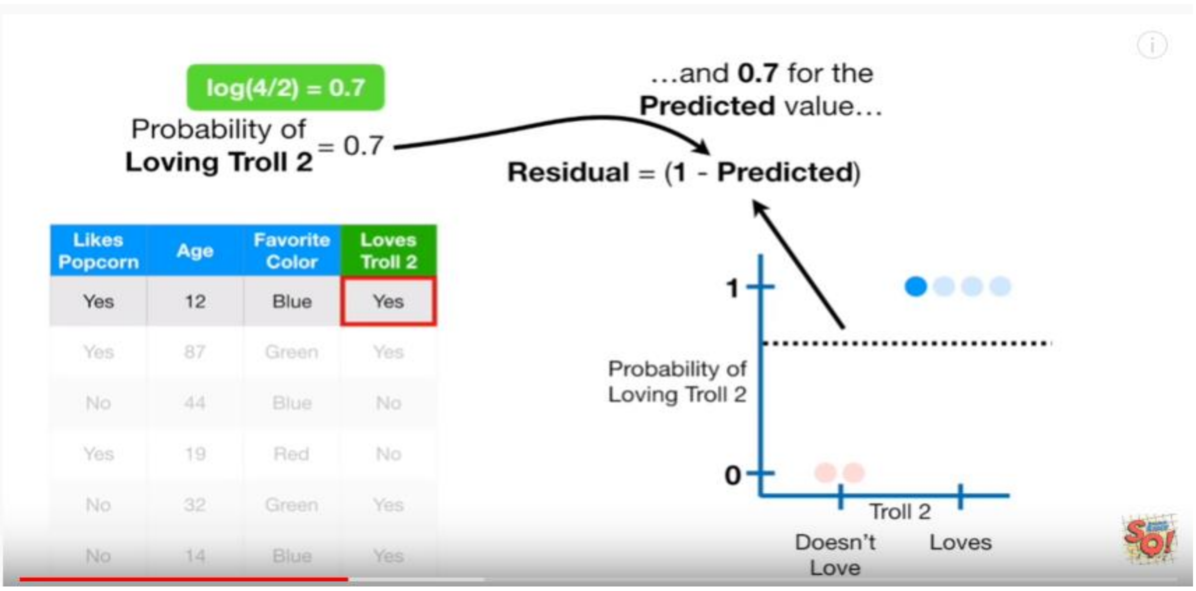
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
In other words, the Red and Blue dots are the Observed values…
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

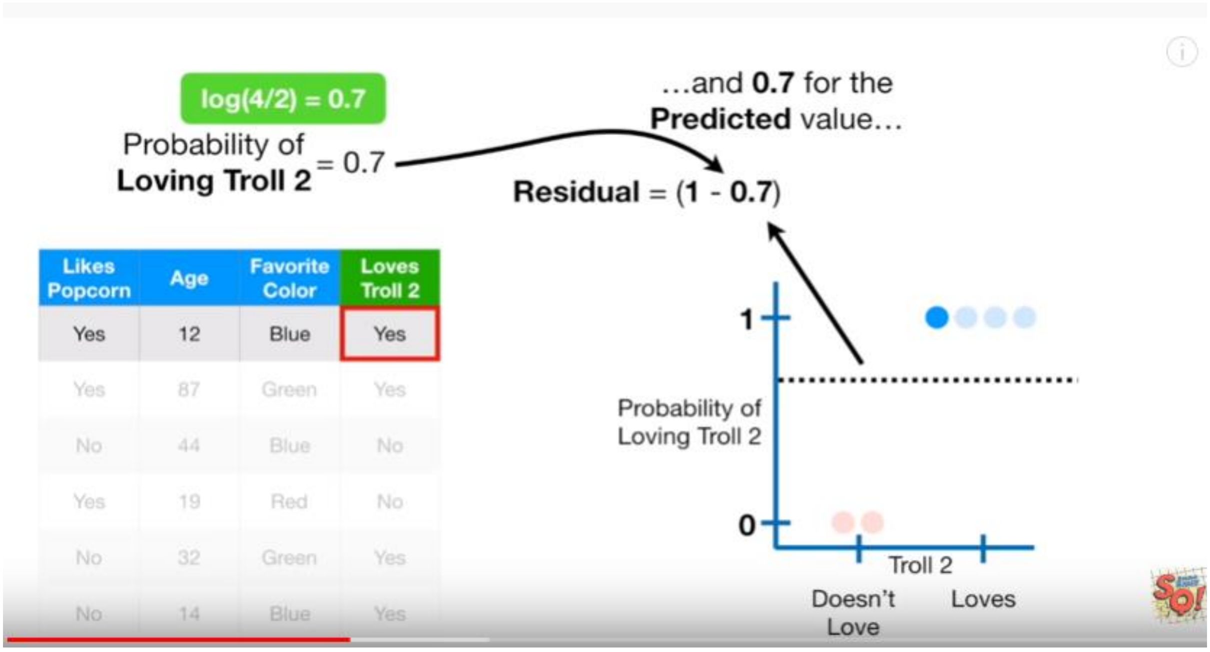

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
...and the dotted line is the Predicted value.
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves

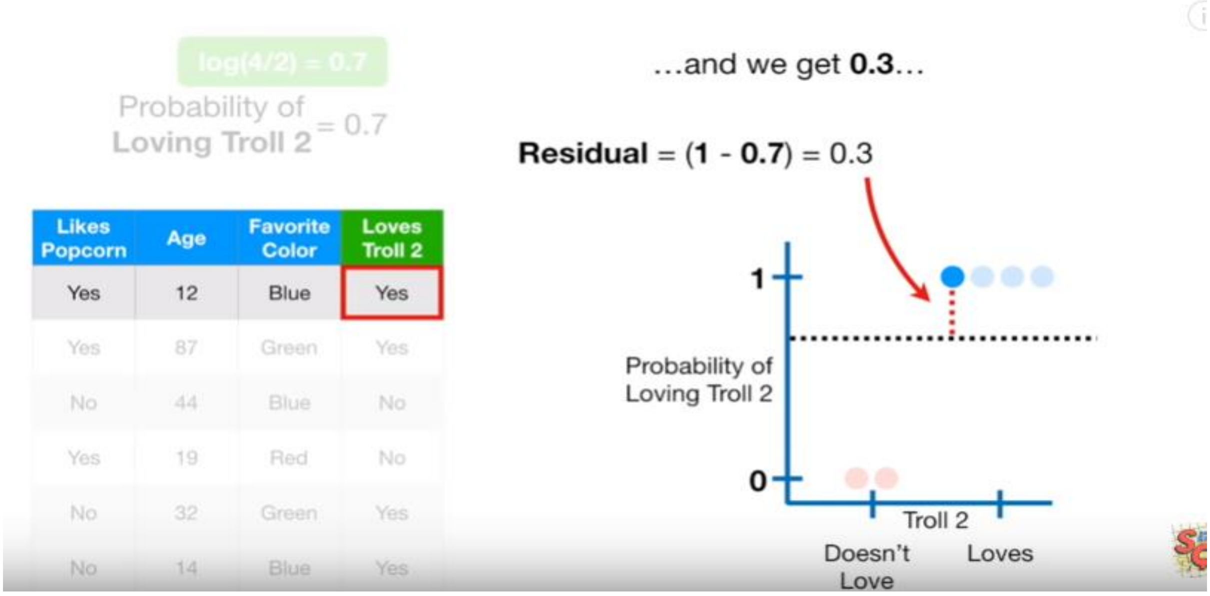

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
So for this sample...
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
...we plug in 1 for the Observed value...
Residual = (Observed - Predicted)
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
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves

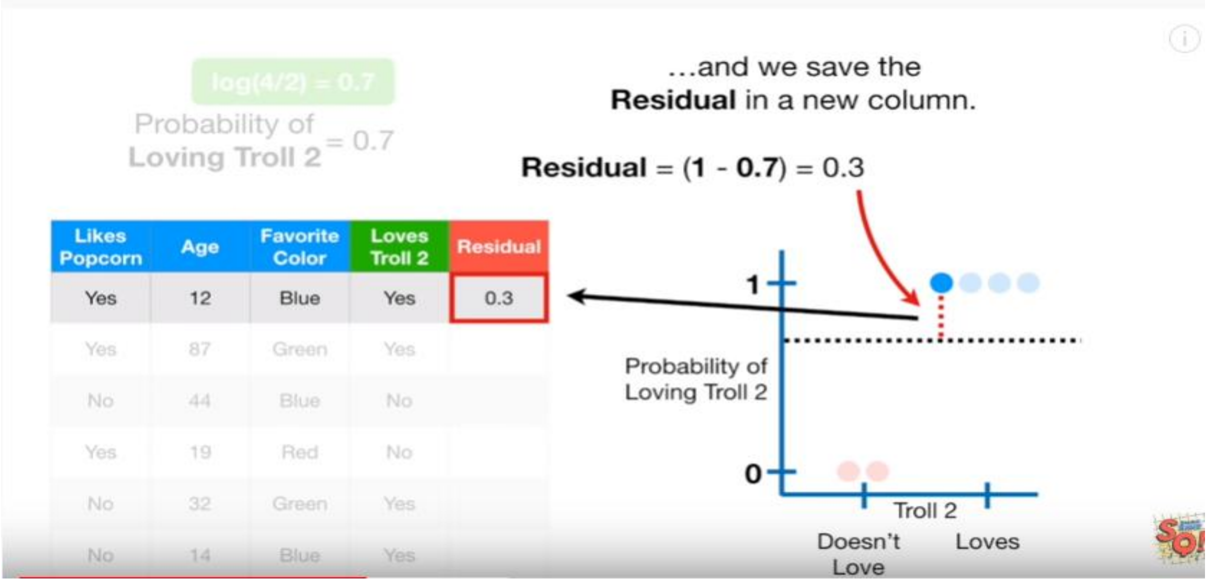

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
…we plug in 1 for the Observed value…
Residual = (1 - Predicted)
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
…and 0.7 for the Predicted value…
Residual = (1 - Predicted)
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
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

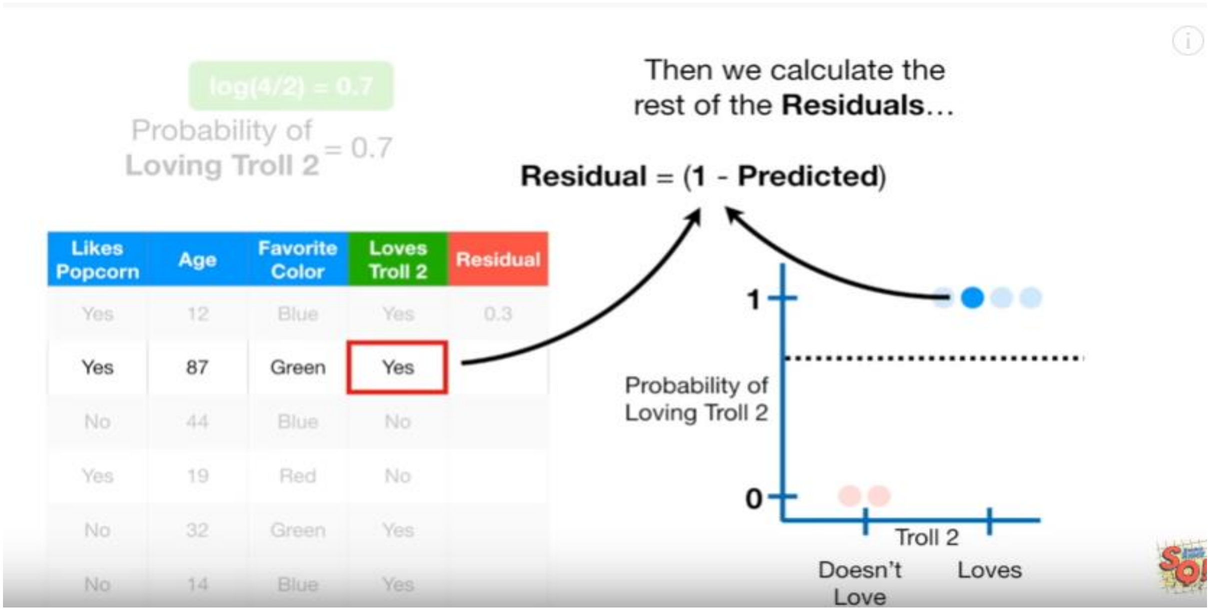

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
…and 0.7 for the Predicted value…
Residual = (1 - 0.7)
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves



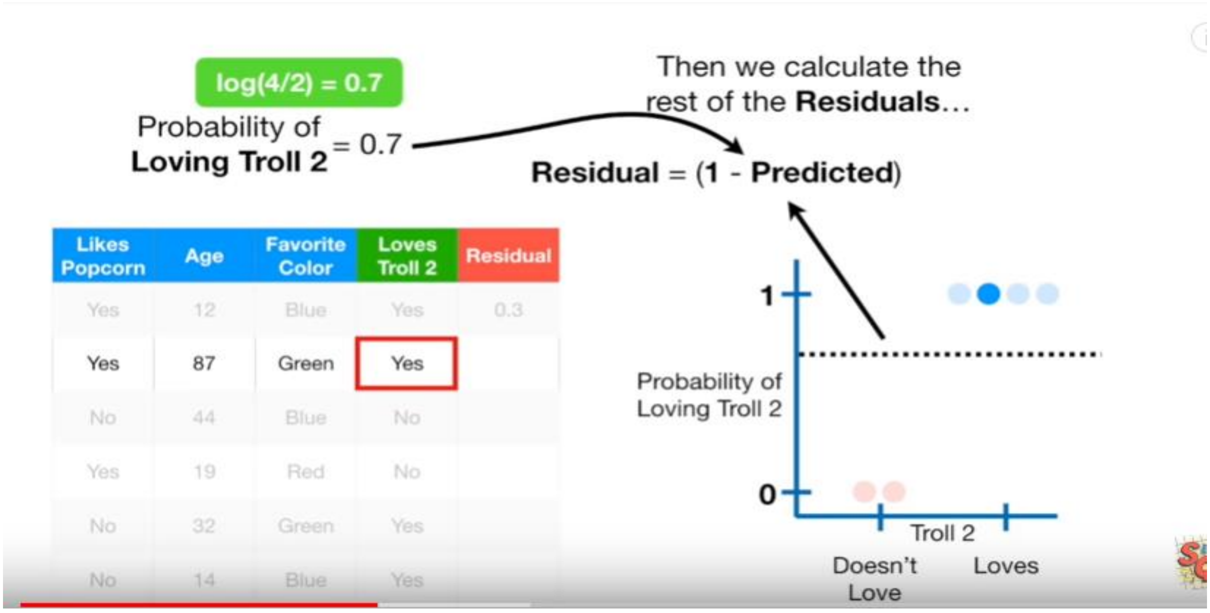
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
...and we get 0.3...
Residual = (1 - 0.7) = 0.3
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

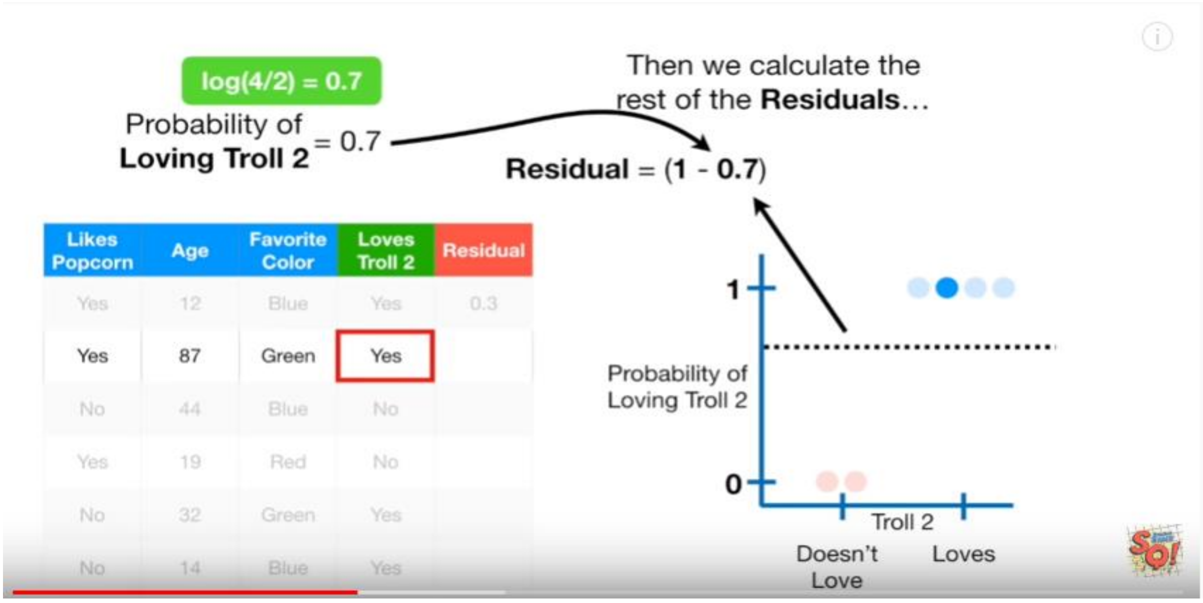



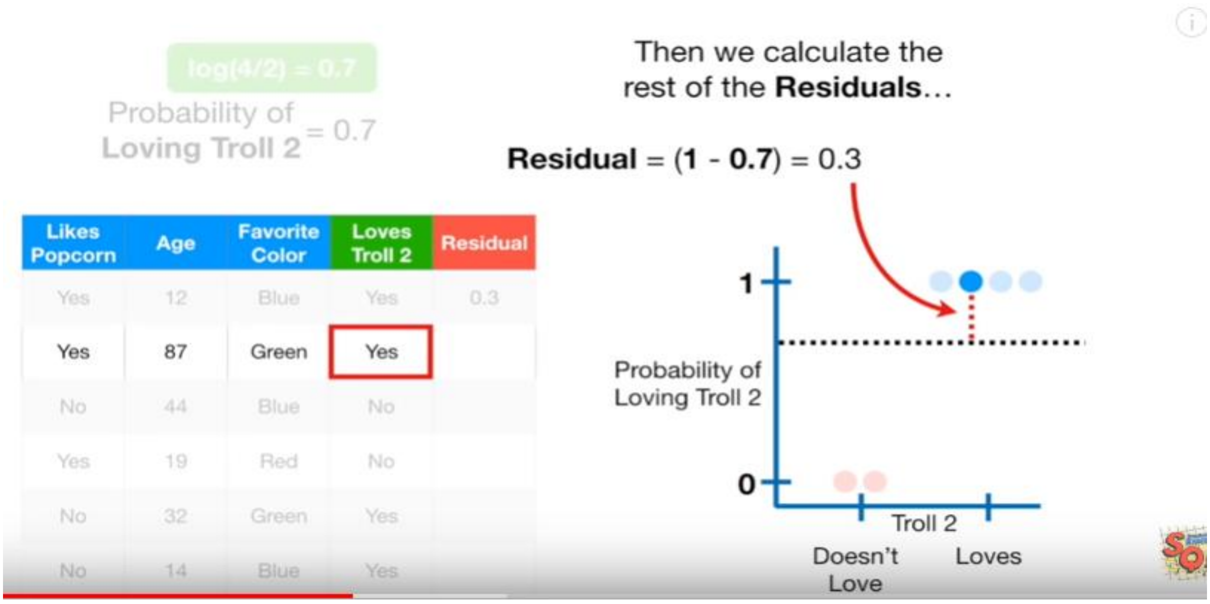
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
...and we save the Residual in a new column.
Residual = (1 - 0.7) = 0.3
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual
Yes | 12 | Blue | Yes | 0.3
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

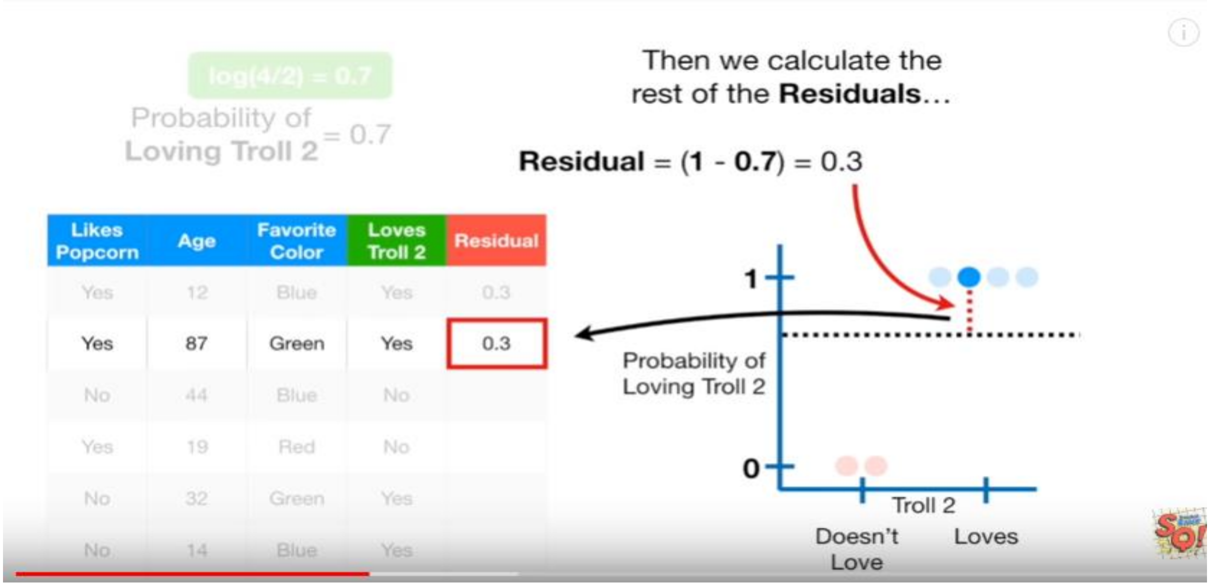



log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (1 - Predicted)
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual
Yes | 12 | Blue | Yes | 0.3
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

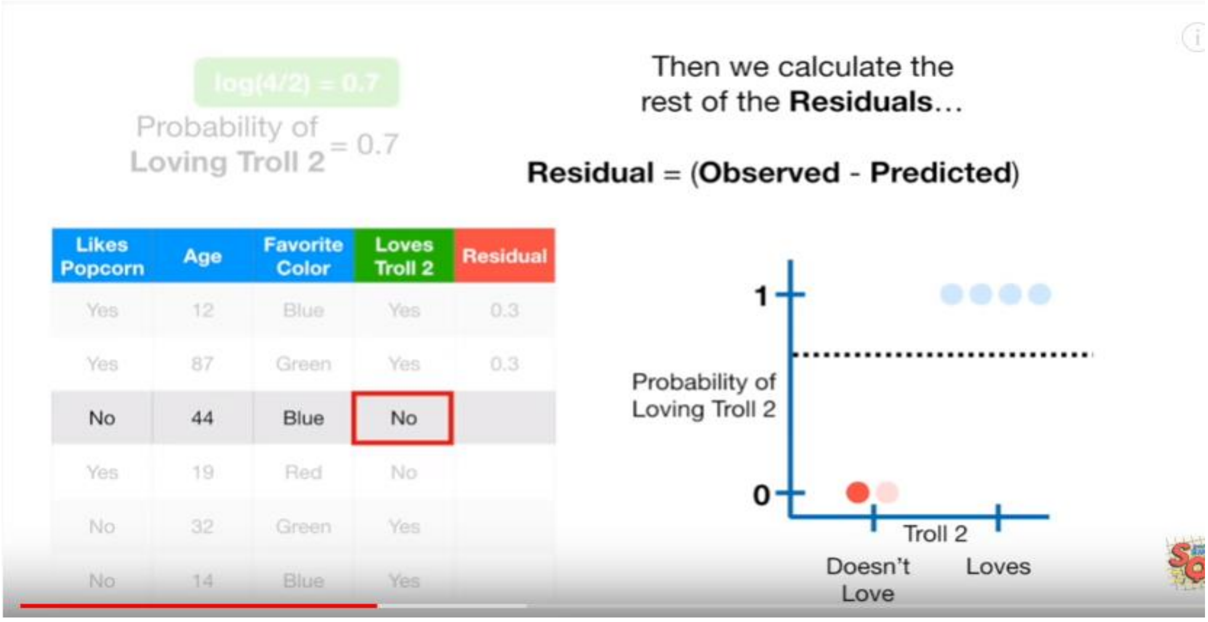

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (1 - Predicted)
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Residual
Yes 12 Blue Yes 0.3
Yes 87 Green Yes
No 44 Blue No
Yes 19 Red No
No 32 Green Yes
No 14 Blue Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (1 - 0.7)
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
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

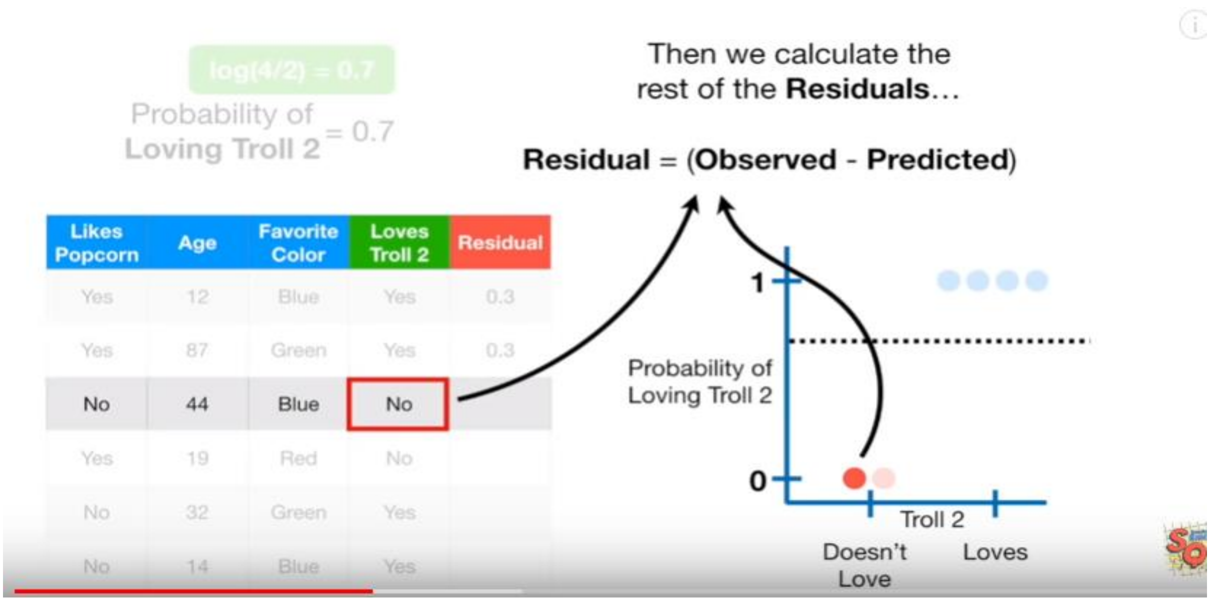

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (1 - 0.7) = 0.3
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Residual
Yes 12 Blue Yes 0.3
Yes 87 Green Yes
No 44 Blue No
Yes 19 Red No
No 32 Green Yes
No 14 Blue Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves



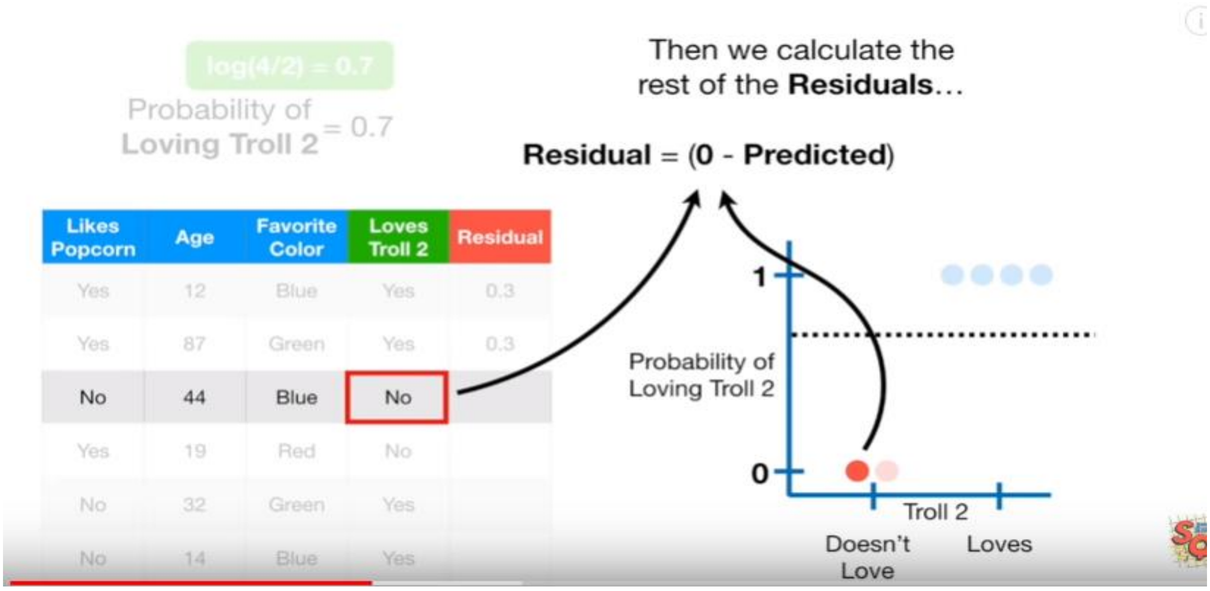
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals…
Residual = (1 - 0.7) = 0.3
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Residual
Yes 12 Blue Yes 0.3
Yes 87 Green Yes 0.3
No 44 Blue No
Yes 19 Red No
No 32 Green Yes
No 14 Blue Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

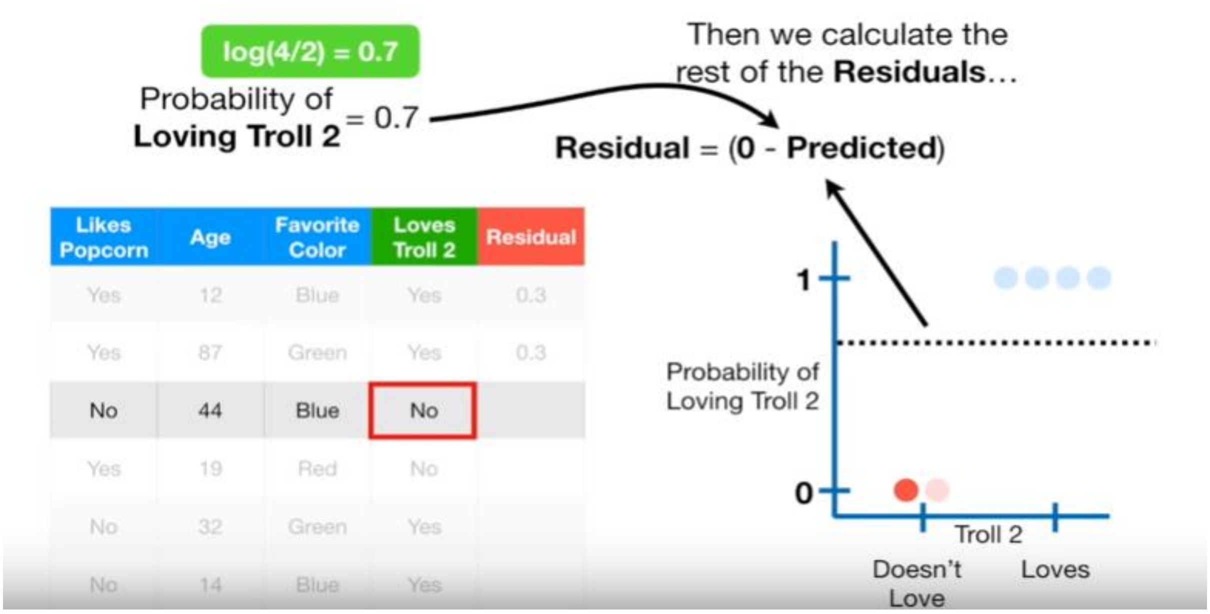



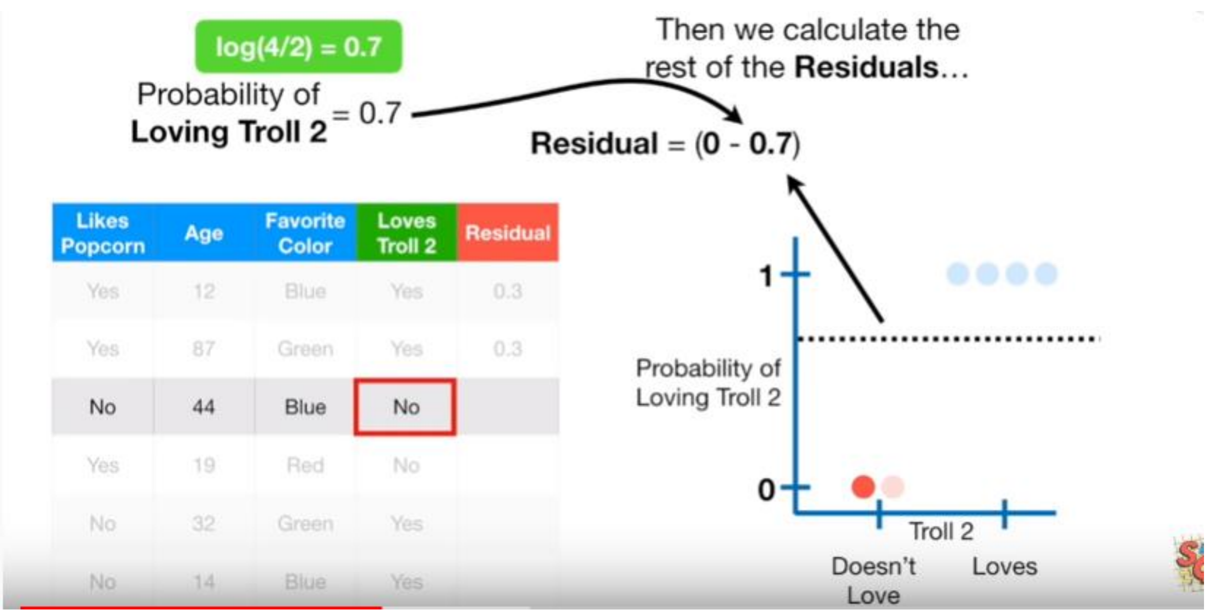
log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals…
Residual = (Observed - Predicted)
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Residual
Yes 12 Blue Yes 0.3
Yes 87 Green Yes 0.3
No 44 Blue No
Yes 19 Red No
No 32 Green Yes
No 14 Blue Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals…
Residual = (Observed - Predicted)
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
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the **Residuals**…
**Residual** = (**0** - **Predicted**)
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 |
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the **Residuals**…
**Residual** = (**0** - **Predicted**)
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 |
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the **Residuals**…
**Residual** = (**0** - **0.7**)
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 |
1
Probability of Loving Troll 2
0
Troll 2
Doesn't Love
Loves

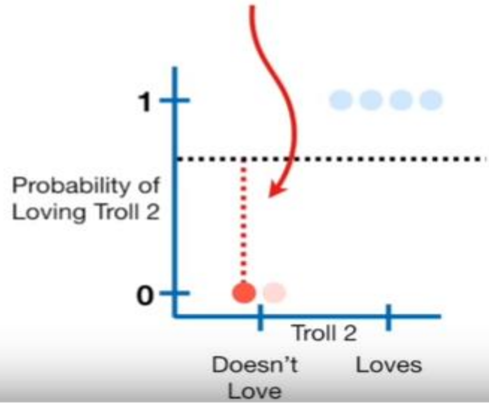
log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

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

**Residual** = (**0** - **0.7**) = -0.7

| 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 | |



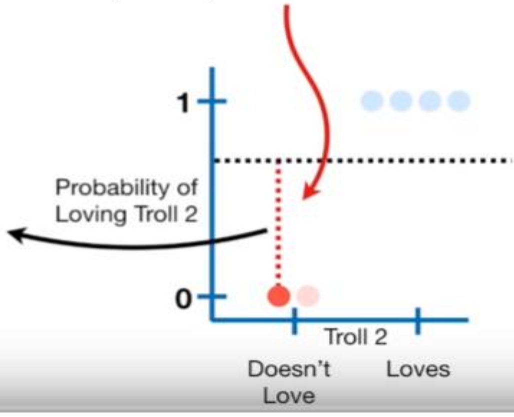

log(4/2) = 0.7

Probability of Loving Troll 2 = 0.7

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

**Residual** = (**0** - **0.7**) = -0.7

| 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 | |
| No | 32 | Green | Yes | |
| No | 14 | Blue | Yes | |



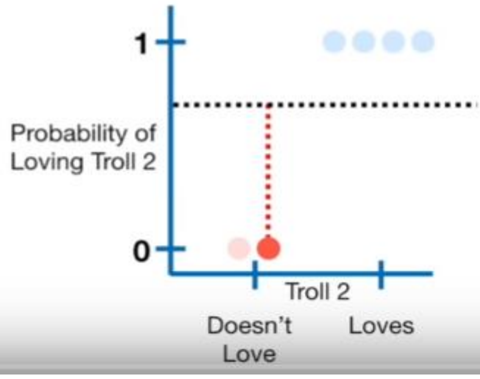

log(4/2) = 0.7

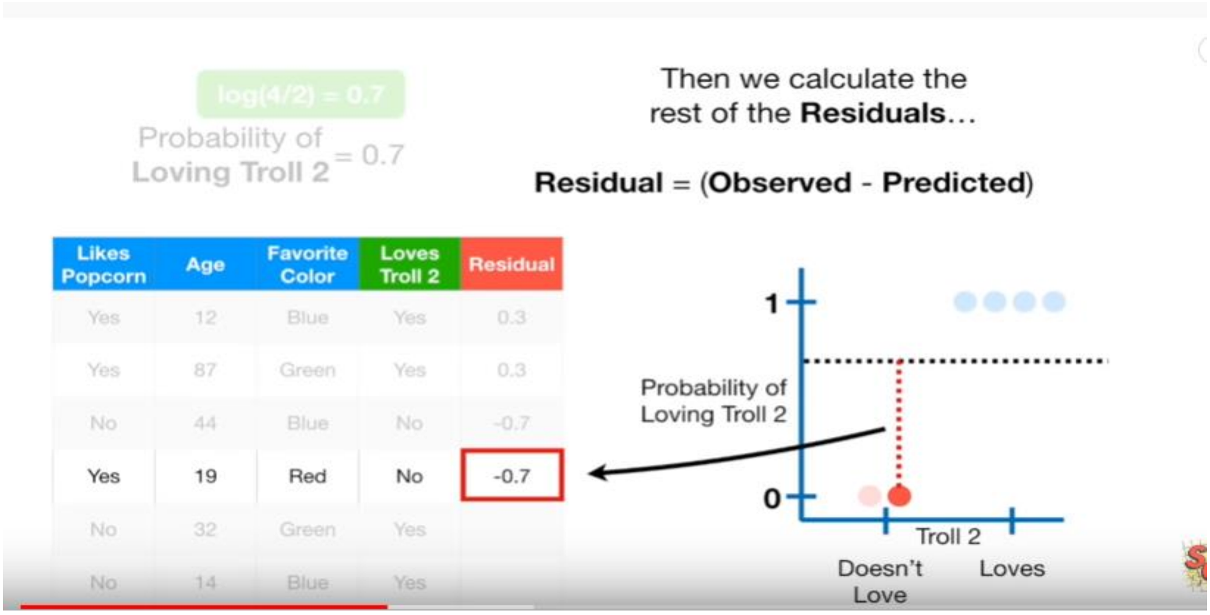
Probability of Loving Troll 2 = 0.7

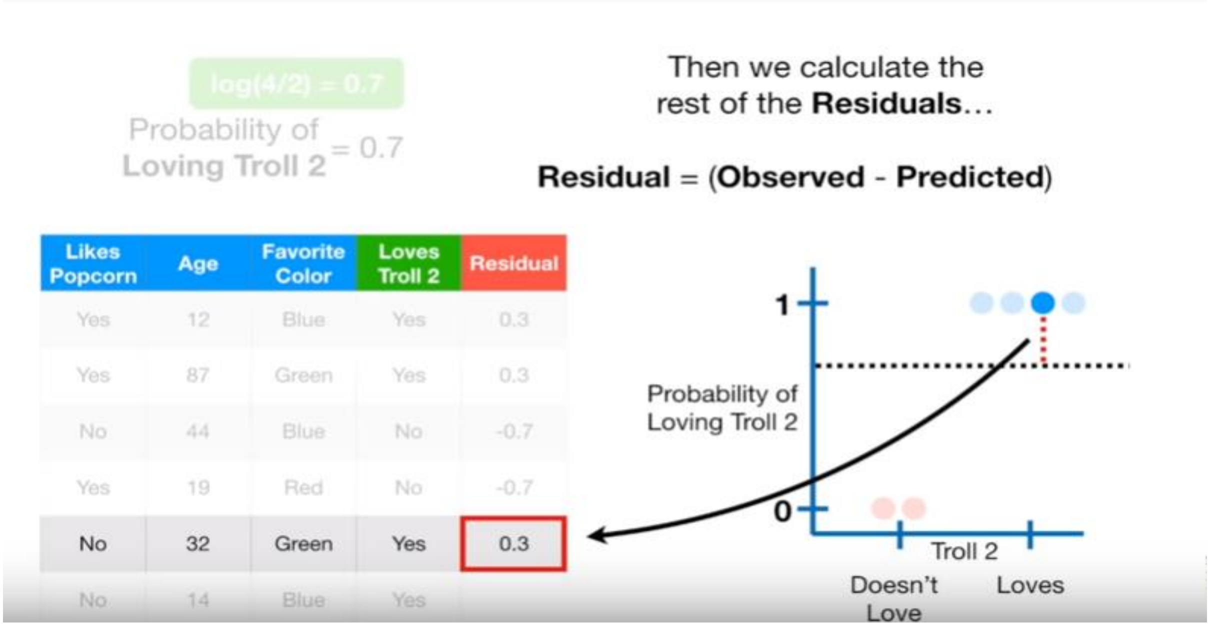
Then we calculate the rest of the **Residuals**...

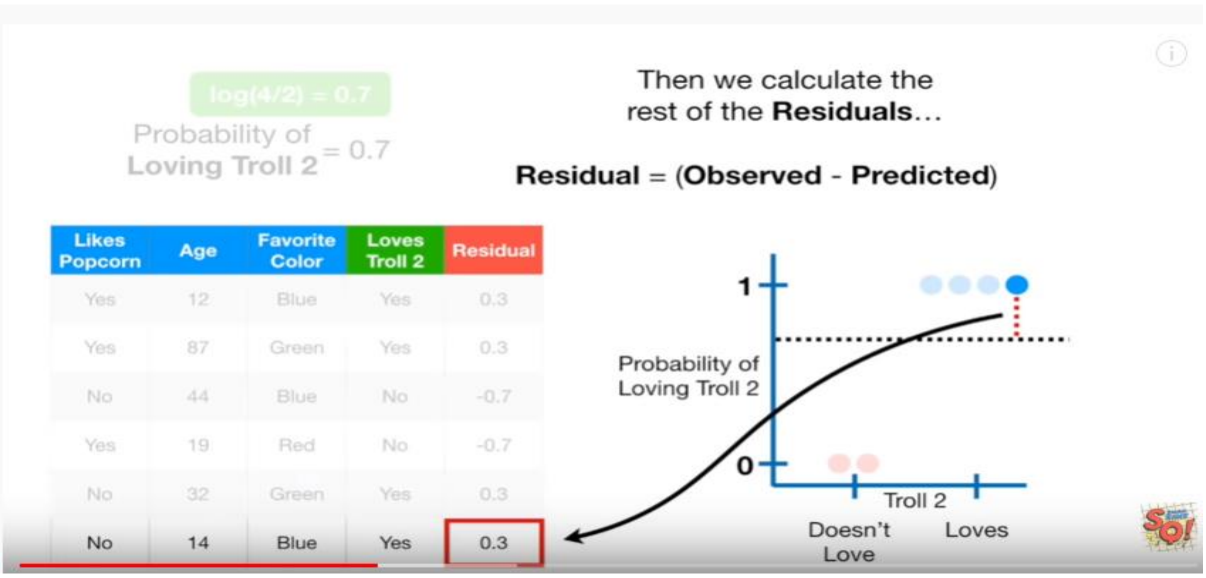
**Residual** = (**Observed** - **Predicted**)

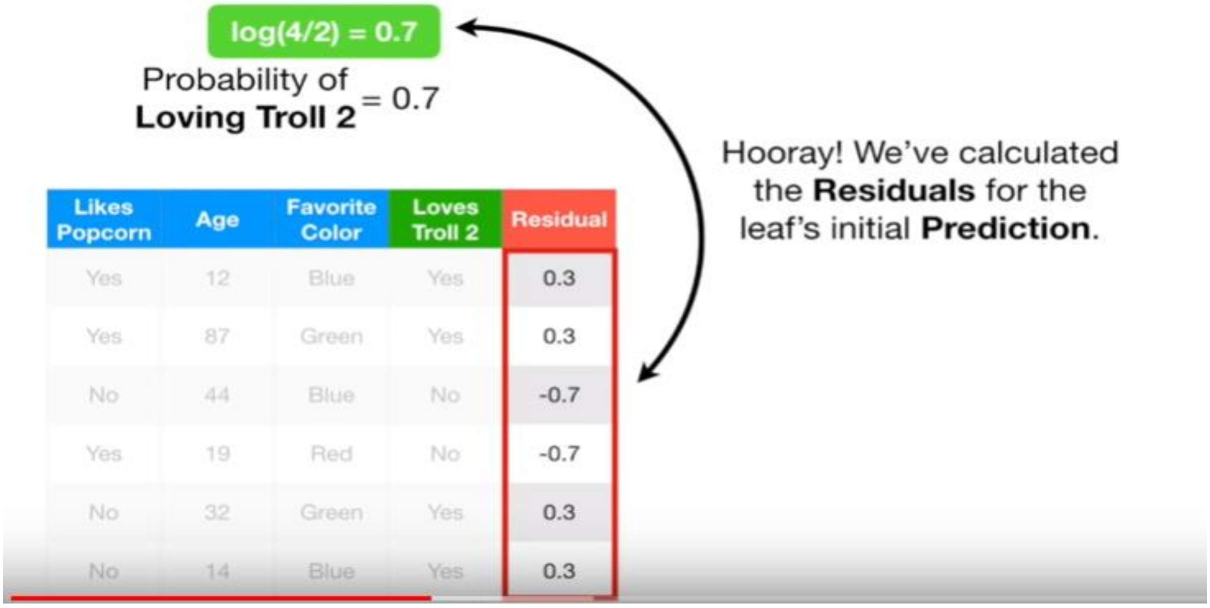
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual |
|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.3 |
| Yes | 87 | Green | Yes | 0.3 |
| No | 44 | Blue | No | -0.7 |
| Yes | 19 | Red | No | |
| No | 32 | Green | Yes | |
| No | 14 | Blue | Yes | |



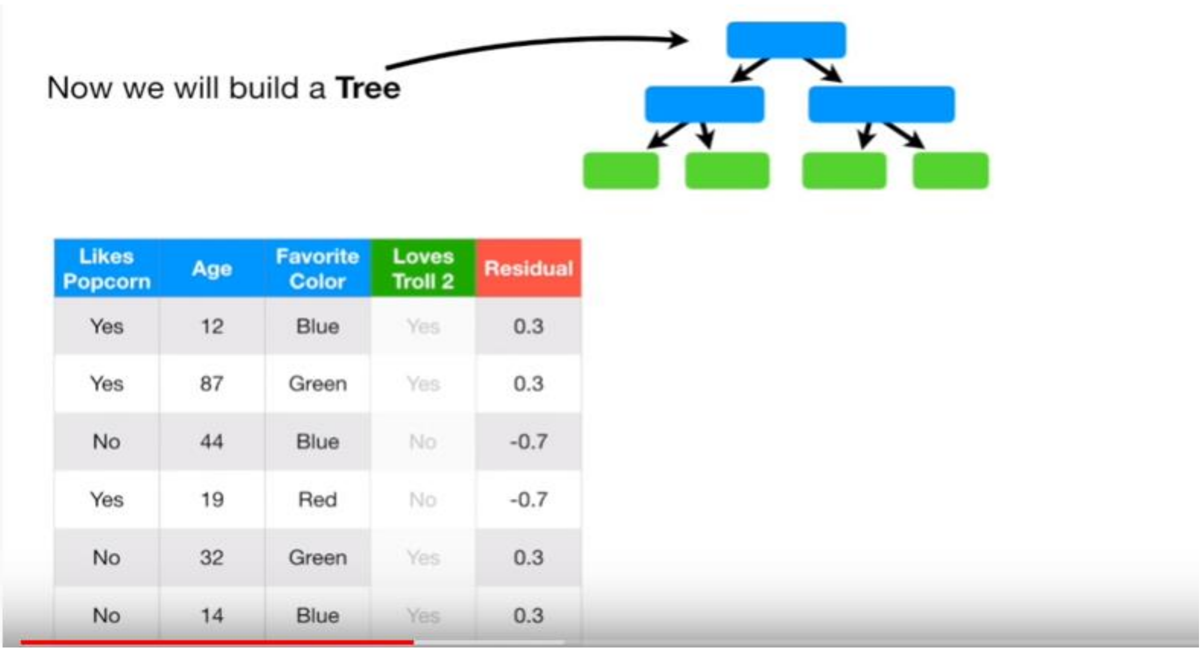

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (Observed - Predicted)
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual
Yes | 12 | Blue | Yes | 0.3
Yes | 87 | Green | Yes | 0.3
No | 44 | Blue | No | -0.7
Yes | 19 | Red | No | -0.7
No | 32 | Green | Yes
No | 14 | Blue | Yes
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves


log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (Observed - Predicted)
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
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves

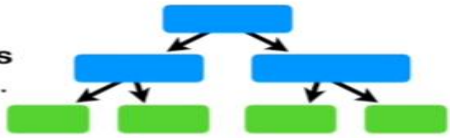

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Then we calculate the rest of the Residuals...
Residual = (Observed - Predicted)
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual
Yes | 12 | Blue | Yes | 0.3
Yes | 87 | Green | Yes | 0.3
No | 44 | Blue | No | -0.7
Yes | 19 | Red | No | -0.7
No | 32 | Green | Yes | 0.3
No | 14 | Blue | Yes | 0.3
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves




| Actual_Probability | Initial_prediction[Log(odds)=log(4/2=0.7)] | Predicted_Probability(1st) | Ressiduals(Actual_Prob - Predicted Prob) |
|---|---|---|---|
| 1 | 0.7 | 0.7 | 1-0.7 = 0.3 |
| 1 | 0.7 | 0.7 | 1-0.7 = 0.3 |
| 0 | 0.7 | 0.7 | 0-0.7 = -0.7 |
| 0 | 0.7 | 0.7 | 0-0.7 = -0.7 |
| 1 | 0.7 | 0.7 | 1-0.7 = 0.3 |
| 1 | 0.7 | 0.7 | 1-0.7 = 0.3 |

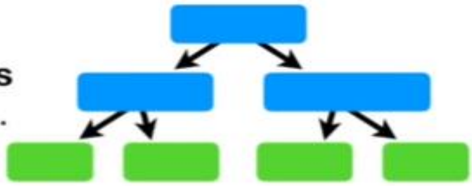


Now we will build a **Tree**, using **Likes Popcorn**, **Age** and **Favorite Color**...



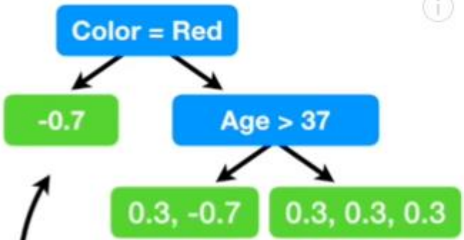
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual |
|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.3 |
| Yes | 87 | Green | Yes | 0.3 |
| No | 44 | Blue | No | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |

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 |

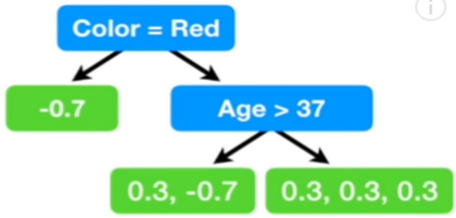
...to **Predict** the **Residuals**.




| 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 |

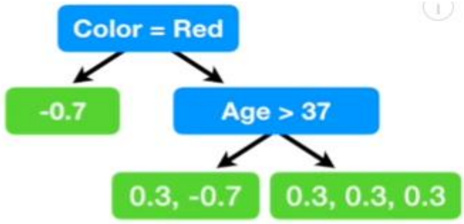
And here's the tree!

| 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.

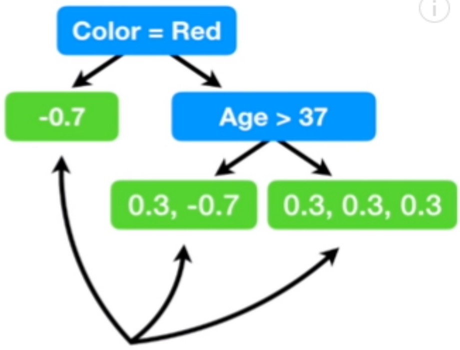
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual |
|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.3 |
| Yes | 87 | Green | Yes | 0.3 |
| No | 44 | Blue | No | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |




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**

| 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 |

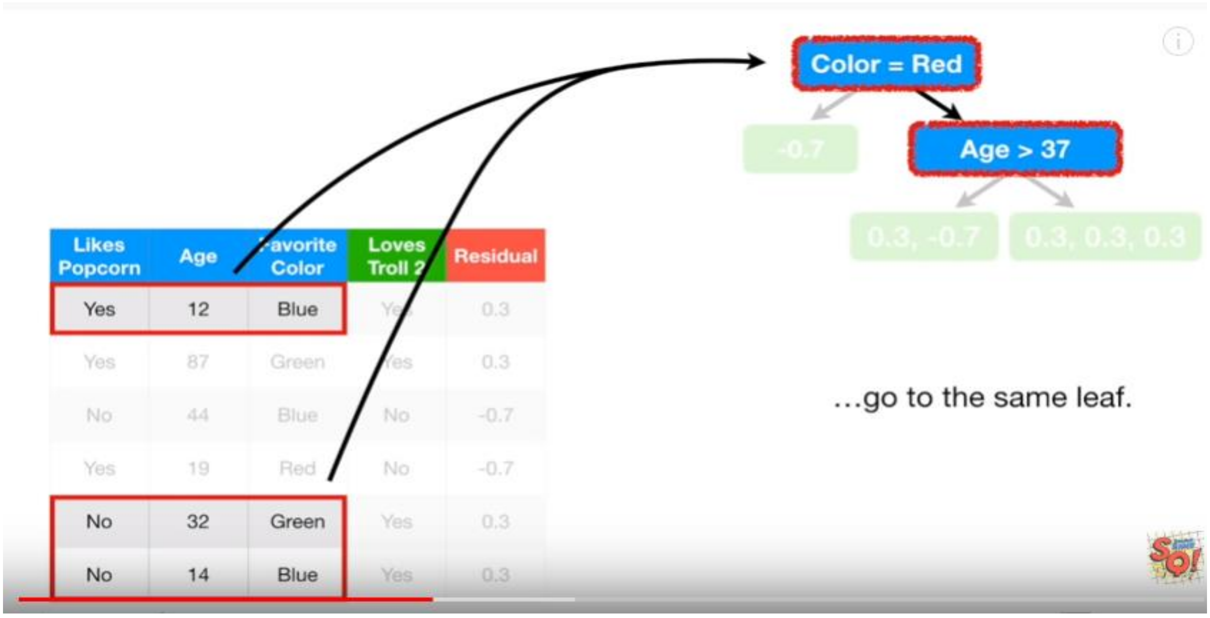



Now let's calculate the **Output Values** for the leaves.

Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 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 | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |

**NOTE:** These three rows of data…

Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 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 | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |

…go to the same leaf.


Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 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 | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |

…go to the same leaf.

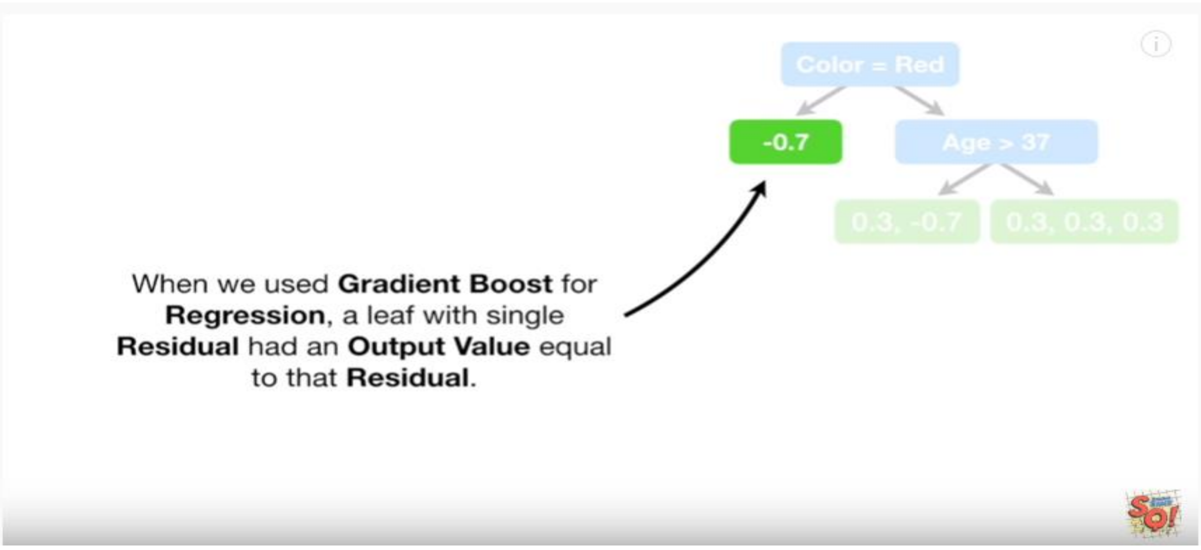

Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual
Yes | 12 | Blue | Yes | 0.3
Yes | 87 | Green | Yes | 0.3
No | 44 | Blue | No | -0.7
Yes | 19 | Red | No | -0.7
No | 32 | Green | Yes | 0.3
No | 14 | Blue | Yes | 0.3
These two rows of data…

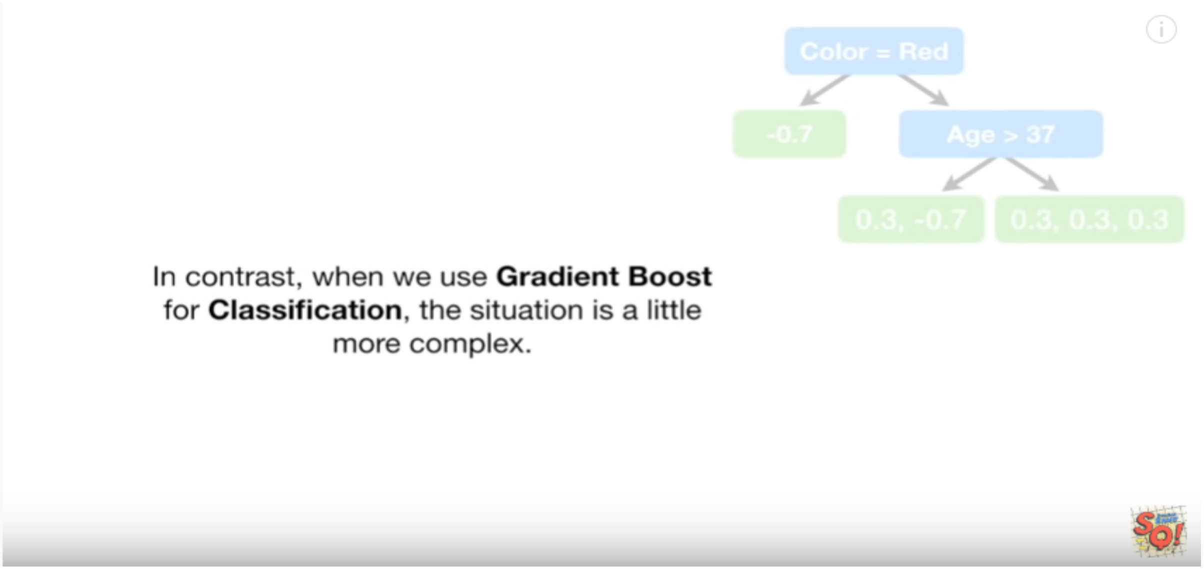

Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual
Yes | 12 | Blue | Yes | 0.3
Yes | 87 | Green | Yes | 0.3
No | 44 | Blue | No | -0.7
Yes | 19 | Red | No | -0.7
No | 32 | Green | Yes | 0.3
No | 14 | Blue | Yes | 0.3
…go to the same leaf.


Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 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 | -0.7
Yes | 19 | Red | No | -0.7
No | 32 | Green | Yes | 0.3
No | 14 | Blue | Yes | 0.3
…go to the same leaf.

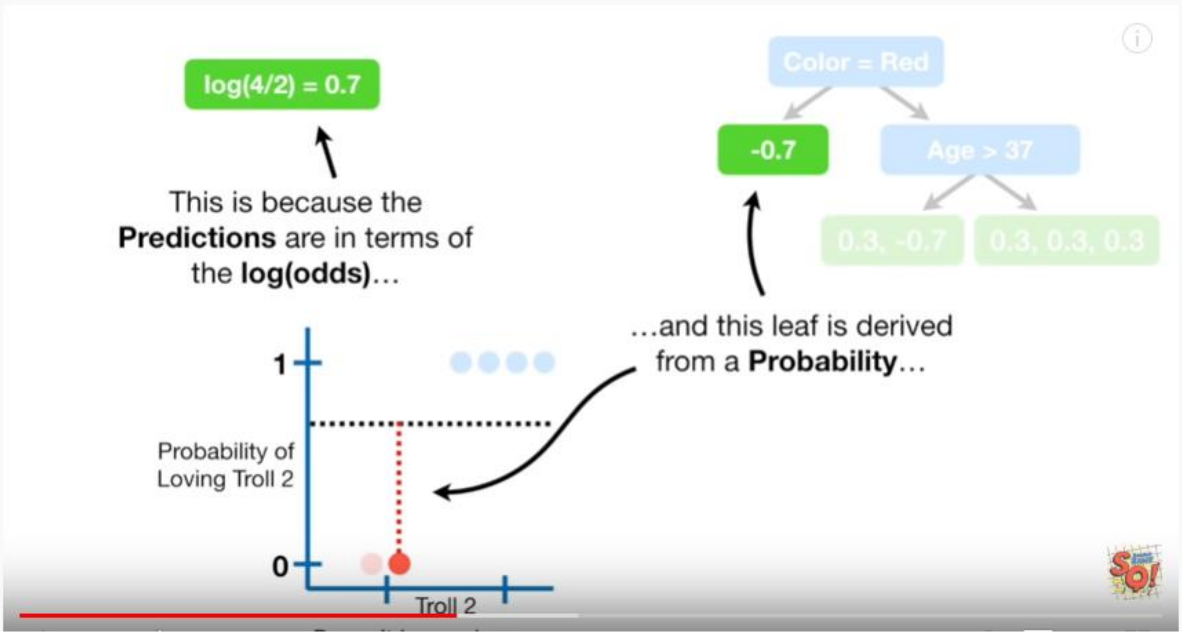

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual |
|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.3 |
| Yes | 87 | Green | Yes | 0.3 |
| No | 44 | Blue | No | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |

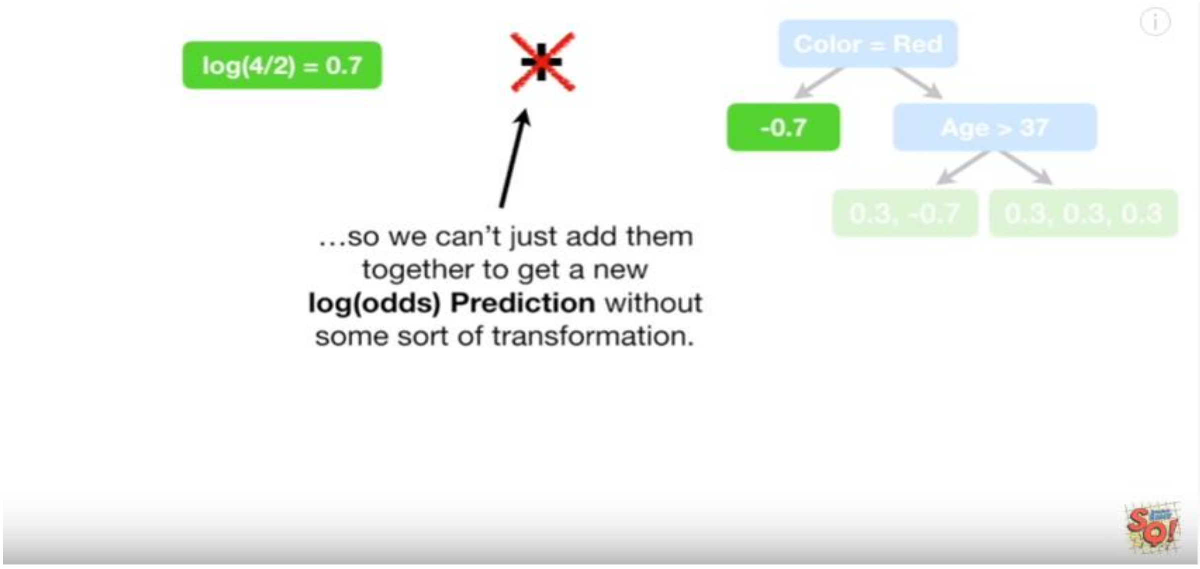

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Residual |
|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.3 |
| Yes | 87 | Green | Yes | 0.3 |
| No | 44 | Blue | No | -0.7 |
| Yes | 19 | Red | No | -0.7 |
| No | 32 | Green | Yes | 0.3 |
| No | 14 | Blue | Yes | 0.3 |

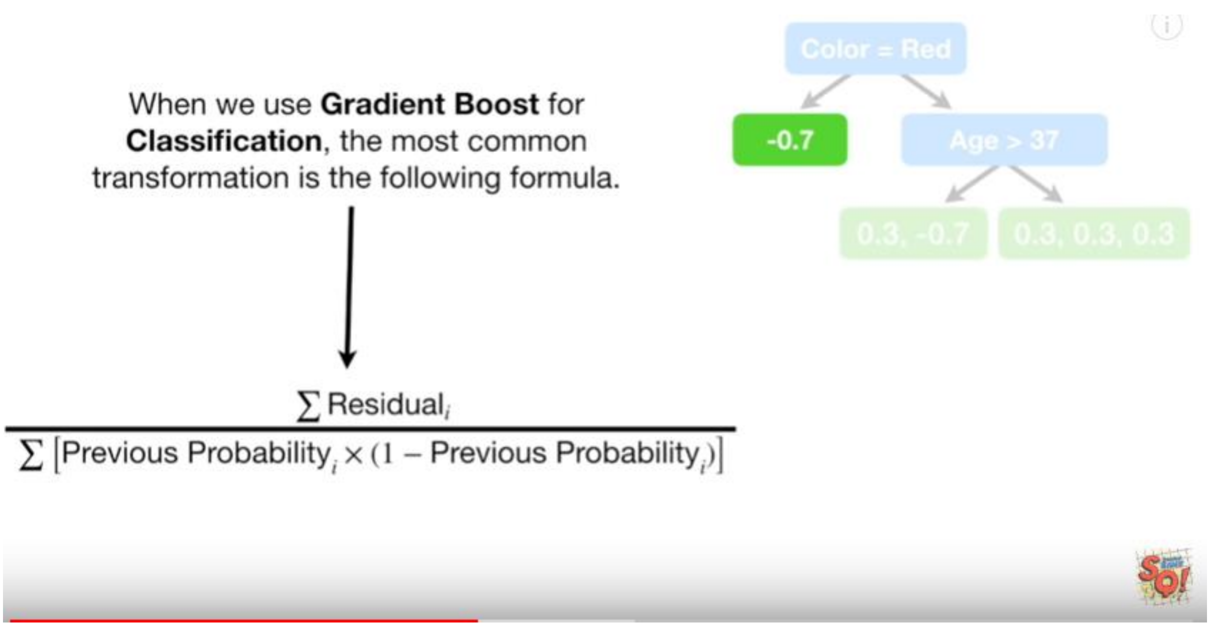


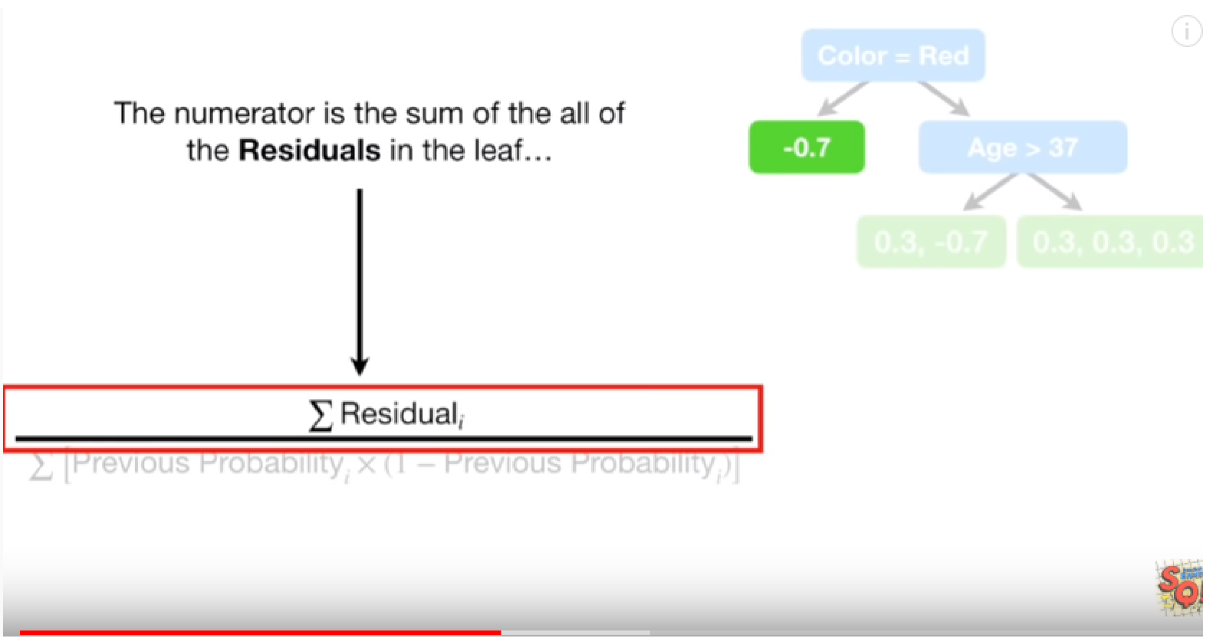

Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
In contrast, when we use **Gradient Boost** for **Classification**, the situation is a little more complex.


log(4/2) = 0.7
This is because the **Predictions** are in terms of the **log(odds)**…
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3


log(4/2) = 0.7
This is because the **Predictions** are in terms of the **log(odds)**…
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
…and this leaf is derived from a **Probability**…
1
0
Probability of Loving Troll 2
Troll 2


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.
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3

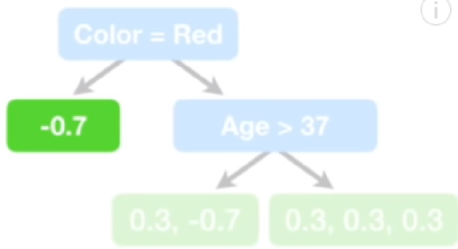

When we use **Gradient Boost** for **Classification**, the most common transformation is the following formula.
$$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$$
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3


The numerator is the sum of the all of the **Residuals** in the leaf…
$$\frac{\boxed{\sum \text{Residual}_i}}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$$
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3

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


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

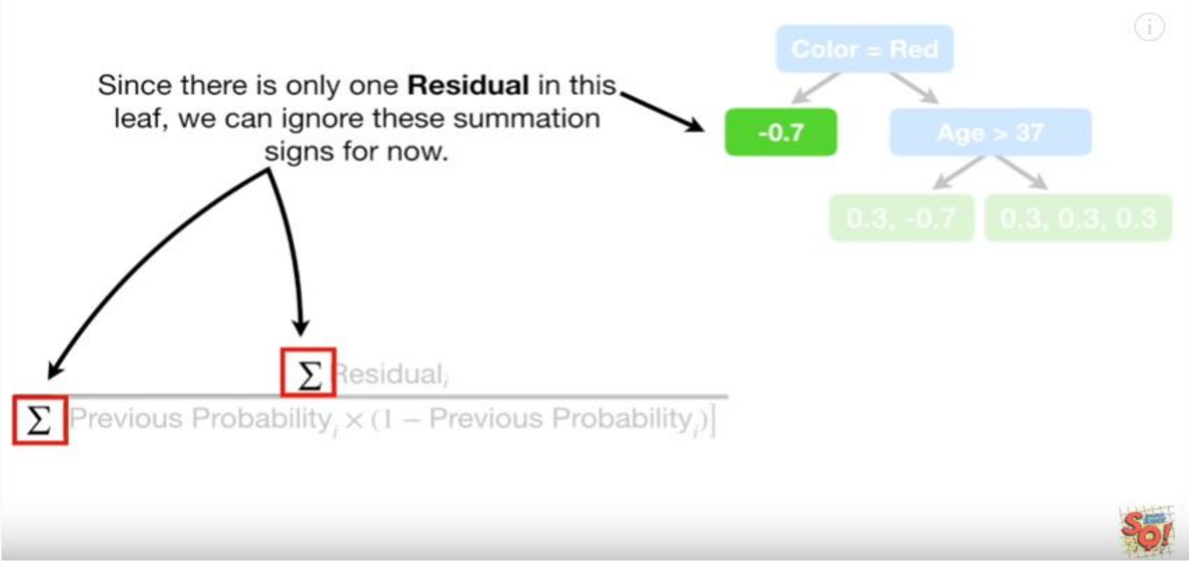
**NOTE:** The derivation of this formula is quite technical, so I'm saving it for **Part 4** of this series when we get into the nitty gritty details of **Gradient Boost for Classification**.

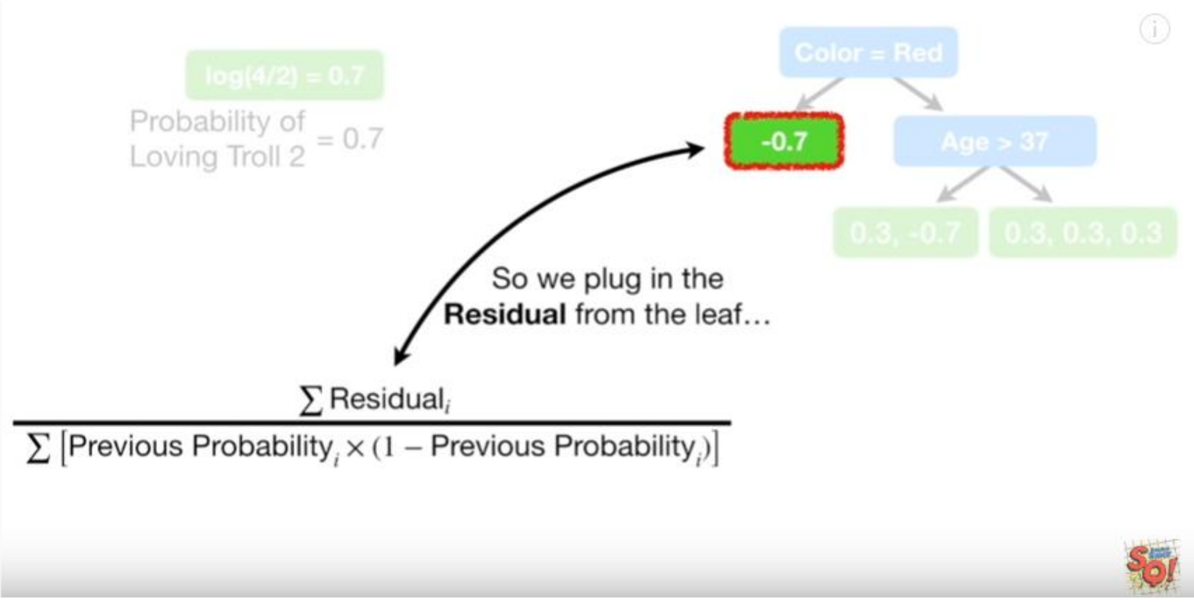


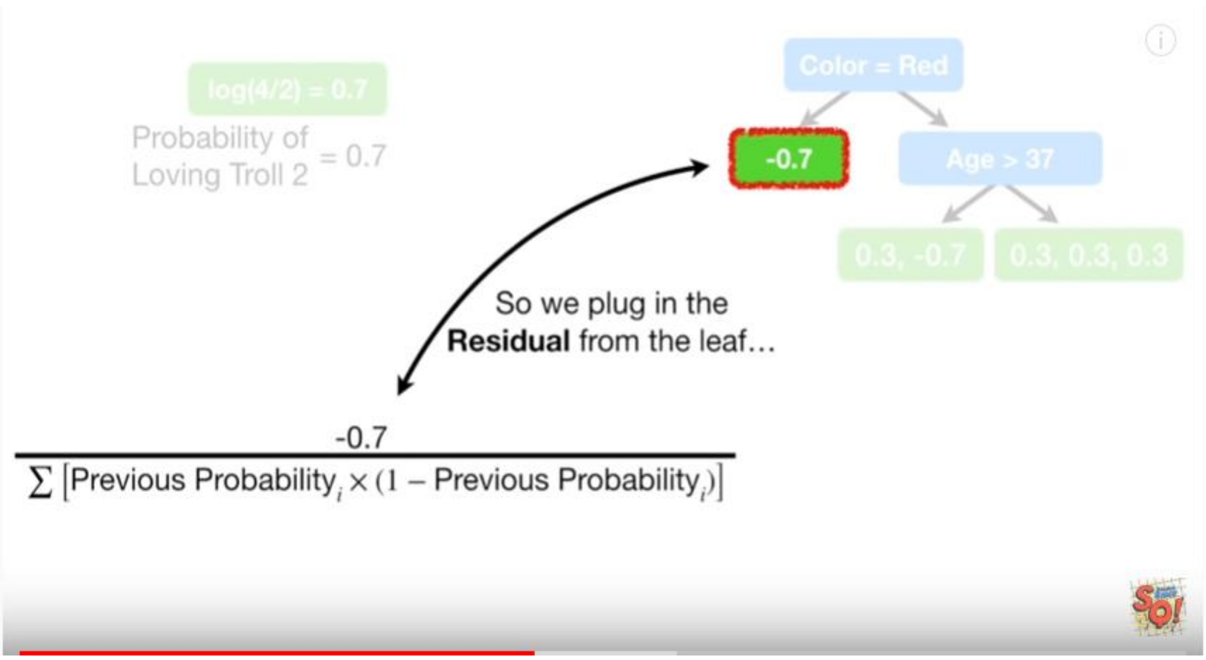

$$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$$

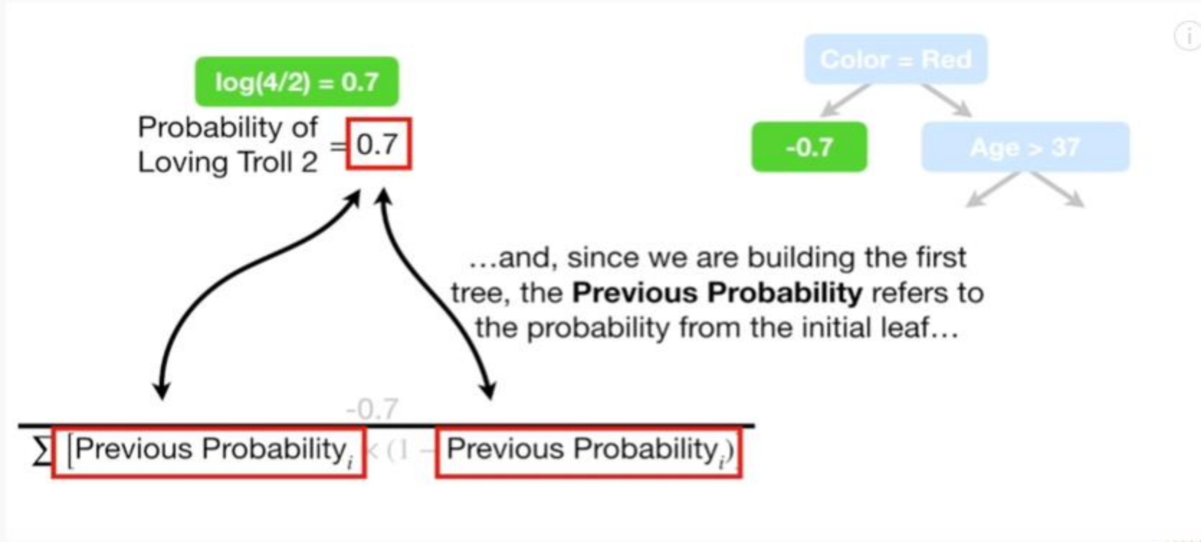
So for now, let's just use the formula to calculate the **Output Value** for this leaf.

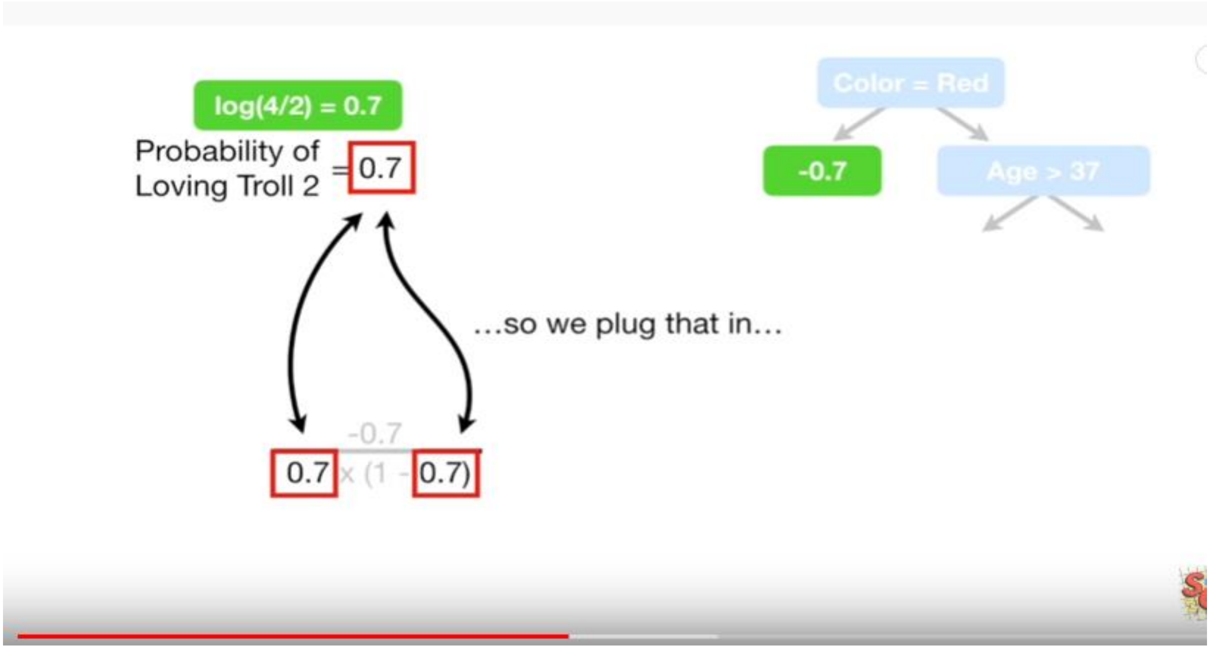

$$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$$

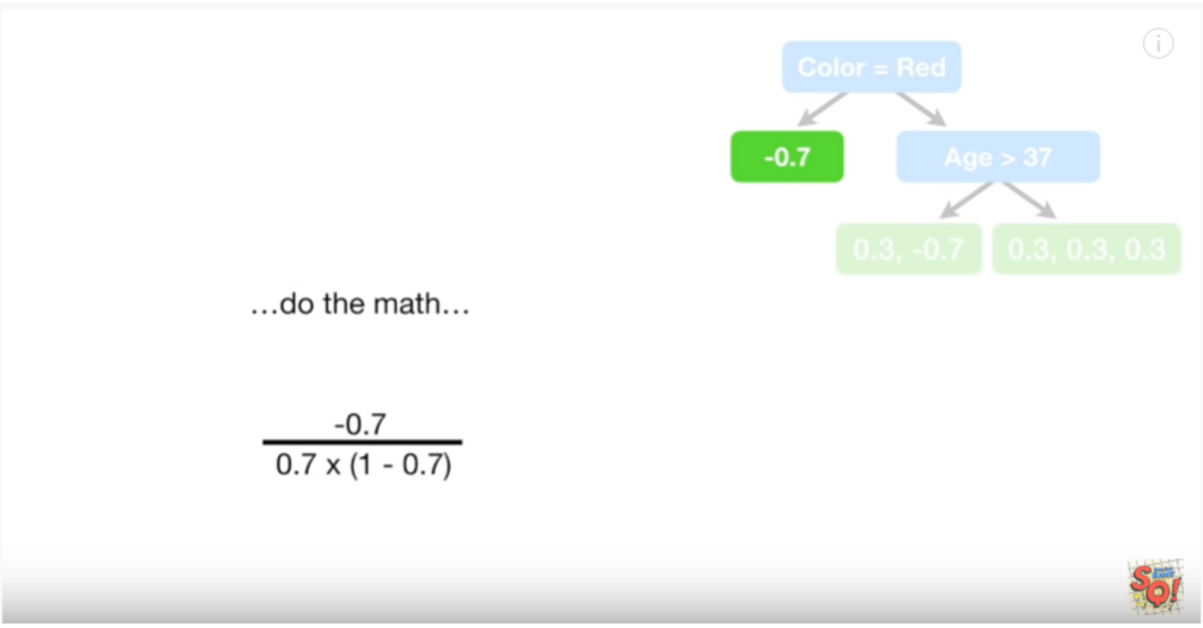

Color = Red
Since there is only one Residual in this leaf, we can ignore these summation signs for now.
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$

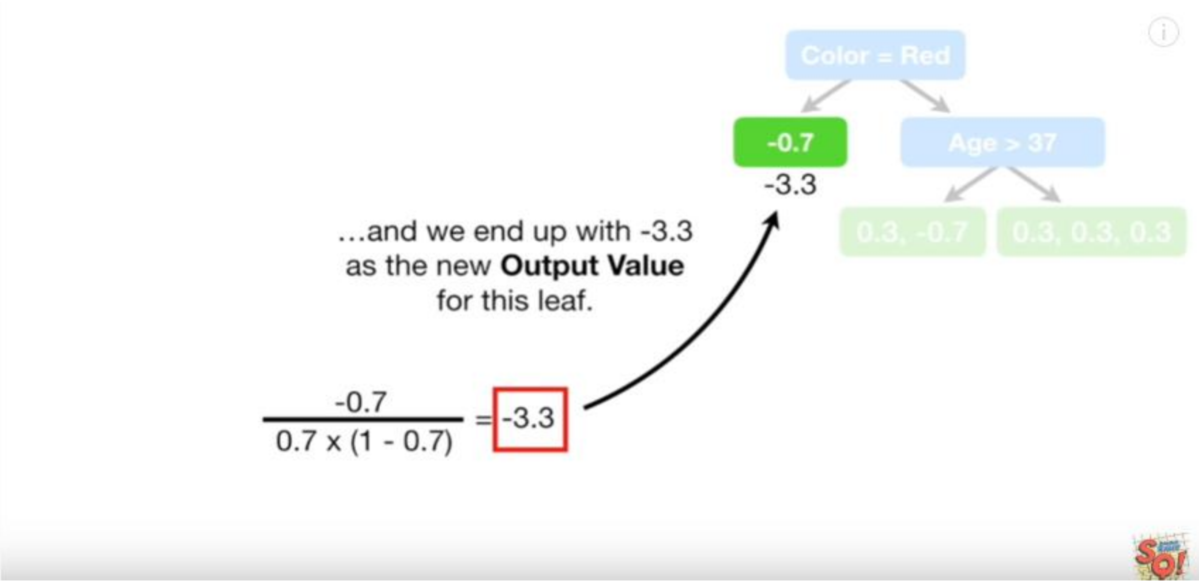

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
So we plug in the Residual from the leaf…
$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$

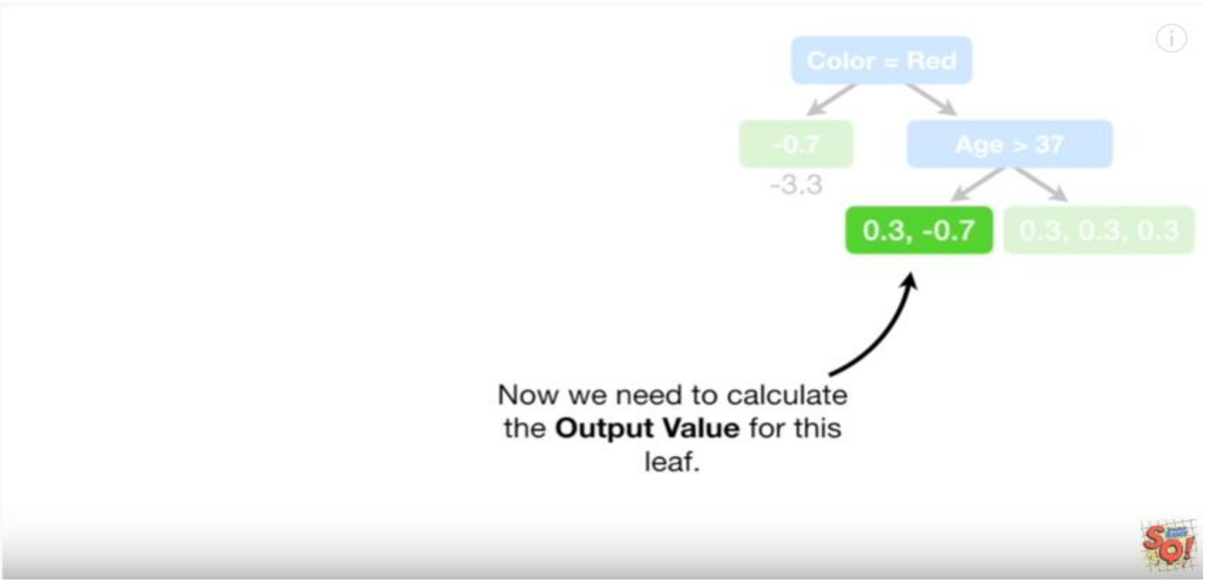

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
So we plug in the Residual from the leaf…
$\frac{-0.7}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$

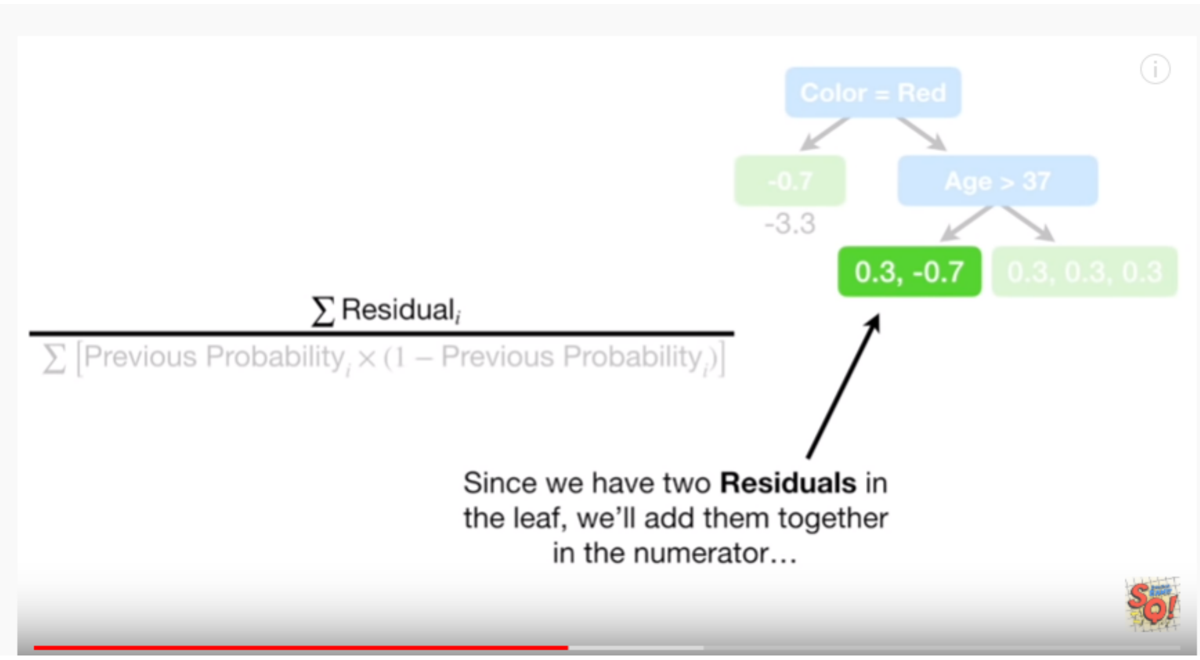

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Color = Red
-0.7
Age > 37
…and, since we are building the first tree, the **Previous Probability** refers to the probability from the initial leaf…
-0.7
∑ [Previous Probability$_i$ x (1 - Previous Probability$_i$)]

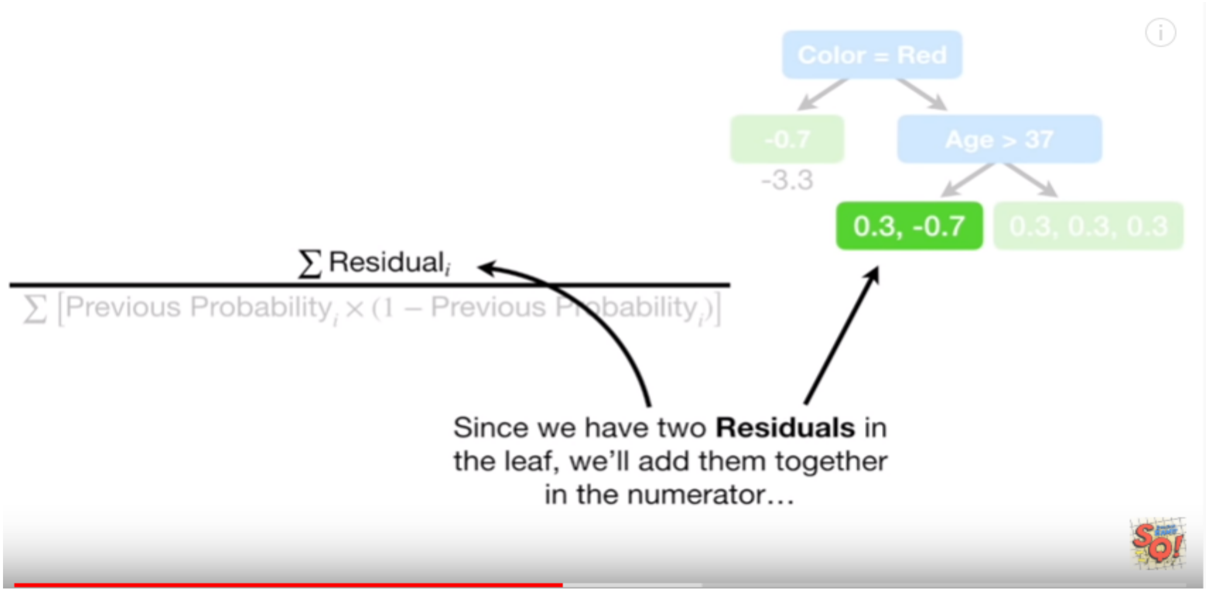

log(4/2) = 0.7
Probability of Loving Troll 2 = 0.7
Color = Red
-0.7
Age > 37
…so we plug that in…
-0.7
0.7 x (1 - 0.7)

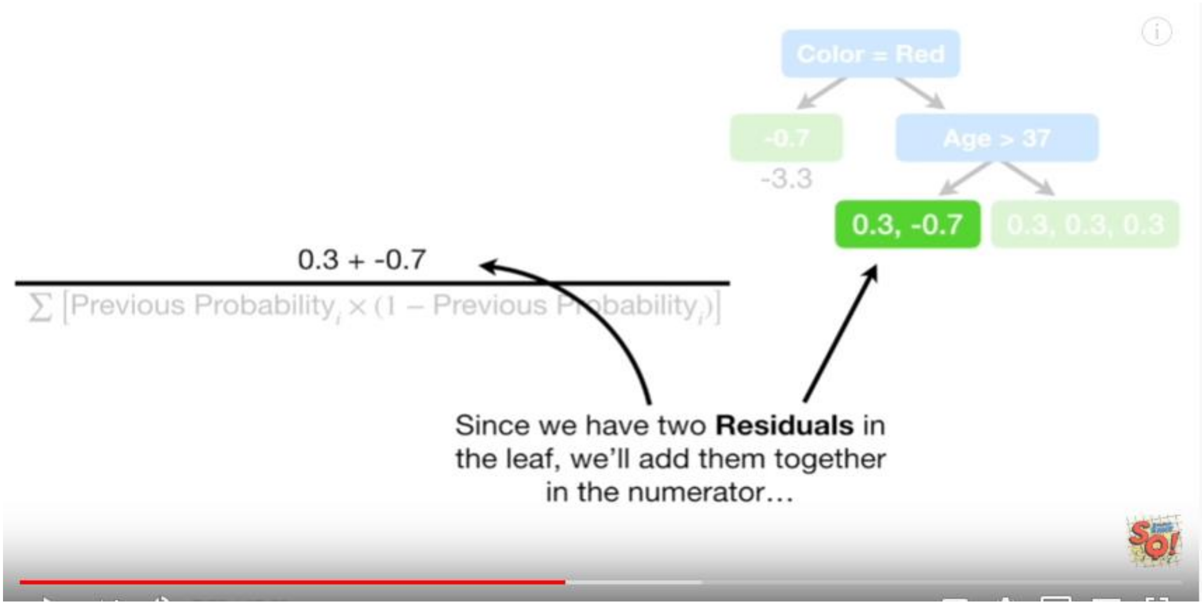

Color = Red
-0.7
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
…do the math…
-0.7
0.7 x (1 - 0.7)

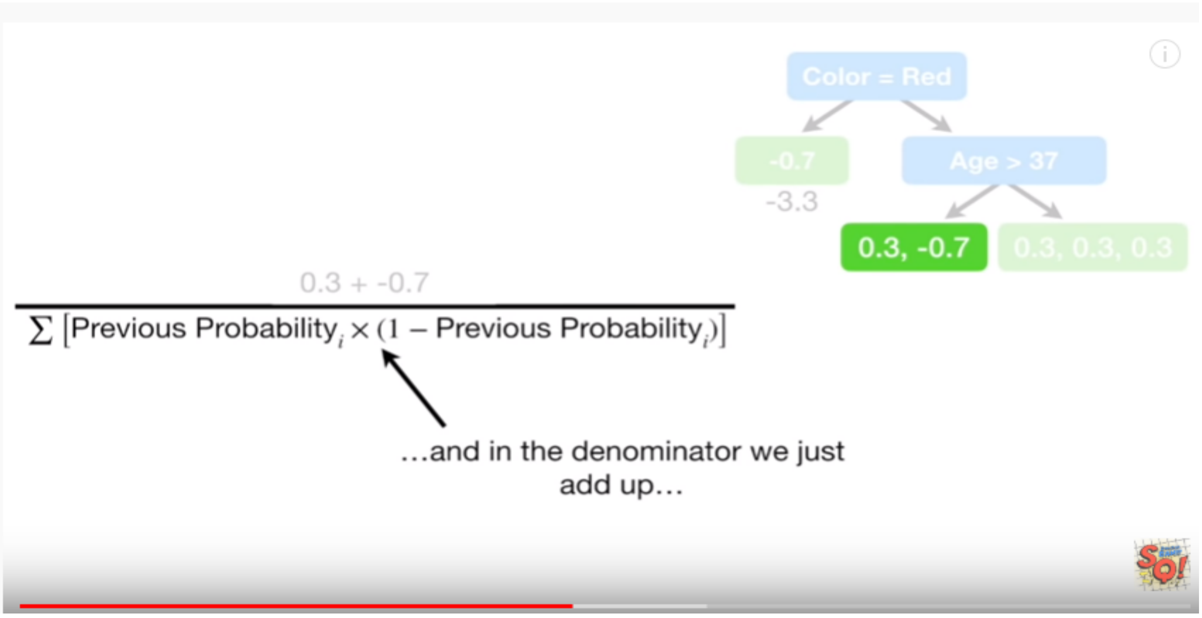

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
...and we end up with -3.3 as the new **Output Value** for this leaf.
$$\frac{-0.7}{0.7 \times (1 - 0.7)} = \boxed{-3.3}$$

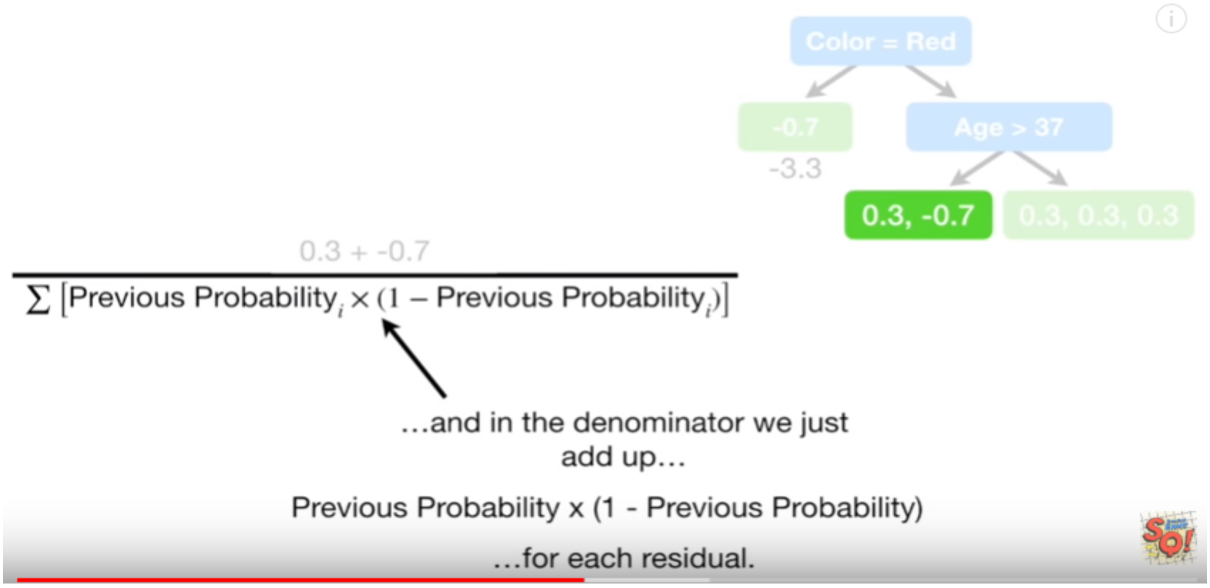

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
Now we need to calculate the **Output Value** for this leaf.

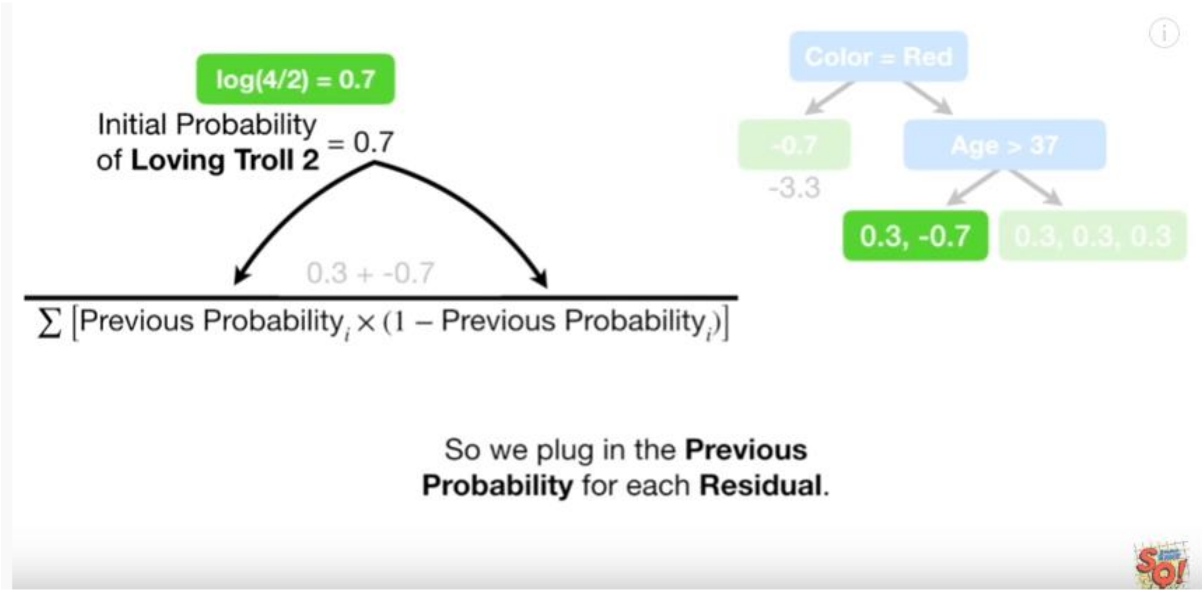

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
$$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$$
Since we have two **Residuals** in the leaf, we'll add them together in the numerator...

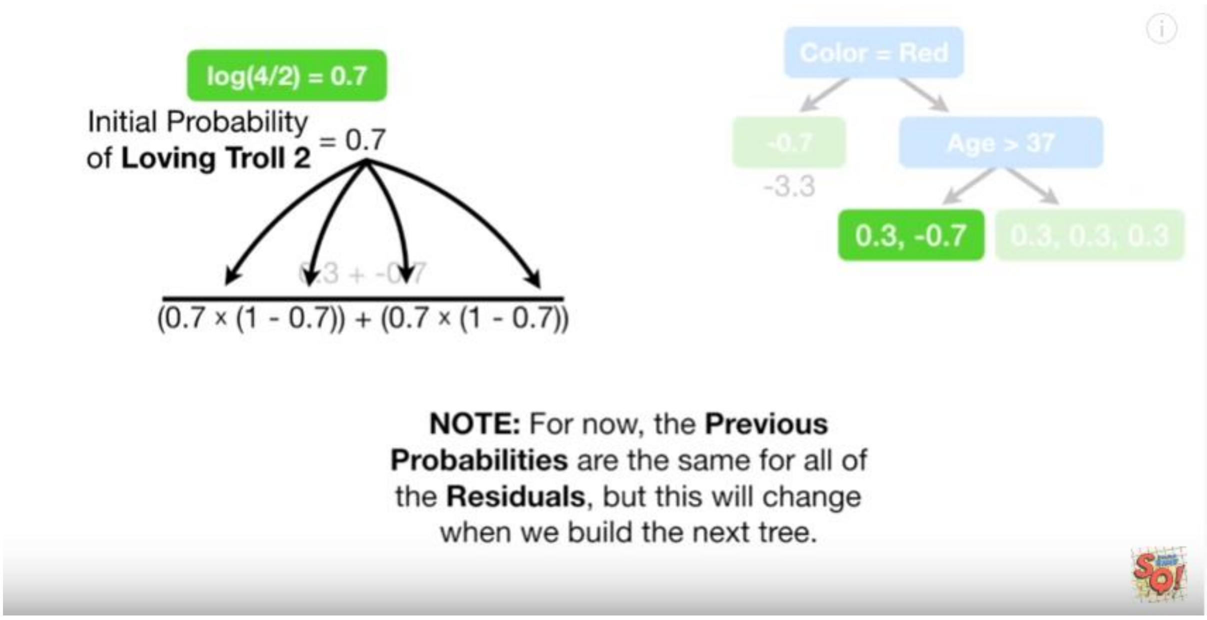

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
$\frac{\sum \text{Residual}_i}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$
Since we have two **Residuals** in the leaf, we'll add them together in the numerator…

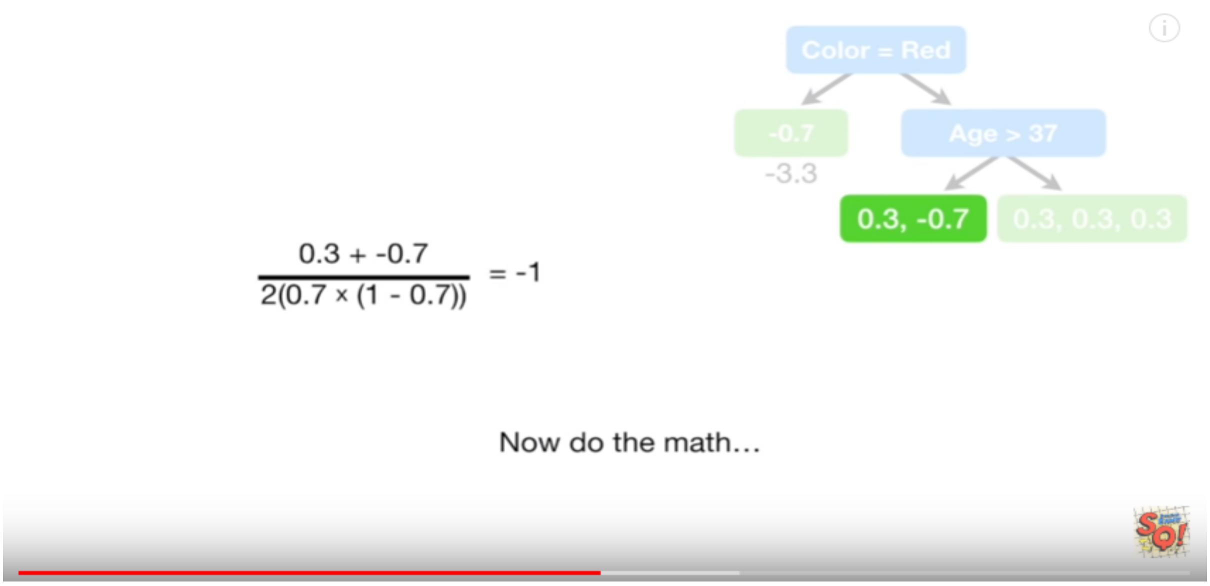

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
$\frac{0.3 + -0.7}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$
Since we have two **Residuals** in the leaf, we'll add them together in the numerator…

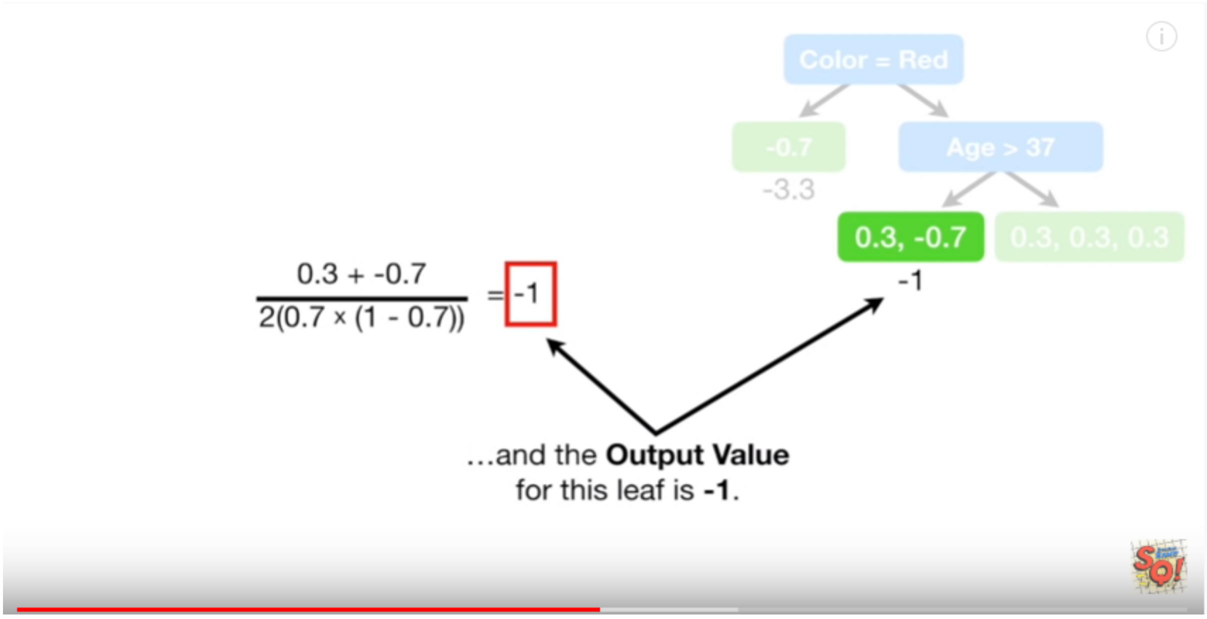

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
$\frac{0.3 + -0.7}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)]}$
…and in the denominator we just add up…

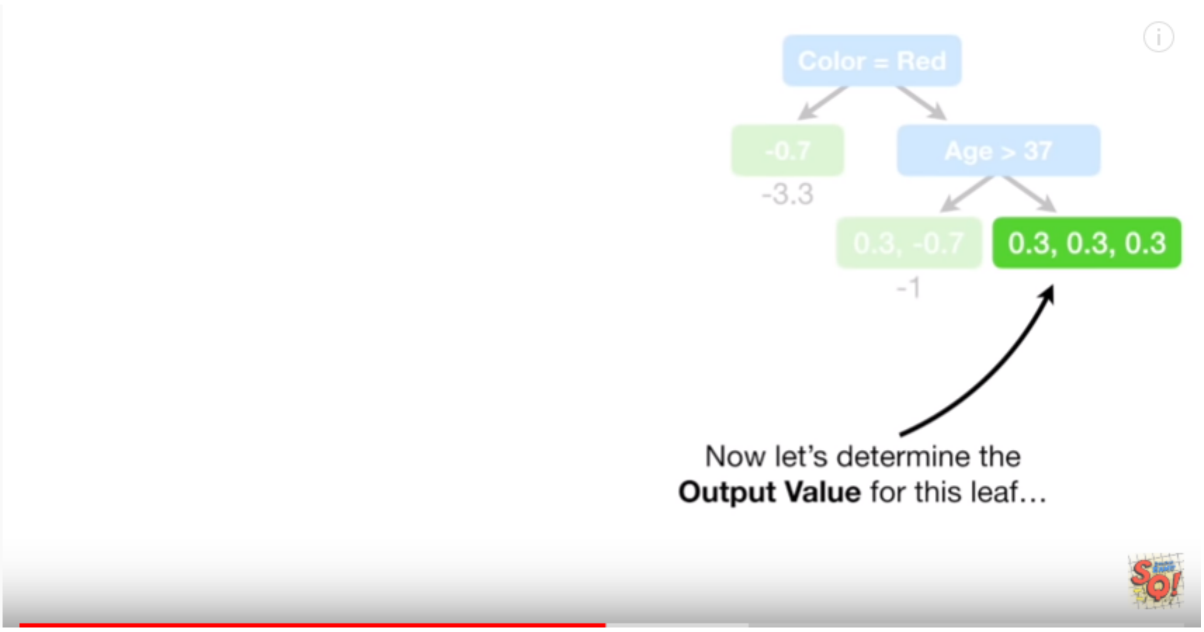

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
0.3 + -0.7
$\sum$ [Previous Probability$_i$ × (1 – Previous Probability$_i$)]
…and in the denominator we just add up…
Previous Probability x (1 - Previous Probability)
…for each residual.

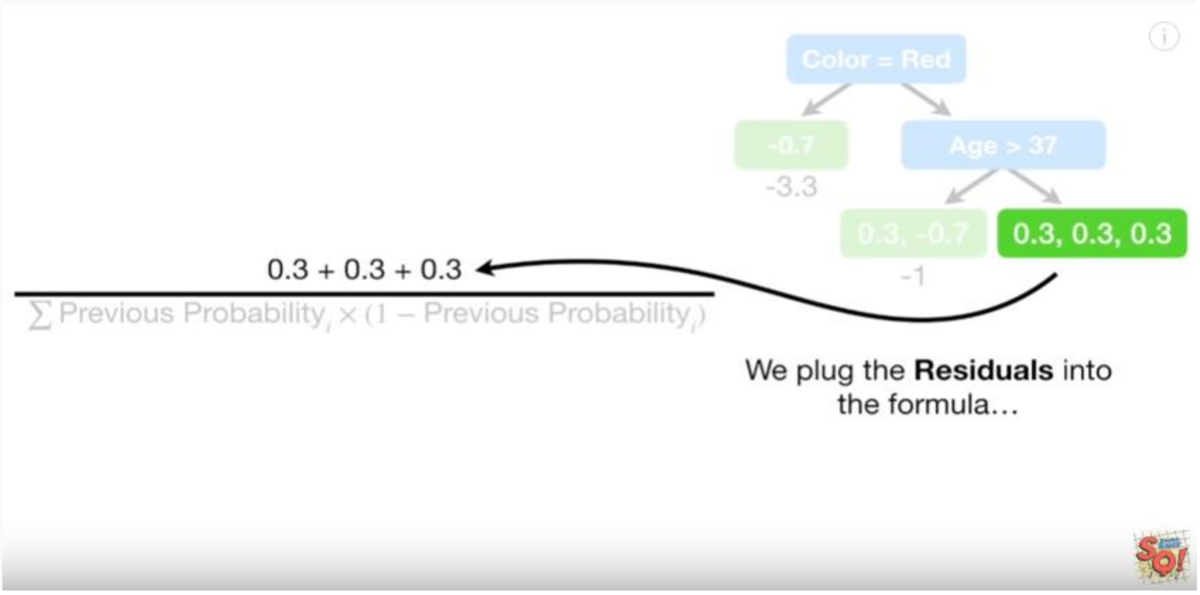

log(4/2) = 0.7
Initial Probability of **Loving Troll 2** = 0.7
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
0.3 + -0.7
$\sum$ [Previous Probability$_i$ × (1 – Previous Probability$_i$)]
So we plug in the **Previous Probability** for each **Residual**.


log(4/2) = 0.7
Initial Probability of **Loving Troll 2** = 0.7
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
0.3 + -0.7
(0.7 × (1 - 0.7)) + (0.7 × (1 - 0.7))
**NOTE:** For now, the **Previous Probabilities** are the same for all of the **Residuals**, but this will change when we build the next tree.

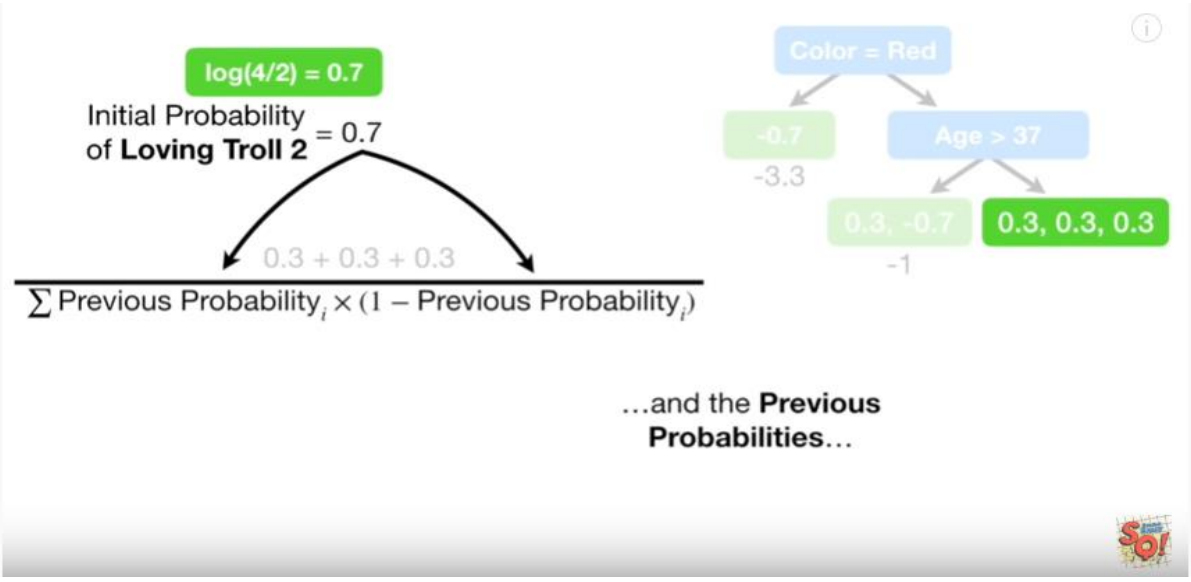

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
$\frac{0.3 + -0.7}{2(0.7 \times (1 - 0.7))} = -1$
Now do the math…


Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
-1
$\frac{0.3 + -0.7}{2(0.7 \times (1 - 0.7))} = \boxed{-1}$
…and the **Output Value**
for this leaf is **-1**.

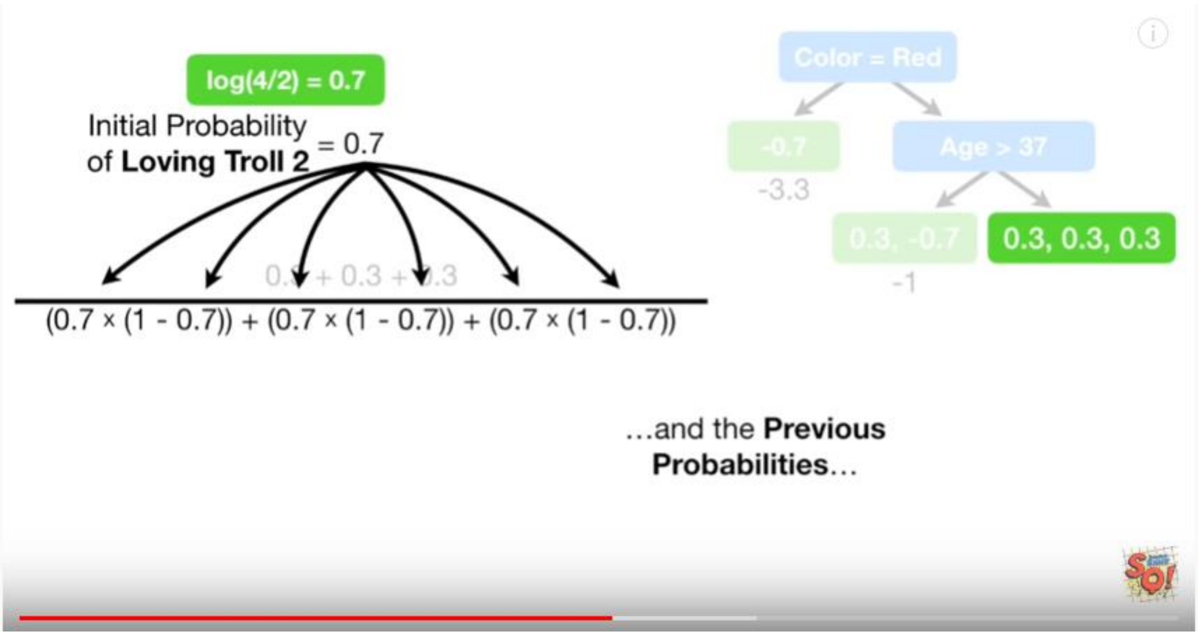

Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
-1
Now let's determine the
**Output Value** for this leaf…



Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
0.3 + 0.3 + 0.3
∑ Previous Probability$_i$ × (1 – Previous Probability$_i$)
We plug the **Residuals** into the formula…

log(4/2) = 0.7
Initial Probability of **Loving Troll 2** = 0.7
0.3 + 0.3 + 0.3
∑ Previous Probability$_i$ × (1 – Previous Probability$_i$)
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
…and the **Previous Probabilities**…

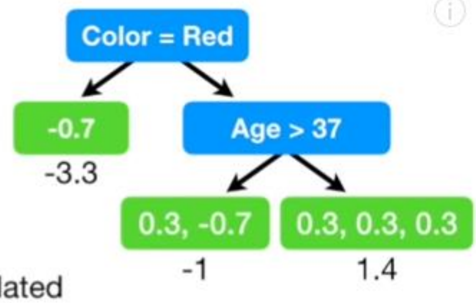
log(4/2) = 0.7
Initial Probability of **Loving Troll 2** = 0.7
0.3 + 0.3 + 0.3
(0.7 × (1 - 0.7)) + (0.7 × (1 - 0.7)) + (0.7 × (1 - 0.7))
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
…and the **Previous Probabilities**…


$$\frac{3 \times 0.3}{3 \times (0.7 \times (1 - 0.7))}$$

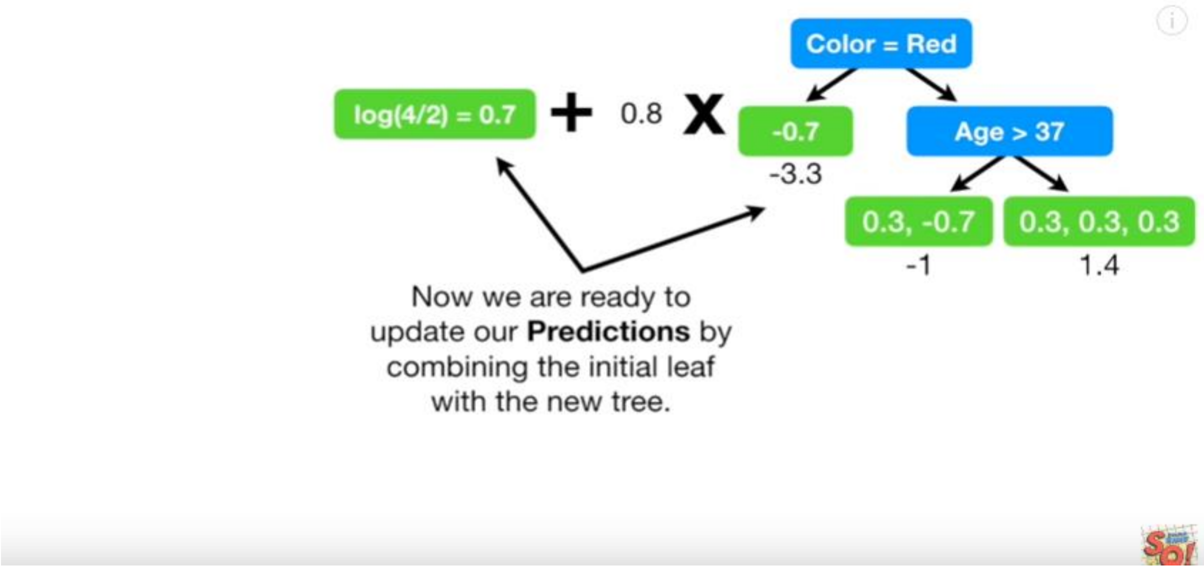
…and do the math…

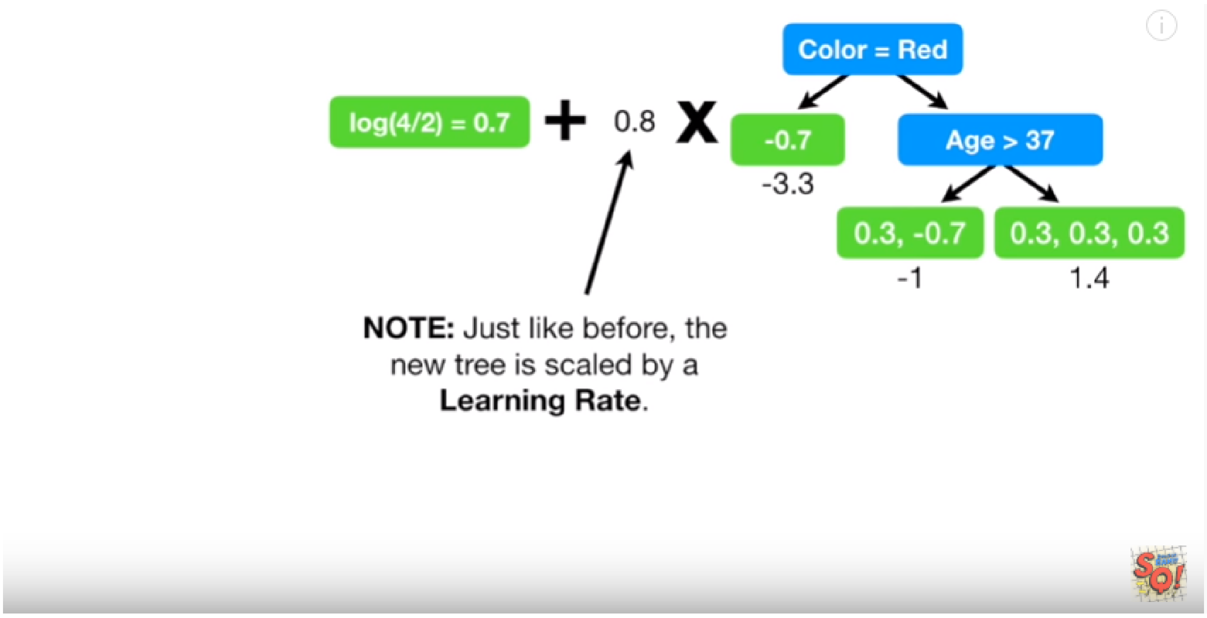

$$\frac{0.3}{(0.7 \times (1 - 0.7))} = \boxed{1.4}$$

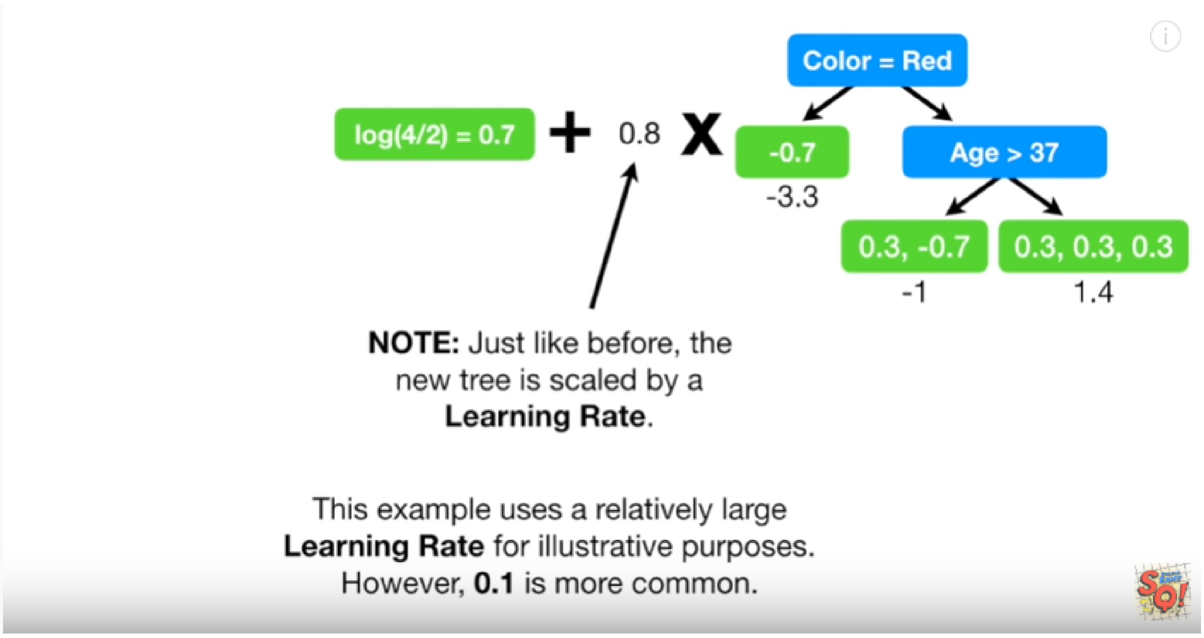
…and the **Output Value** for this leaf is **1.4**.

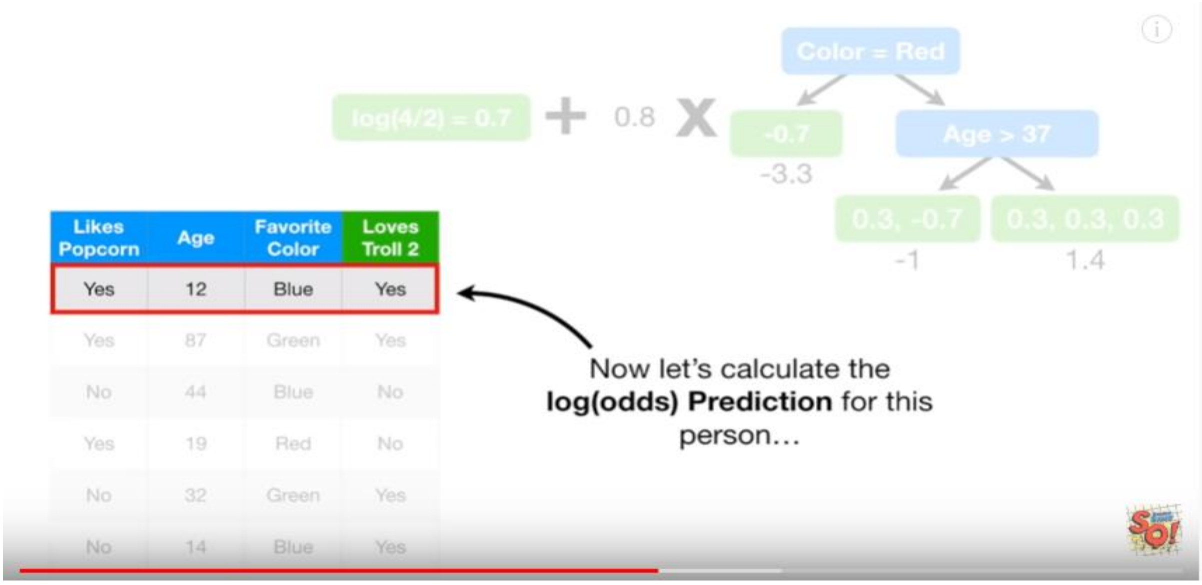


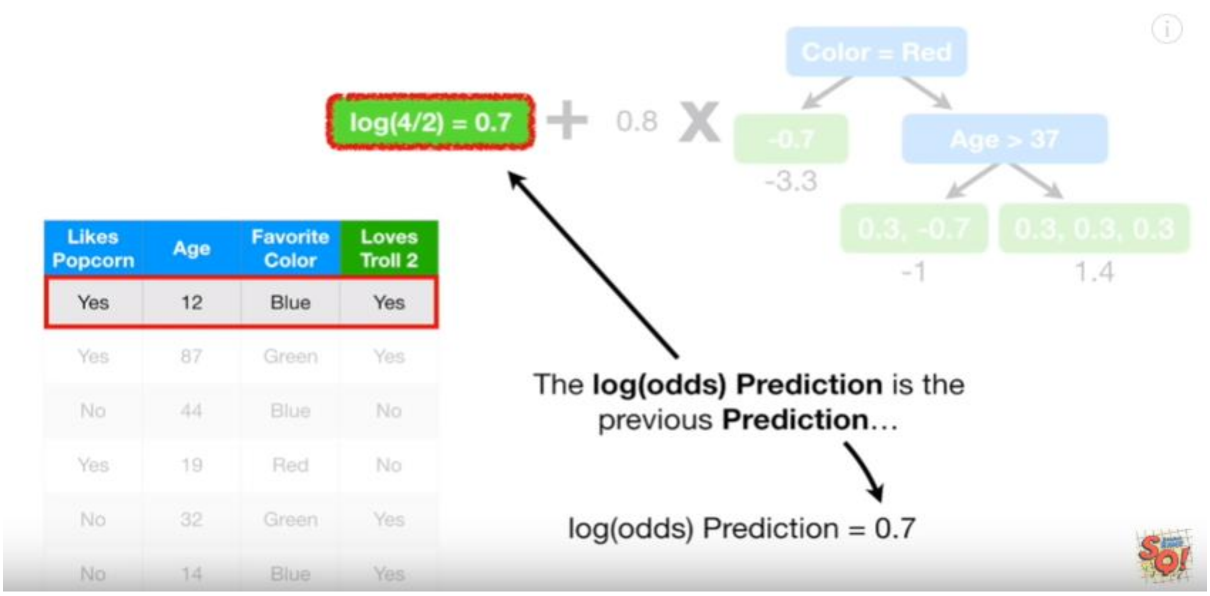

Hooray!!! We've calculated **Output Values** for all three leaves in the tree!

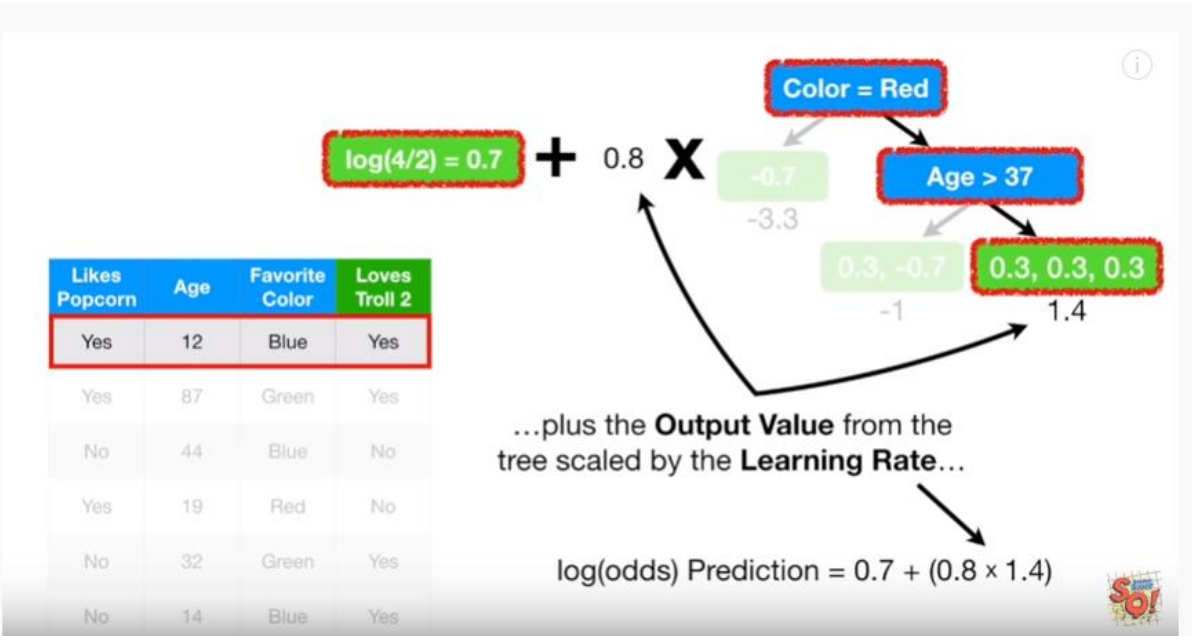

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Now we are ready to update our Predictions by combining the initial leaf with the new tree.

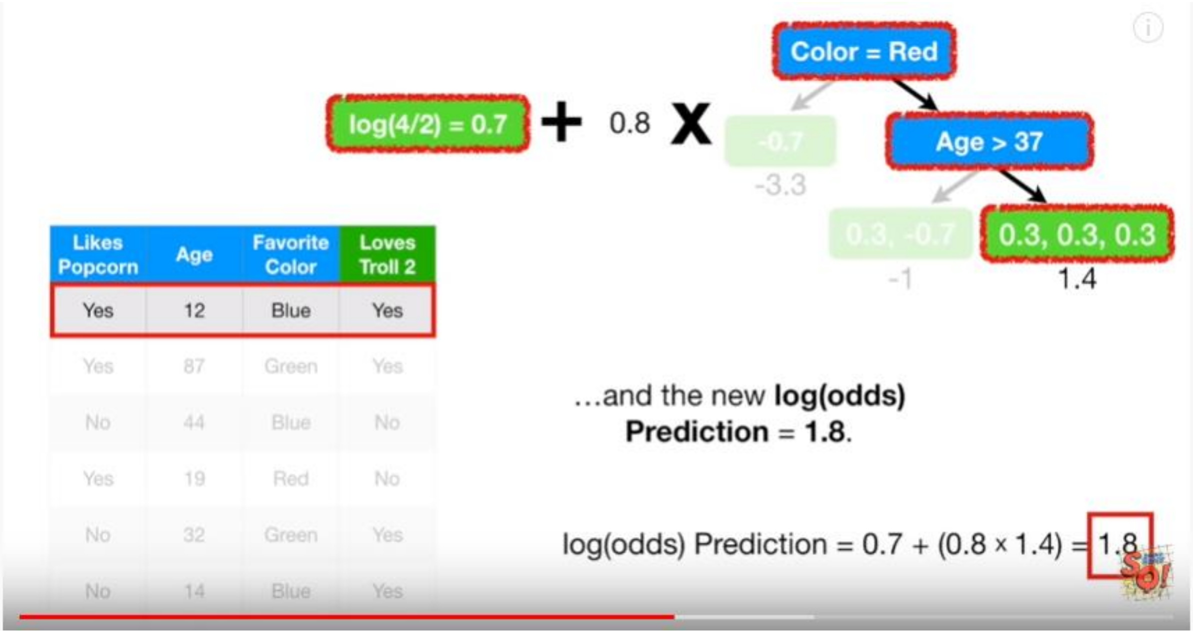

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
NOTE: Just like before, the new tree is scaled by a Learning Rate.


Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
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.

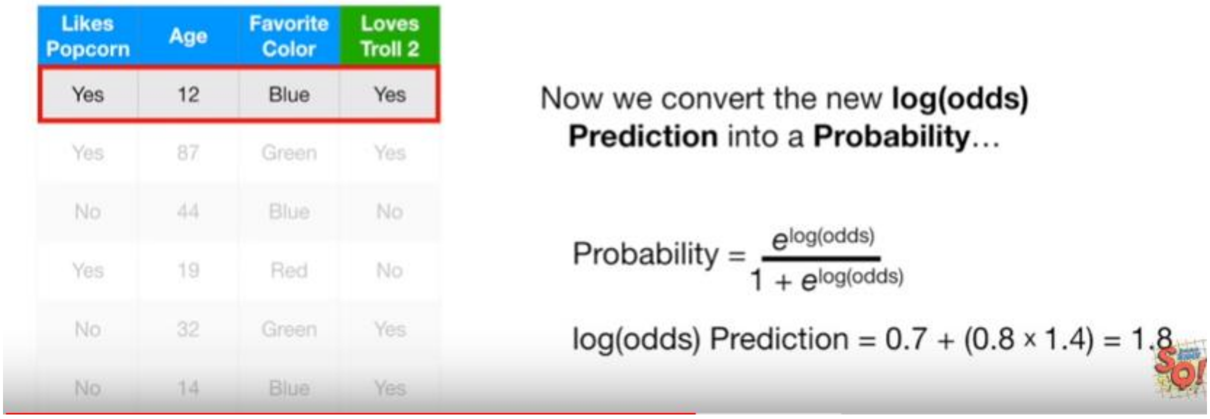

Color = Red
log(4/2) = 0.7
0.8
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now let's calculate the **log(odds) Prediction** for this person…


Color = Red
log(4/2) = 0.7
0.8
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
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

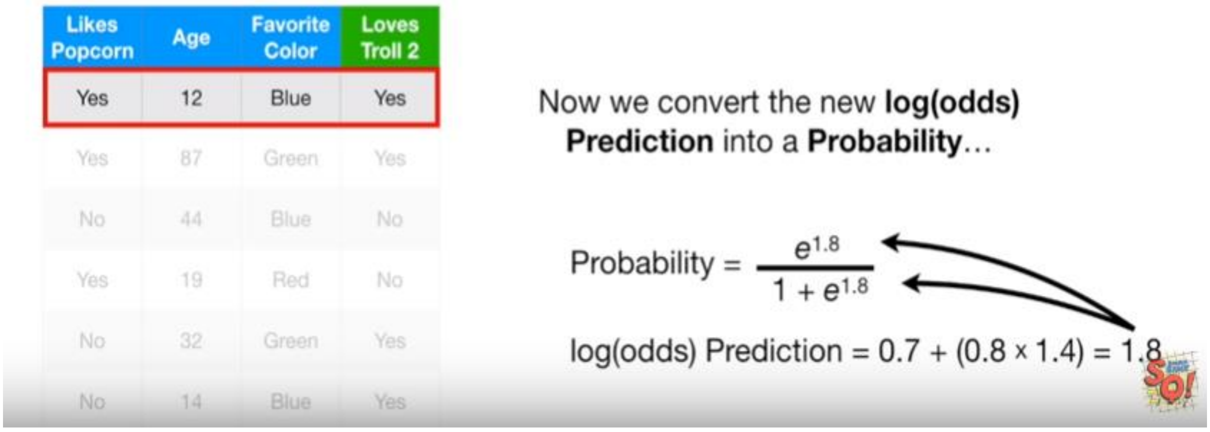

Color = Red
log(4/2) = 0.7
0.8
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
…plus the **Output Value** from the tree scaled by the **Learning Rate**…
log(odds) Prediction = 0.7 + (0.8 × 1.4)



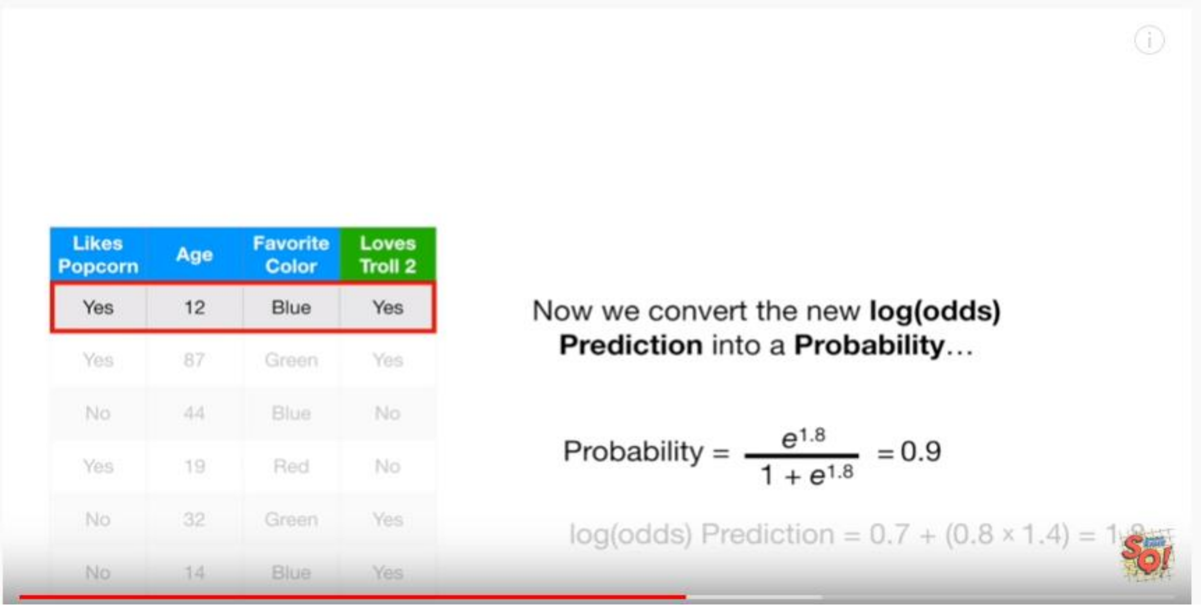
Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
...and the new **log(odds)** **Prediction** = **1.8**.
log(odds) Prediction = 0.7 + (0.8 × 1.4) = 1.8

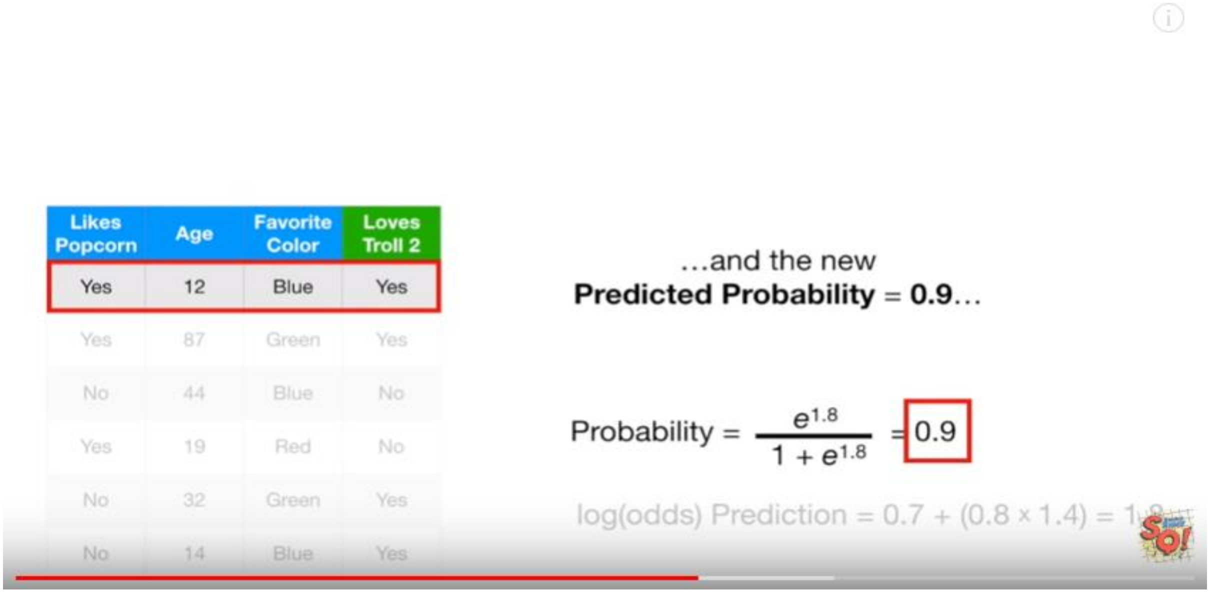
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now we convert the new **log(odds)** **Prediction** into a **Probability**...
$\text{Probability} = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$
log(odds) Prediction = 0.7 + (0.8 × 1.4) = 1.8

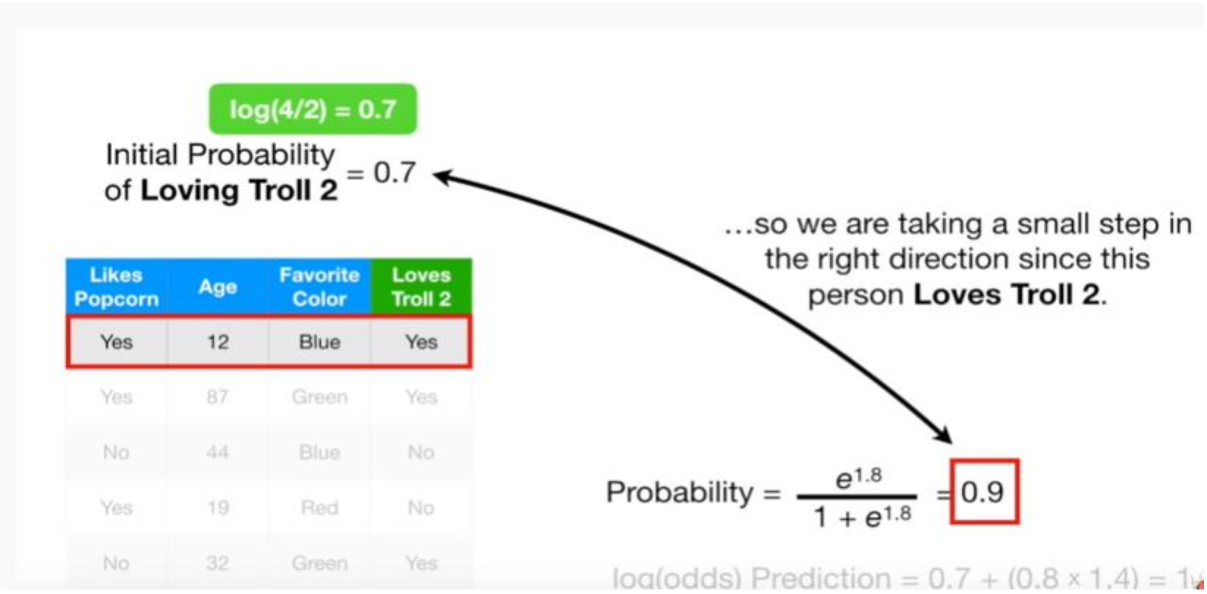
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now we convert the new **log(odds)** **Prediction** into a **Probability**...
$\text{Probability} = \frac{e^{1.8}}{1 + e^{1.8}}$
log(odds) Prediction = 0.7 + (0.8 × 1.4) = 1.8



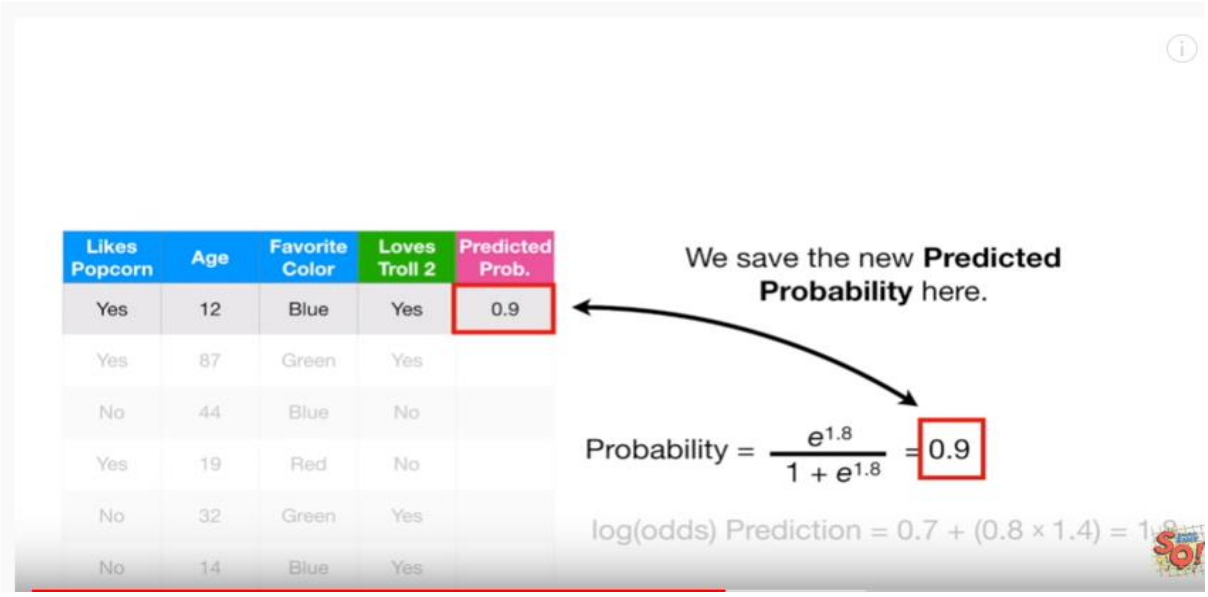
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now we convert the new log(odds) Prediction into a Probability…
Probability = e^1.8 / (1 + e^1.8) = 0.9
log(odds) Prediction = 0.7 + (0.8 x 1.4) = 1.8

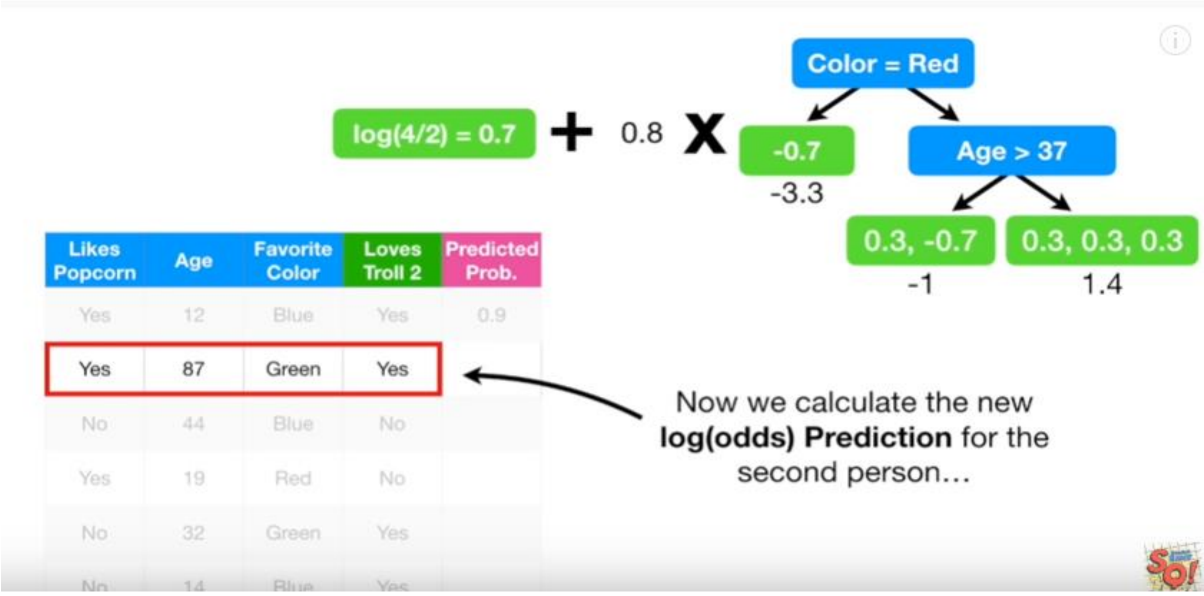



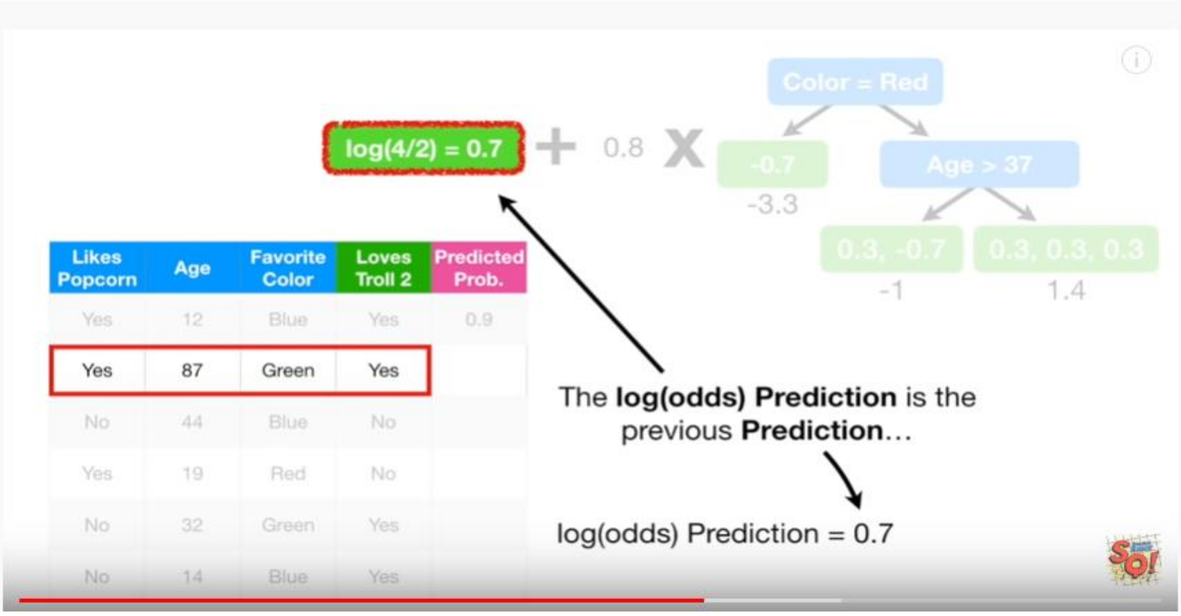
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
…and the new Predicted Probability = 0.9…
Probability = e^1.8 / (1 + e^1.8) = 0.9
log(odds) Prediction = 0.7 + (0.8 x 1.4) = 1.8

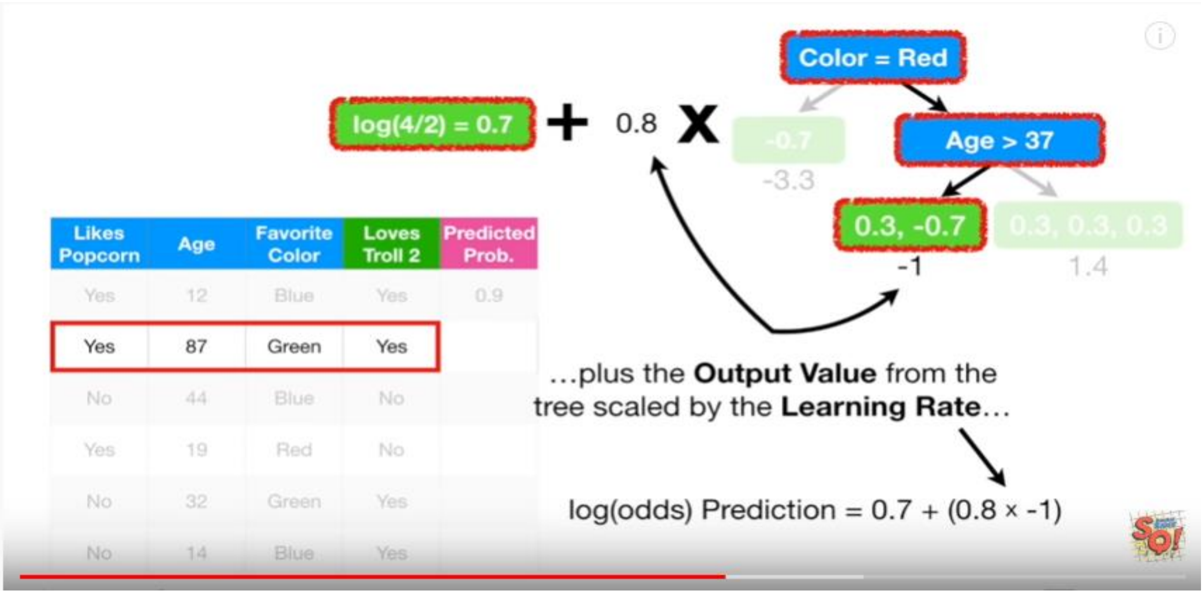



log(4/2) = 0.7
Initial Probability of Loving Troll 2 = 0.7
…so we are taking a small step in the right direction since this person Loves Troll 2.
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 12 | Blue | Yes
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
Probability = e^1.8 / (1 + e^1.8) = 0.9
log(odds) Prediction = 0.7 + (0.8 x 1.4) = 1

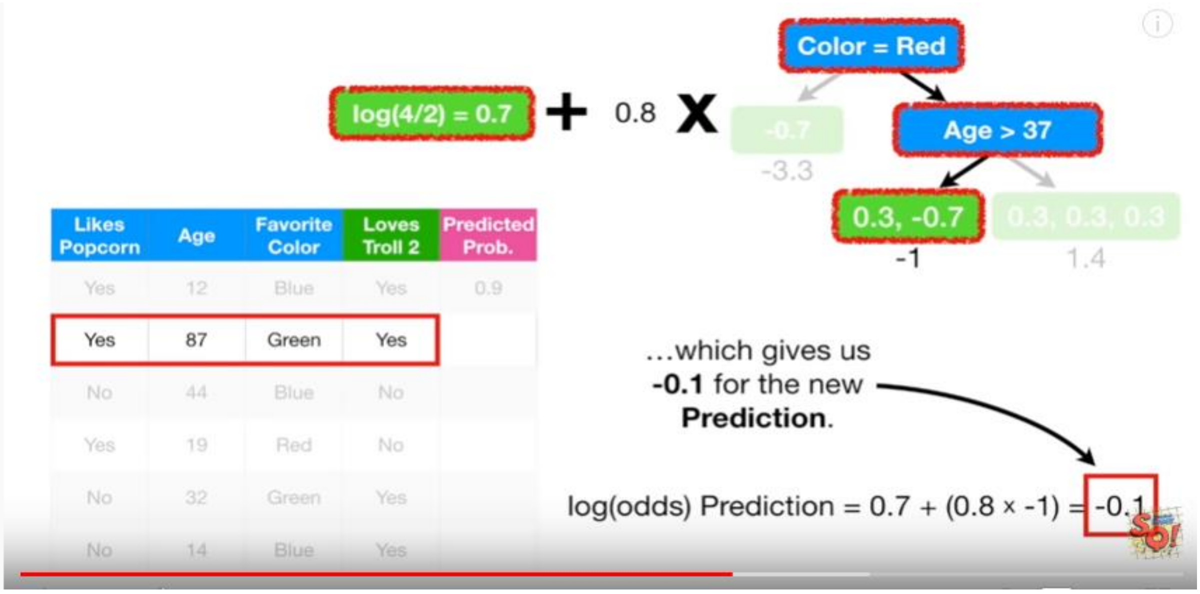

Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
We save the new Predicted Probability here.
Probability = $\frac{e^{1.8}}{1 + e^{1.8}}$ = 0.9
log(odds) Prediction = 0.7 + (0.8 × 1.4) = 1.8


log(4/2) = 0.7 + 0.8 X
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
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
Now we calculate the new log(odds) Prediction for the second person…

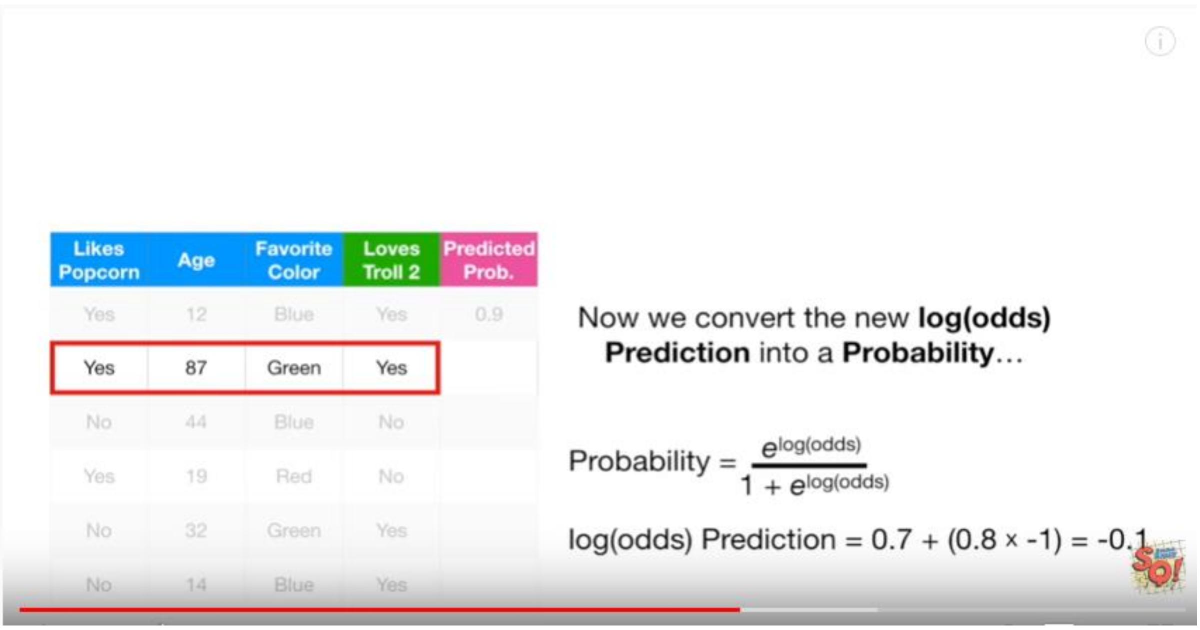

log(4/2) = 0.7 + 0.8 X
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
The log(odds) Prediction is the previous Prediction…
log(odds) Prediction = 0.7



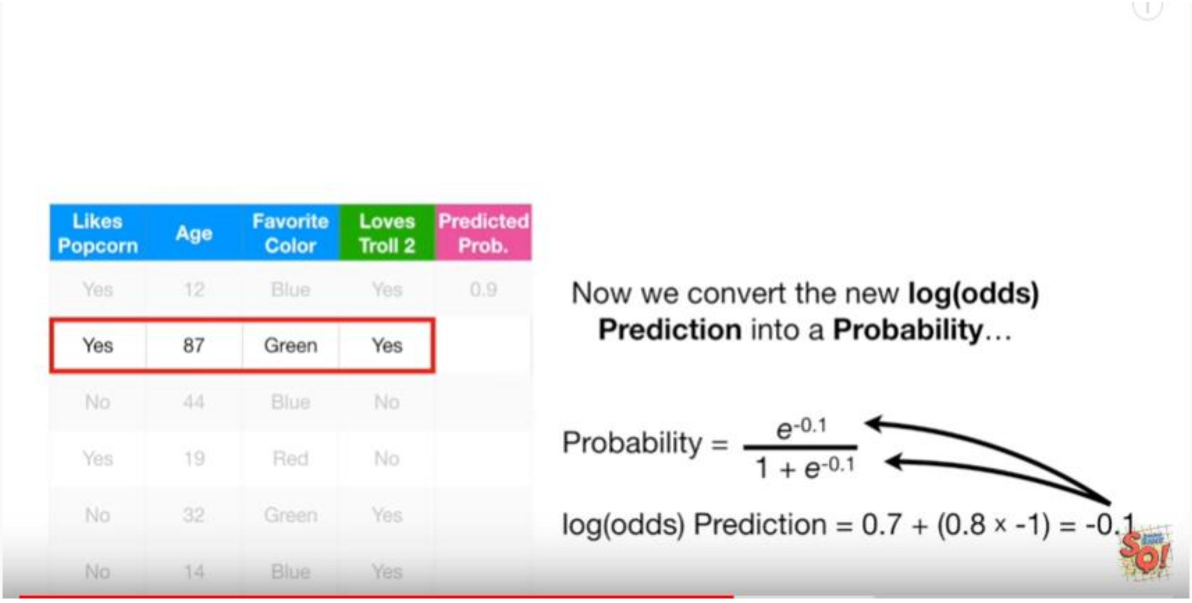
Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
…plus the **Output Value** from the tree scaled by the **Learning Rate**…
log(odds) Prediction = 0.7 + (0.8 × -1)

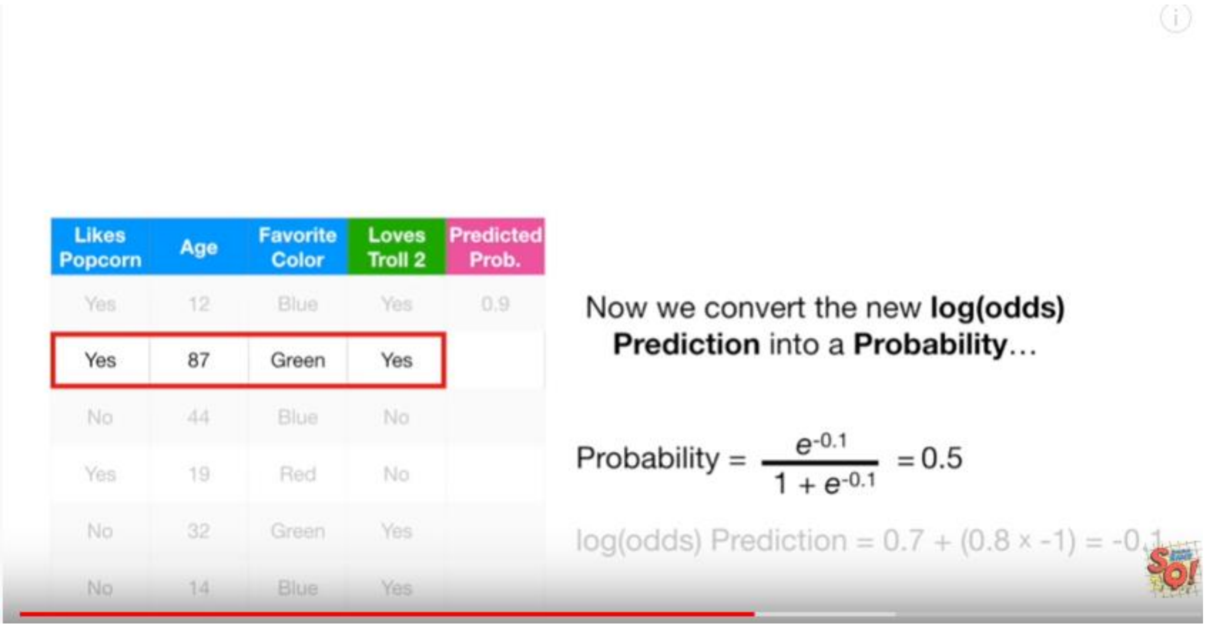



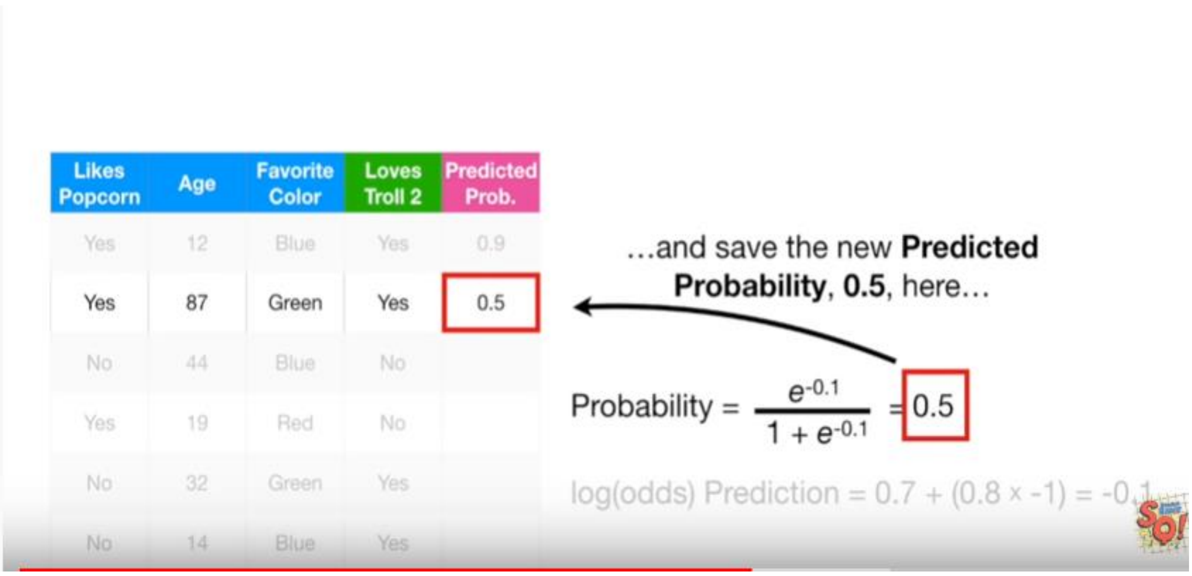
Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
…which gives us **-0.1** for the new **Prediction**.
log(odds) Prediction = 0.7 + (0.8 × -1) = -0.1

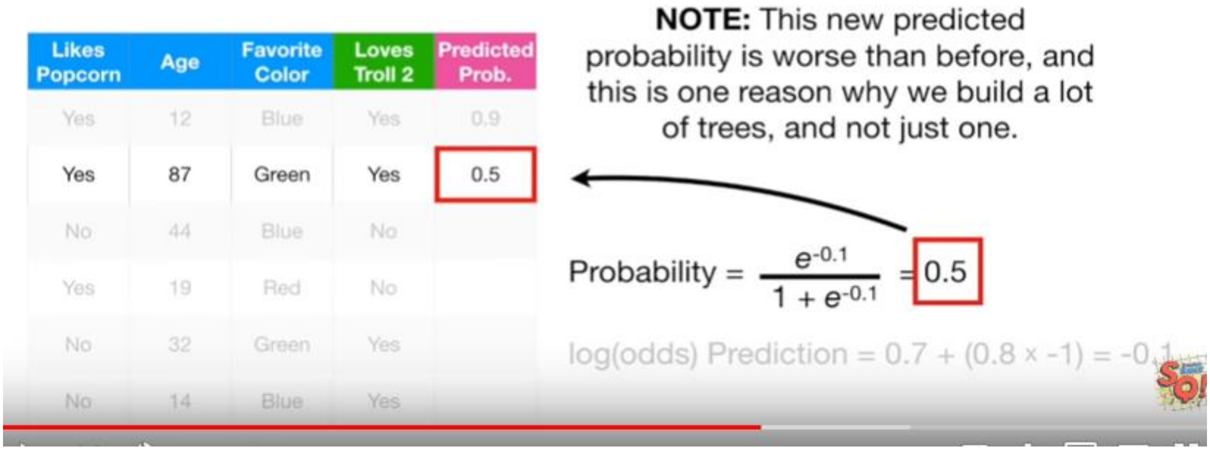



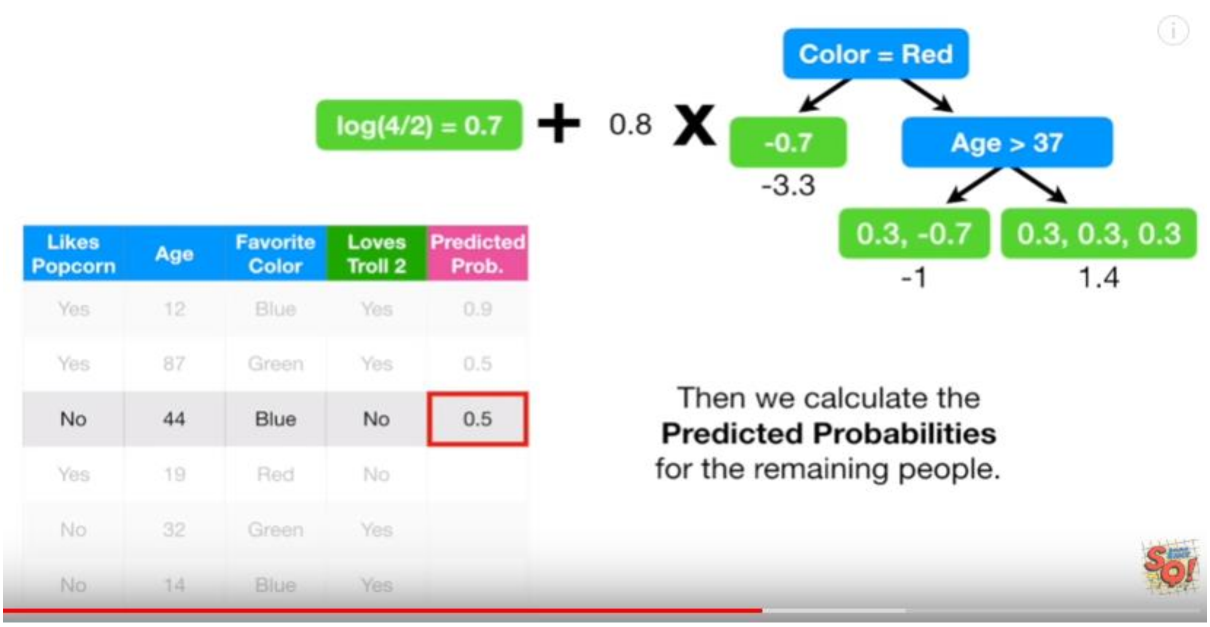
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now we convert the new **log(odds) Prediction** into a **Probability**…
$\text{Probability} = \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}$
log(odds) Prediction = 0.7 + (0.8 × -1) = -0.1

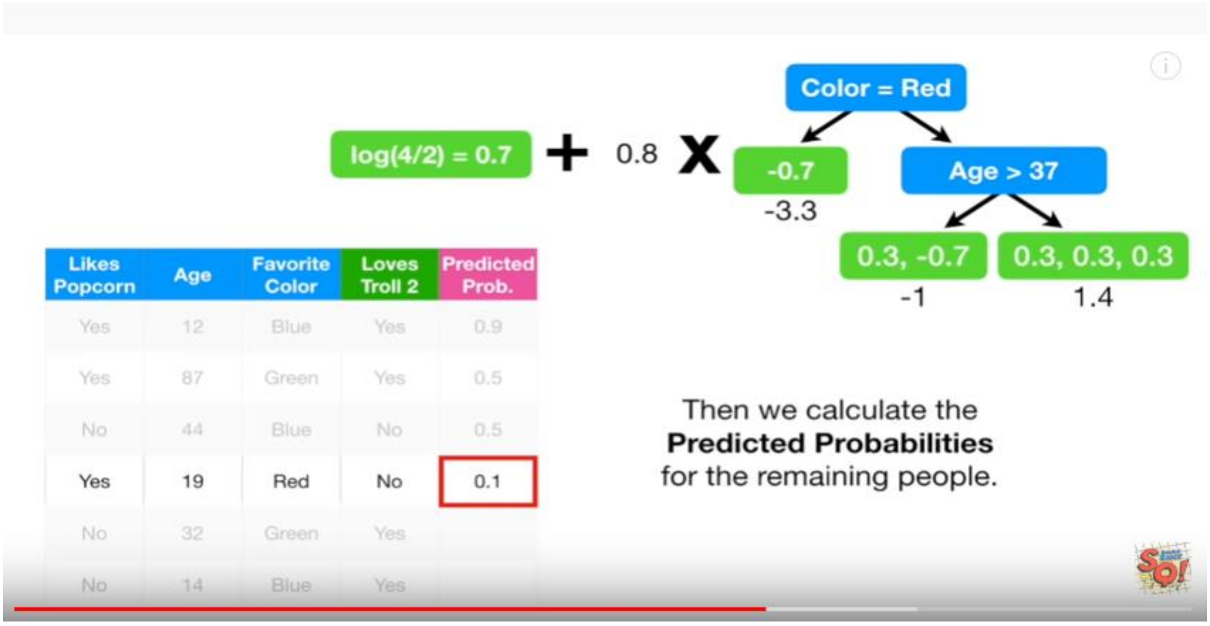

Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now we convert the new log(odds) Prediction into a Probability…
Probability = e^-0.1 / (1 + e^-0.1)
log(odds) Prediction = 0.7 + (0.8 x -1) = -0.1

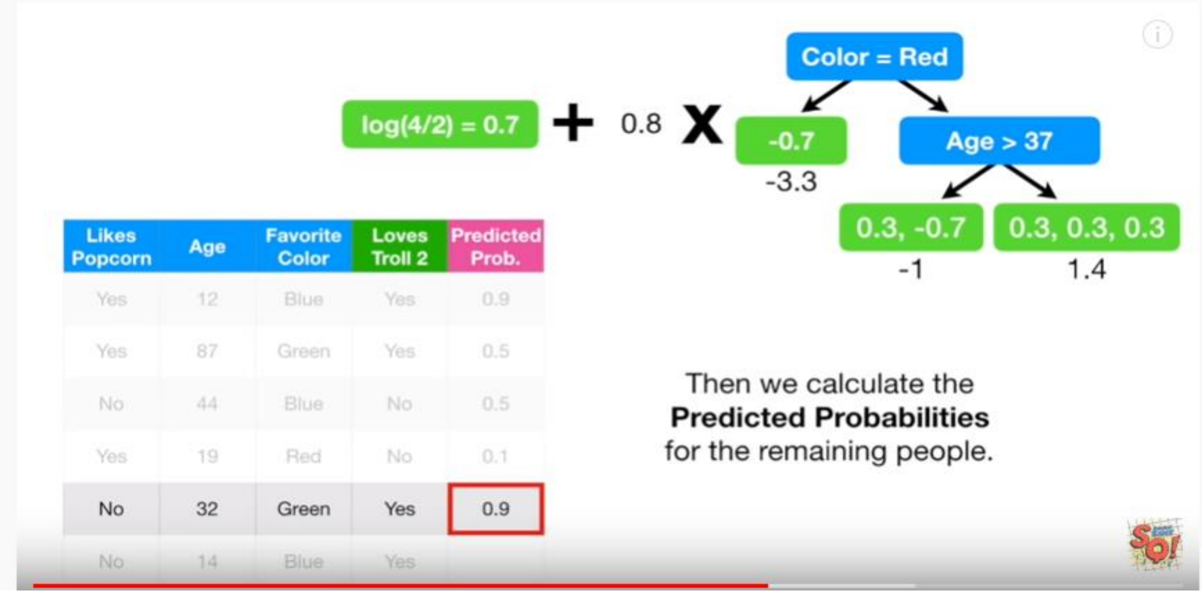

Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
Now we convert the new log(odds) Prediction into a Probability…
Probability = e^-0.1 / (1 + e^-0.1) = 0.5
log(odds) Prediction = 0.7 + (0.8 x -1) = -0.1

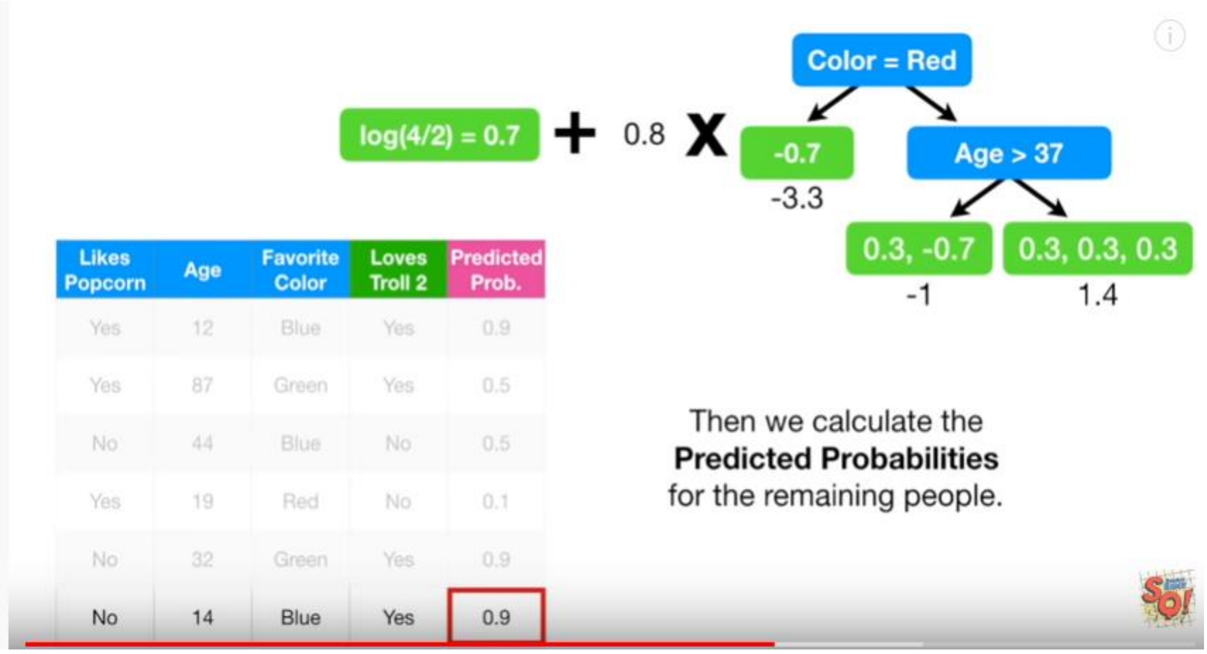

Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob.
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes | 0.5
No | 44 | Blue | No
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
…and save the new Predicted Probability, 0.5, here…
Probability = e^-0.1 / (1 + e^-0.1) = 0.5
log(odds) Prediction = 0.7 + (0.8 x -1) = -0.1


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
Yes | 19 | Red | No
No | 32 | Green | Yes
No | 14 | Blue | Yes
NOTE: This new predicted probability is worse than before, and this is one reason why we build a lot of trees, and not just one.
Probability = $\frac{e^{-0.1}}{1 + e^{-0.1}}$ = 0.5
log(odds) Prediction = 0.7 + (0.8 × -1) = -0.1


log(4/2) = 0.7
+ 0.8 X
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
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
No | 32 | Green | Yes
No | 14 | Blue | Yes
Then we calculate the Predicted Probabilities for the remaining people.


log(4/2) = 0.7
+ 0.8 X
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
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
No | 14 | Blue | Yes
Then we calculate the Predicted Probabilities for the remaining people.




| Initial_prediction 1[Log(odds)=log(4/2=0.7)] | Predicted_Probability(1st) | Residuals(Actual_Prob - Predicted Prob) | Updated_prediction 1 [New log(odds) prediction] | New Predicted_Probability |
|---|---|---|---|---|
| 0.7 | 0.7 | 1-0.7 = 0.3 | 1.8 | 0.9 |
| 0.7 | 0.7 | 1-0.7 = 0.3 | -0.1 | 0.5 |
| 0.7 | 0.7 | 0-0.7 = -0.7 | -0.1 | 0.5 |
| 0.7 | 0.7 | 0-0.7 = -0.7 | -1.94 | 0.1 |
| 0.7 | 0.7 | 1-0.7 = 0.3 | 1.8 | 0.9 |
| 0.7 | 0.7 | 1-0.7 = 0.3 | 1.8 | 0.9 |

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

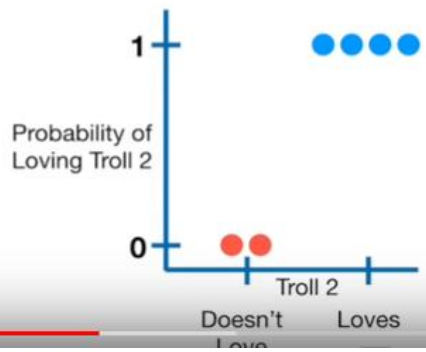
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | |
| Yes | 87 | Green | Yes | 0.5 | |
| No | 44 | Blue | No | 0.5 | |
| Yes | 19 | Red | No | 0.1 | |
| No | 32 | Green | Yes | 0.9 | |
| No | 14 | Blue | Yes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

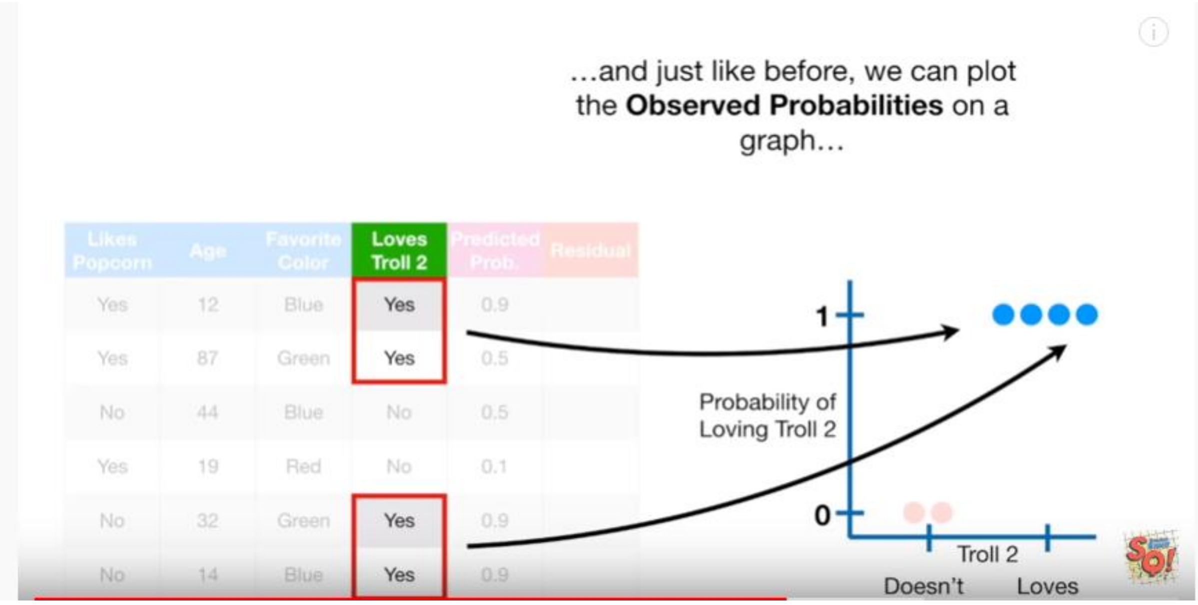
...and just like before, **Residuals** are the difference between the **Observed** and **Predicted Probabilities**... → **Residual** = (**Observed** - **Predicted**)

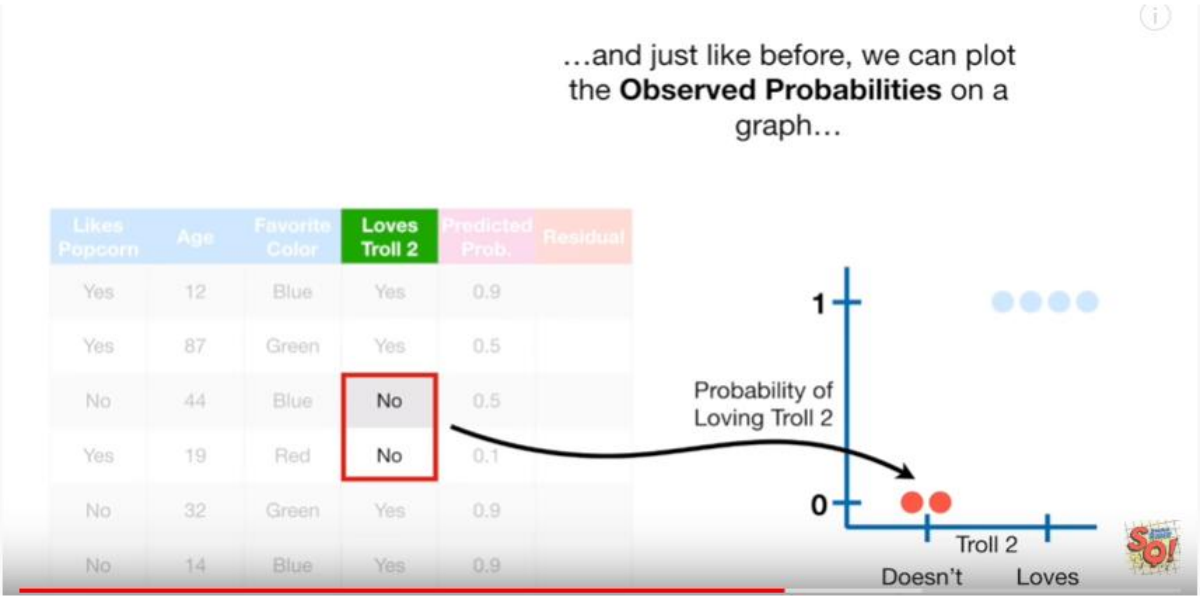
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | |
| Yes | 87 | Green | Yes | 0.5 | |
| No | 44 | Blue | No | 0.5 | |
| Yes | 19 | Red | No | 0.1 | |
| No | 32 | Green | Yes | 0.9 | |
| No | 14 | Blue | Yes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

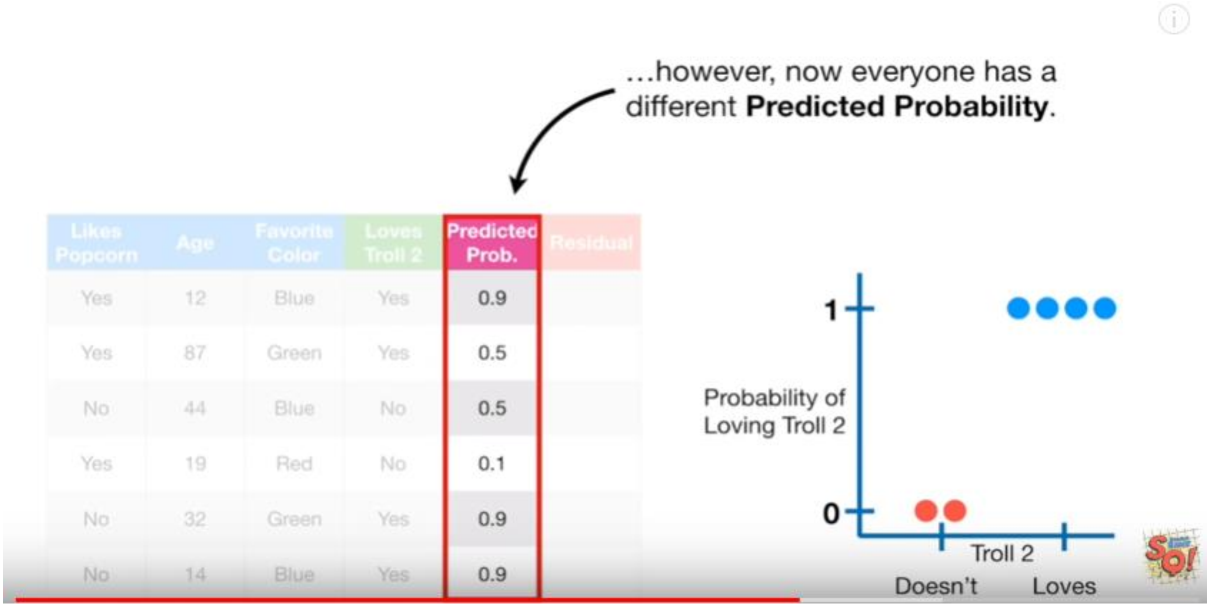
...and just like before, we can plot the **Observed Probabilities** on a graph...

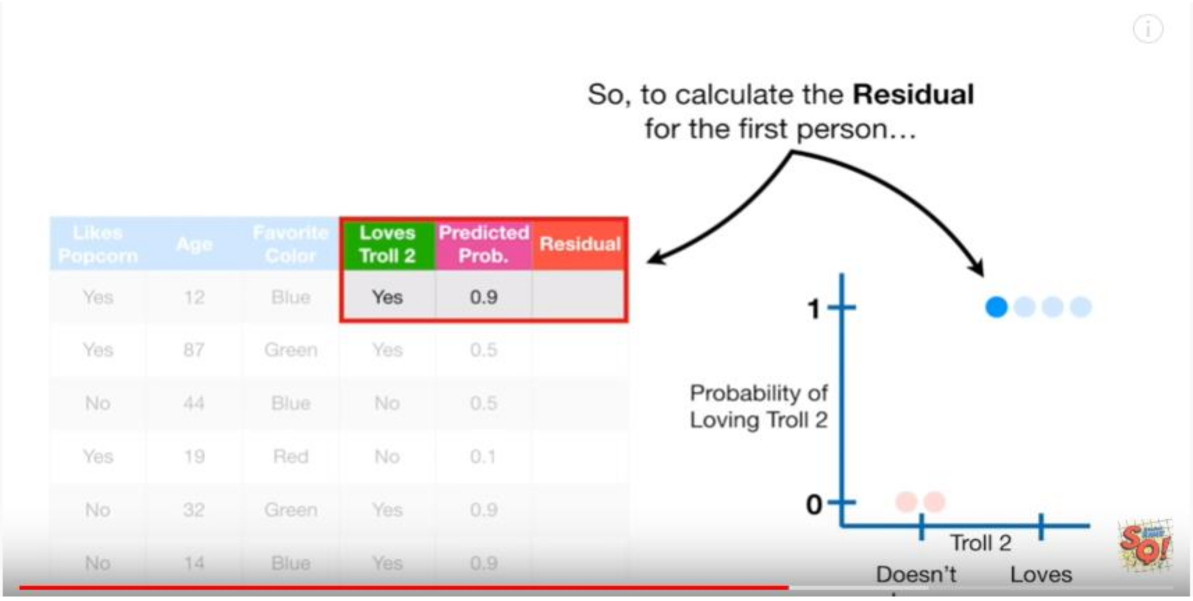
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | |
| Yes | 87 | Green | Yes | 0.5 | |
| No | 44 | Blue | No | 0.5 | |
| Yes | 19 | Red | No | 0.1 | |
| No | 32 | Green | Yes | 0.9 | |
| No | 14 | Blue | Yes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

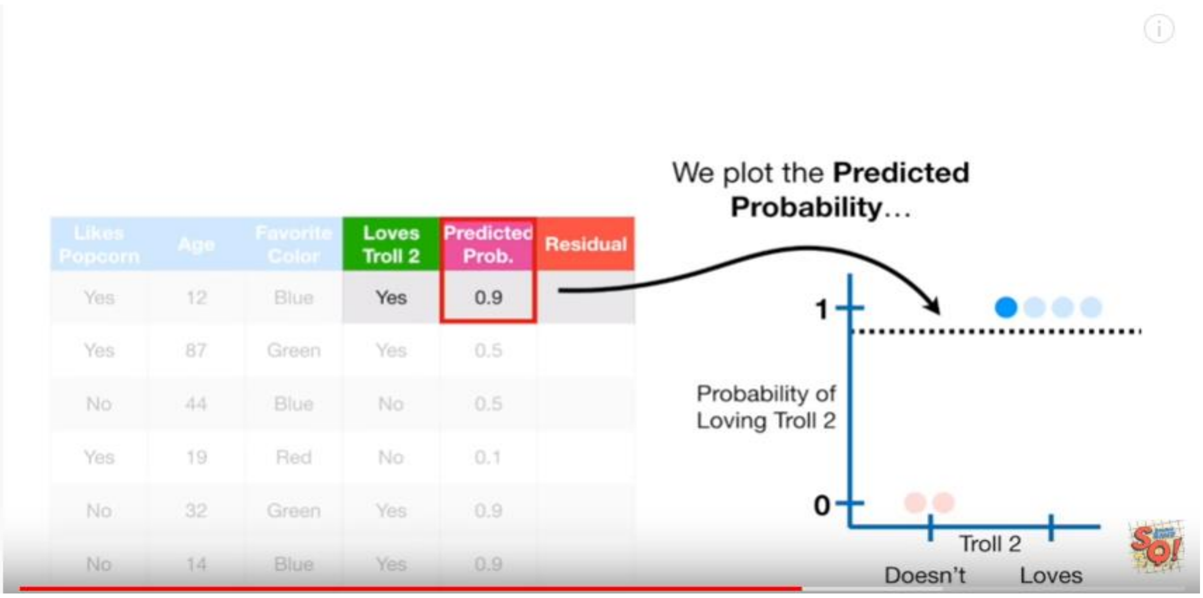


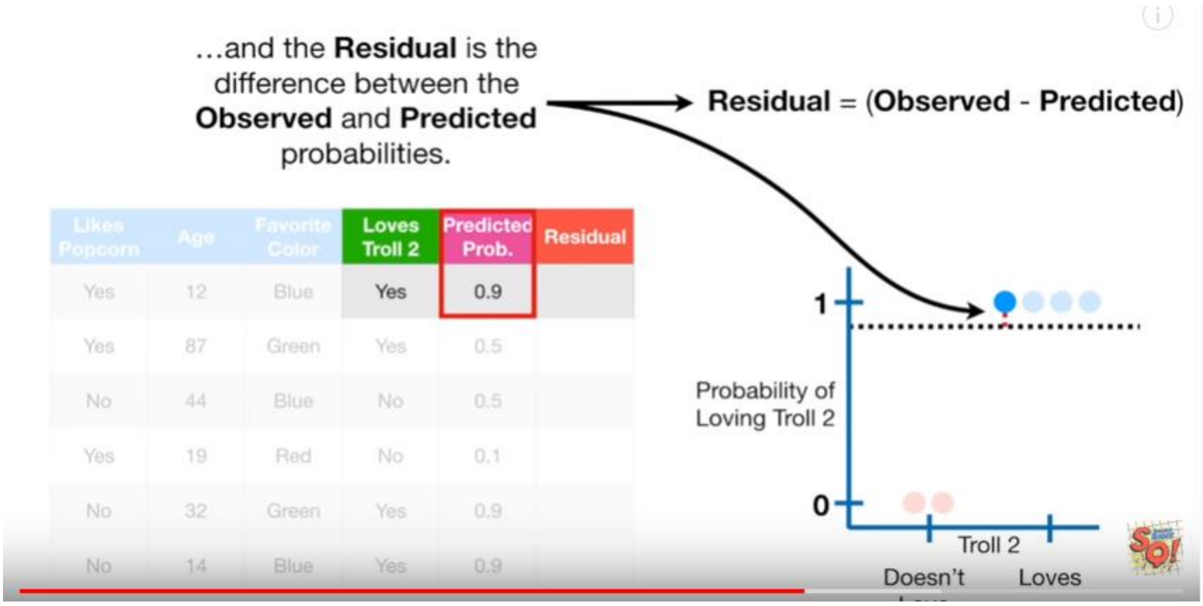

...and just like before, we can plot the Observed Probabilities on a graph...
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes | 0.5
No | 44 | Blue | No | 0.5
Yes | 19 | Red | No | 0.1
No | 32 | Green | Yes | 0.9
No | 14 | Blue | Yes | 0.9
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

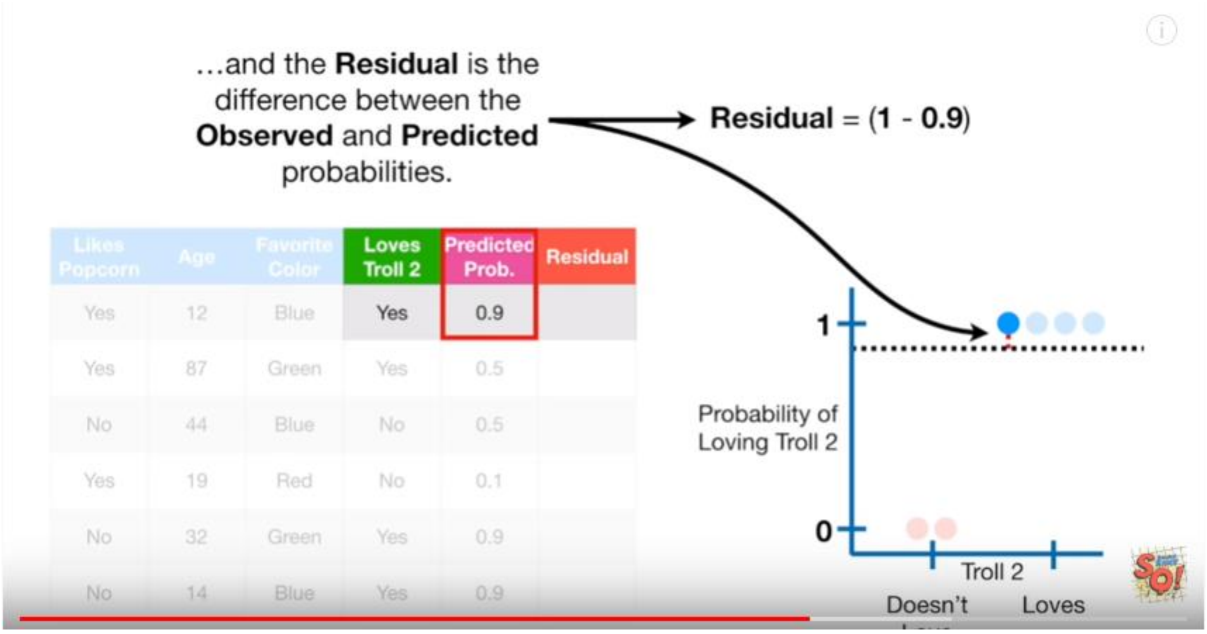

...and just like before, we can plot the Observed Probabilities on a graph...
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes | 0.5
No | 44 | Blue | No | 0.5
Yes | 19 | Red | No | 0.1
No | 32 | Green | Yes | 0.9
No | 14 | Blue | Yes | 0.9
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves


...however, now everyone has a different Predicted Probability.
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
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

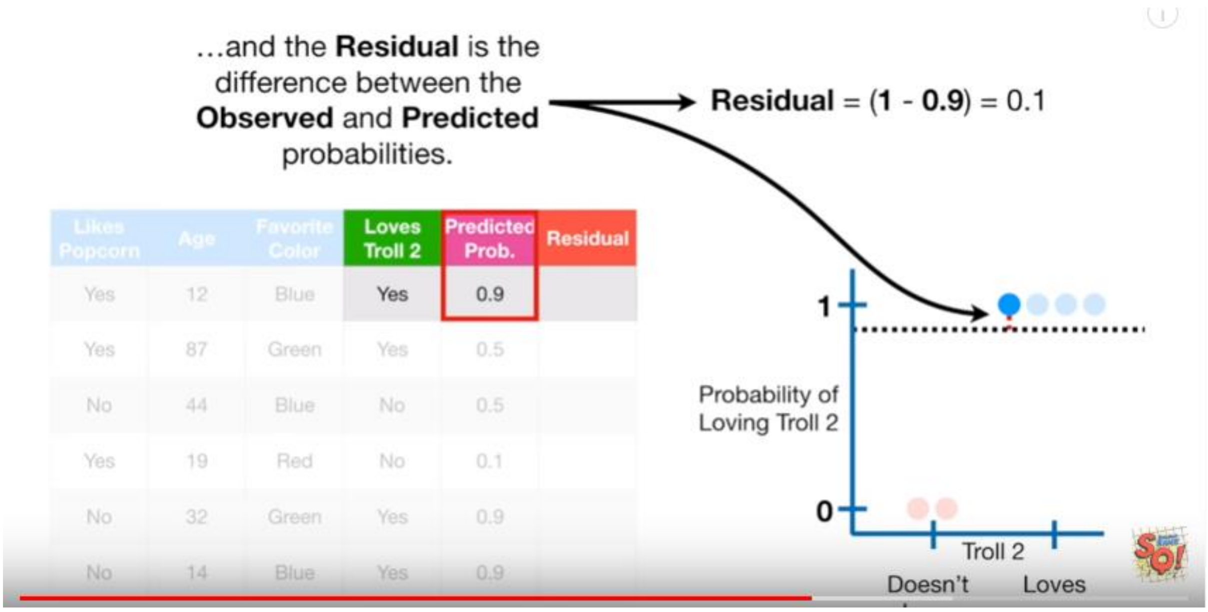

So, to calculate the **Residual** for the first person…
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes | 0.5
No | 44 | Blue | No | 0.5
Yes | 19 | Red | No | 0.1
No | 32 | Green | Yes | 0.9
No | 14 | Blue | Yes | 0.9
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves


We plot the **Predicted Probability**…
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
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

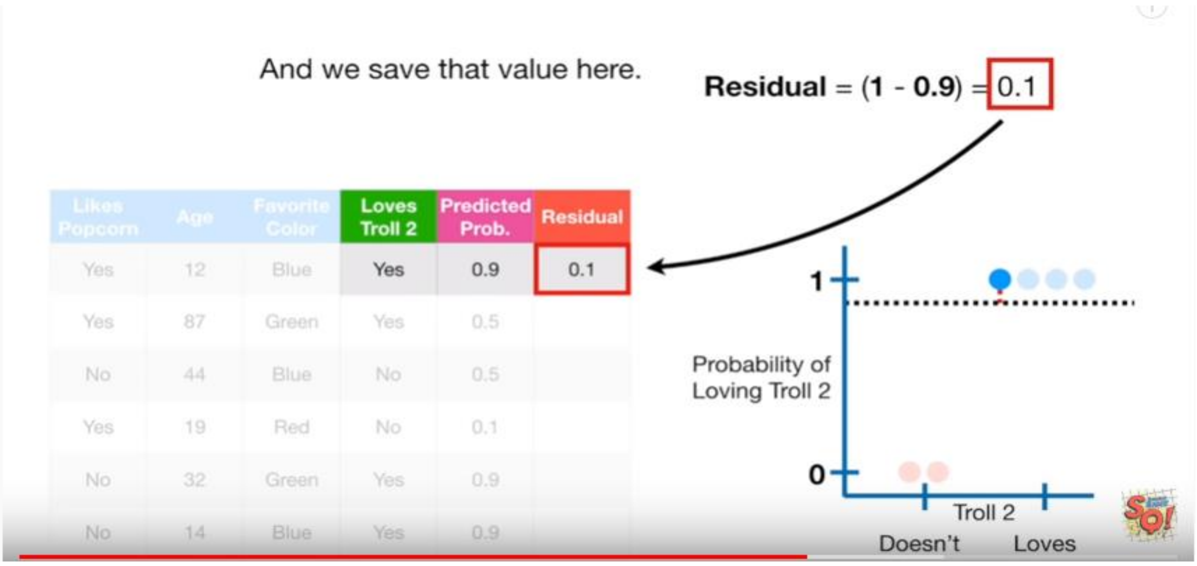

…and the **Residual** is the difference between the **Observed** and **Predicted** probabilities.
**Residual** = (**Observed** - **Predicted**)
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9
Yes | 87 | Green | Yes | 0.5
No | 44 | Blue | No | 0.5
Yes | 19 | Red | No | 0.1
No | 32 | Green | Yes | 0.9
No | 14 | Blue | Yes | 0.9
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves



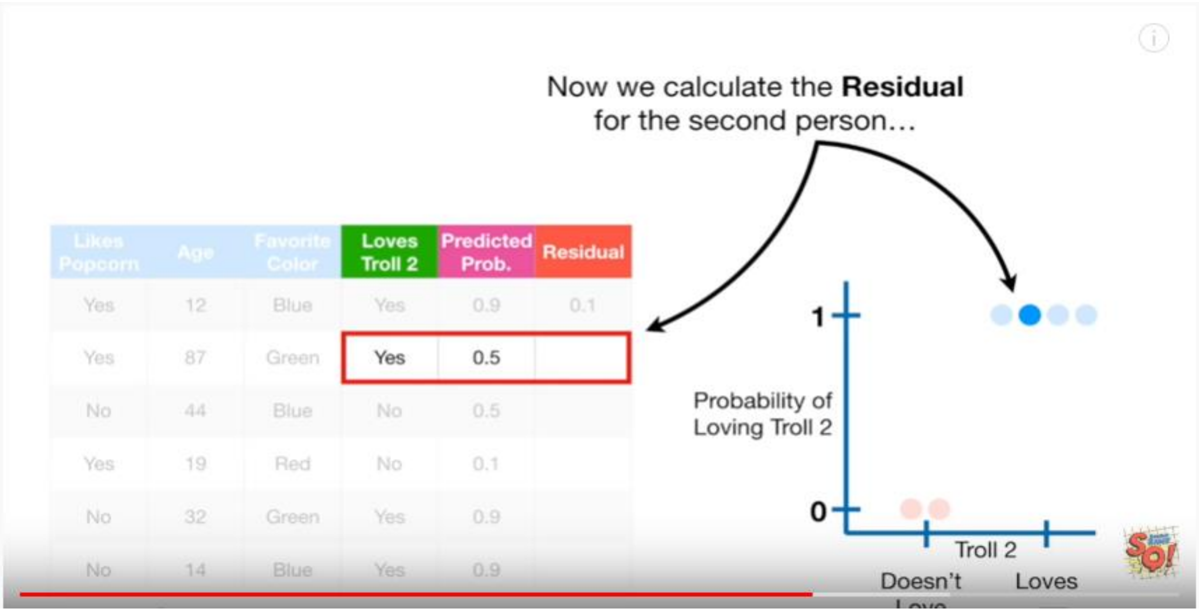
...and the **Residual** is the difference between the **Observed** and **Predicted** probabilities.
**Residual** = (**1** - **0.9**)
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 |
Yes | 87 | Green | Yes | 0.5 |
No | 44 | Blue | No | 0.5 |
Yes | 19 | Red | No | 0.1 |
No | 32 | Green | Yes | 0.9 |
No | 14 | Blue | Yes | 0.9 |
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

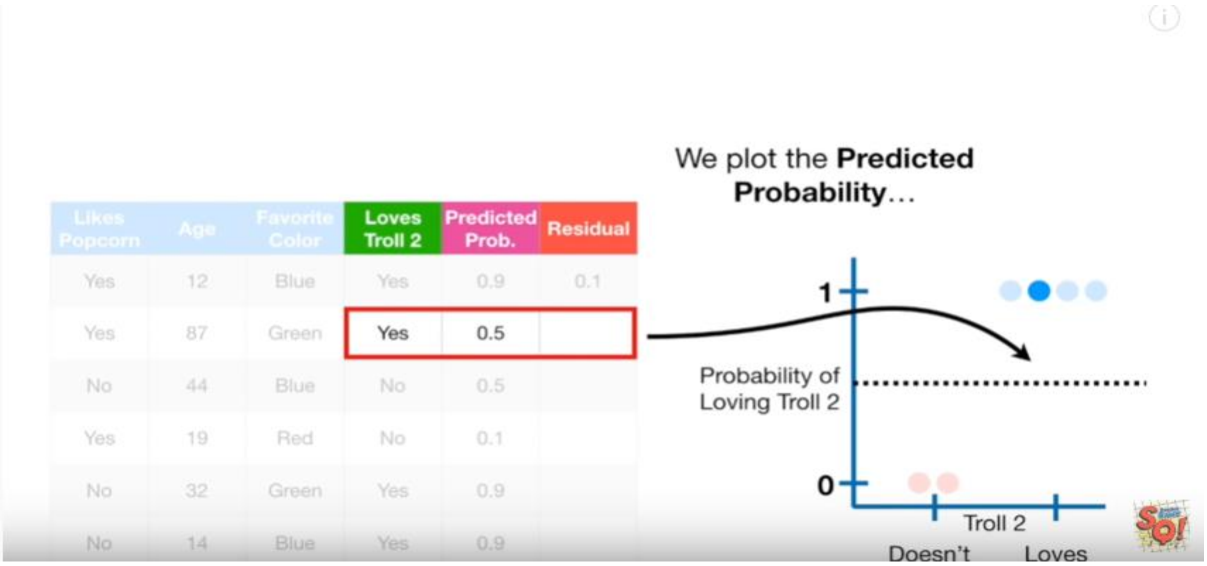



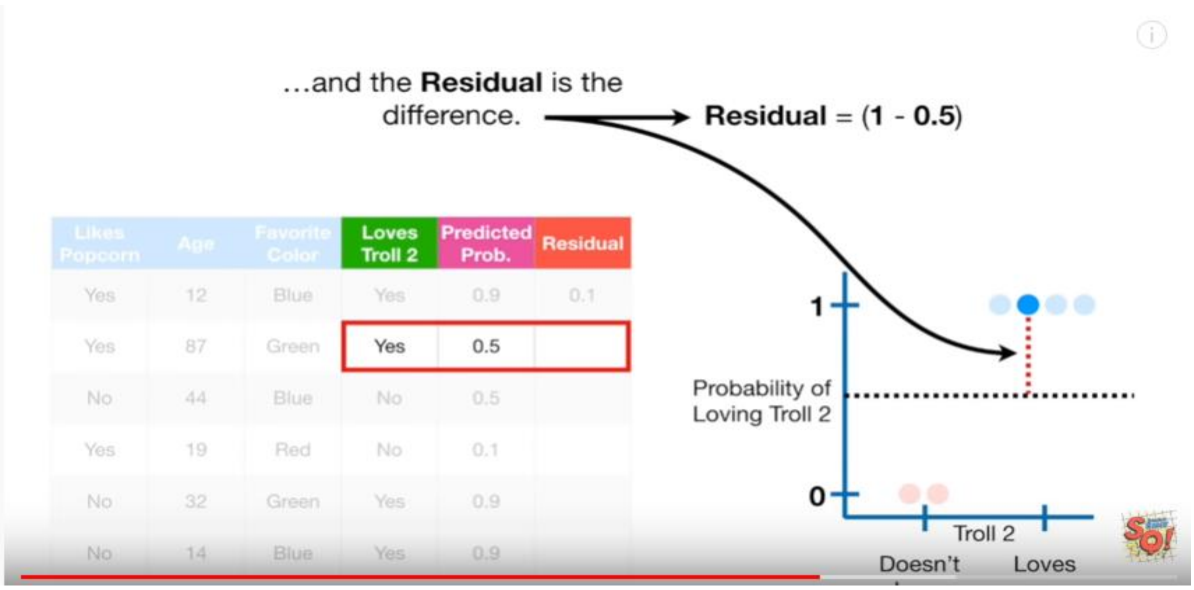
...and the **Residual** is the difference between the **Observed** and **Predicted** probabilities.
**Residual** = (**1** - **0.9**) = 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 |
Yes | 87 | Green | Yes | 0.5 |
No | 44 | Blue | No | 0.5 |
Yes | 19 | Red | No | 0.1 |
No | 32 | Green | Yes | 0.9 |
No | 14 | Blue | Yes | 0.9 |
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

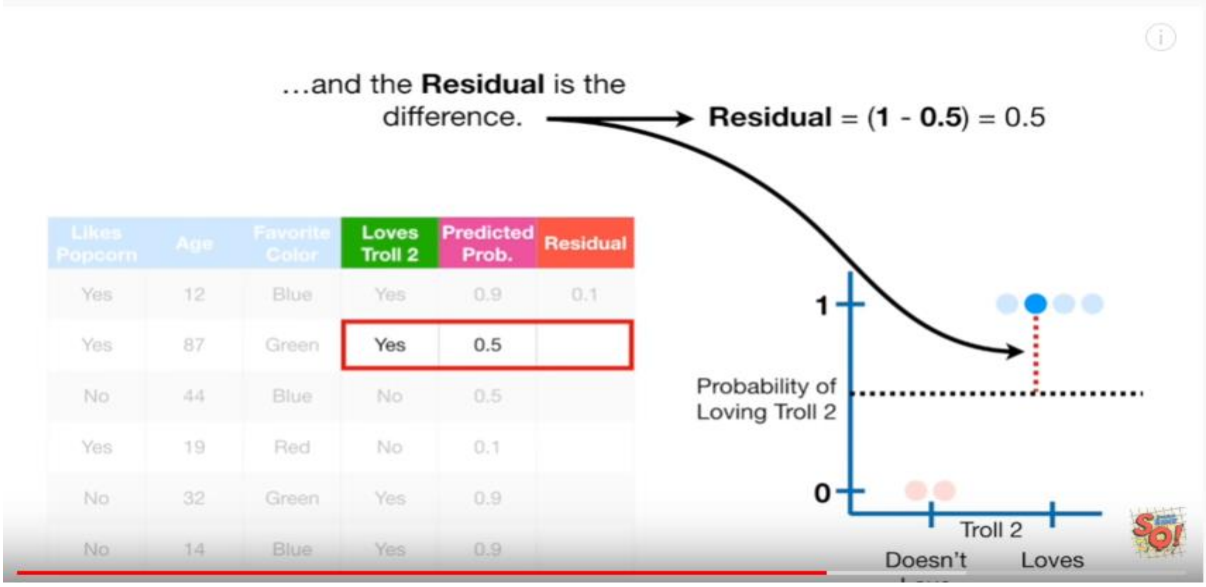



And we save that value here.
**Residual** = (**1** - **0.9**) = 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 |
No | 44 | Blue | No | 0.5 |
Yes | 19 | Red | No | 0.1 |
No | 32 | Green | Yes | 0.9 |
No | 14 | Blue | Yes | 0.9 |
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

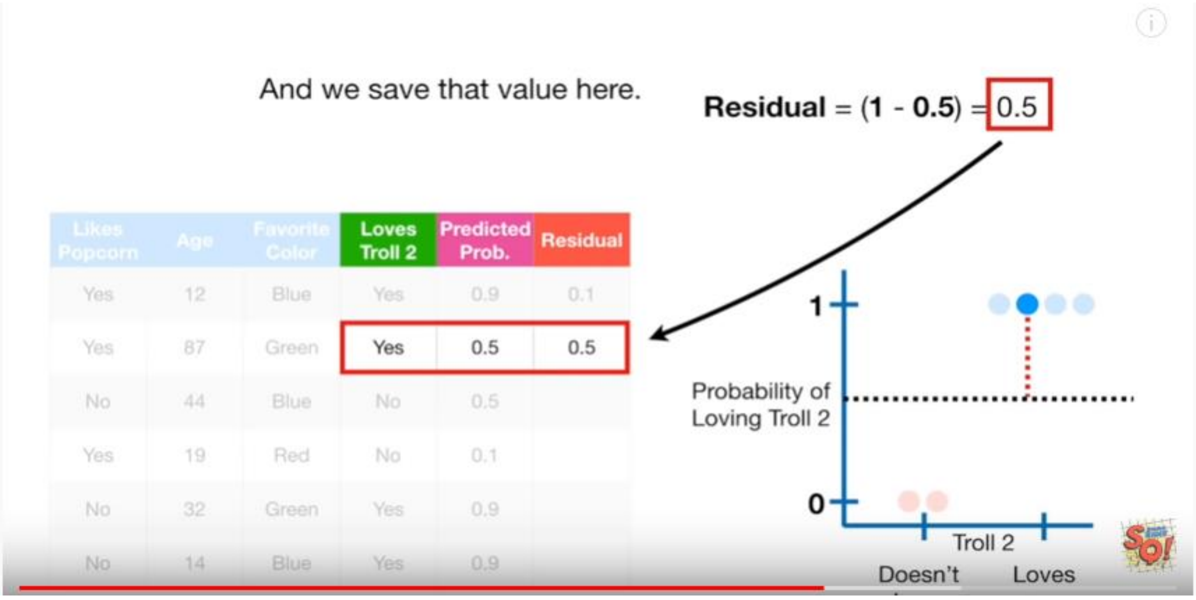

Now we calculate the **Residual** for the second person…
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 |
No | 44 | Blue | No | 0.5 |
Yes | 19 | Red | No | 0.1 |
No | 32 | Green | Yes | 0.9 |
No | 14 | Blue | Yes | 0.9 |
1
0
Probability of Loving Troll 2
Troll 2
Doesn't Love
Loves


We plot the **Predicted Probability**…
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 |
No | 44 | Blue | No | 0.5 |
Yes | 19 | Red | No | 0.1 |
No | 32 | Green | Yes | 0.9 |
No | 14 | Blue | Yes | 0.9 |
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves

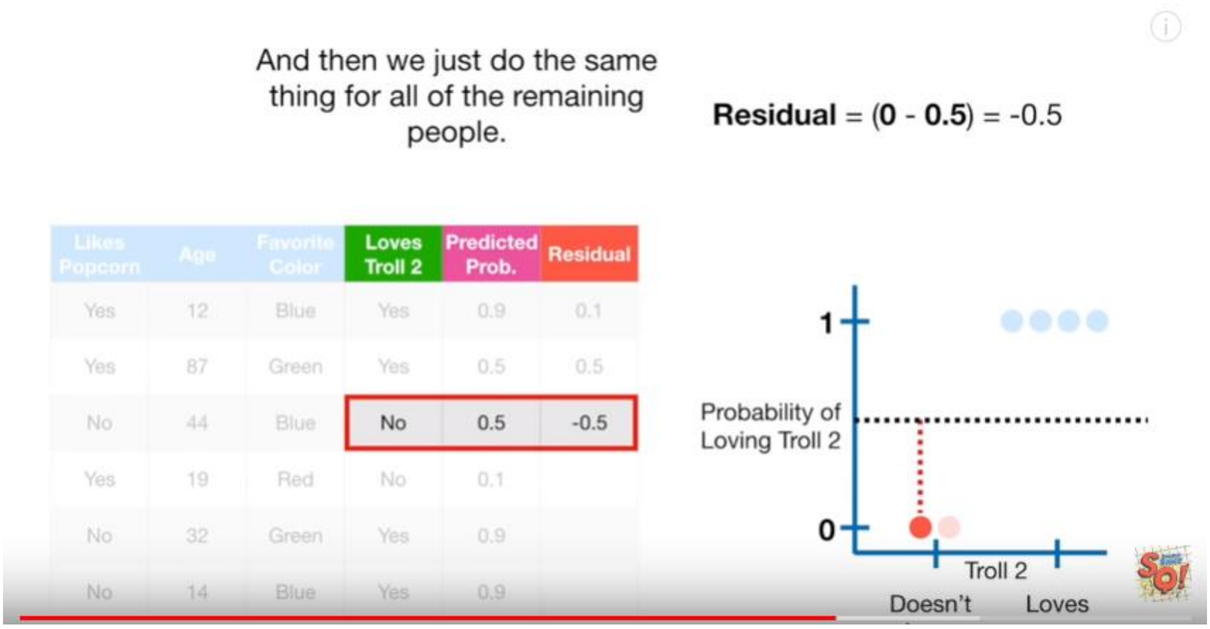

…and the **Residual** is the difference.
**Residual** = (**1** - **0.5**)
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 |
No | 44 | Blue | No | 0.5 |
Yes | 19 | Red | No | 0.1 |
No | 32 | Green | Yes | 0.9 |
No | 14 | Blue | Yes | 0.9 |
1
0
Probability of Loving Troll 2
Troll 2
Doesn't
Loves



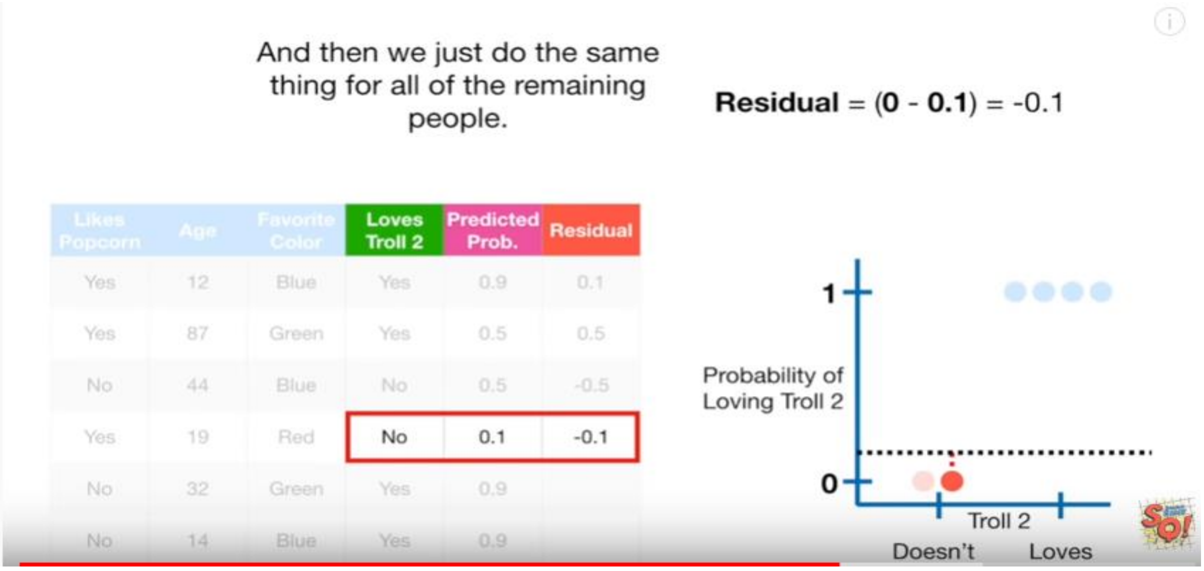

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | |
| No | 44 | Blue | No | 0.5 | |
| Yes | 19 | Red | No | 0.1 | |
| No | 32 | Green | Yes | 0.9 | |
| No | 14 | Blue |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



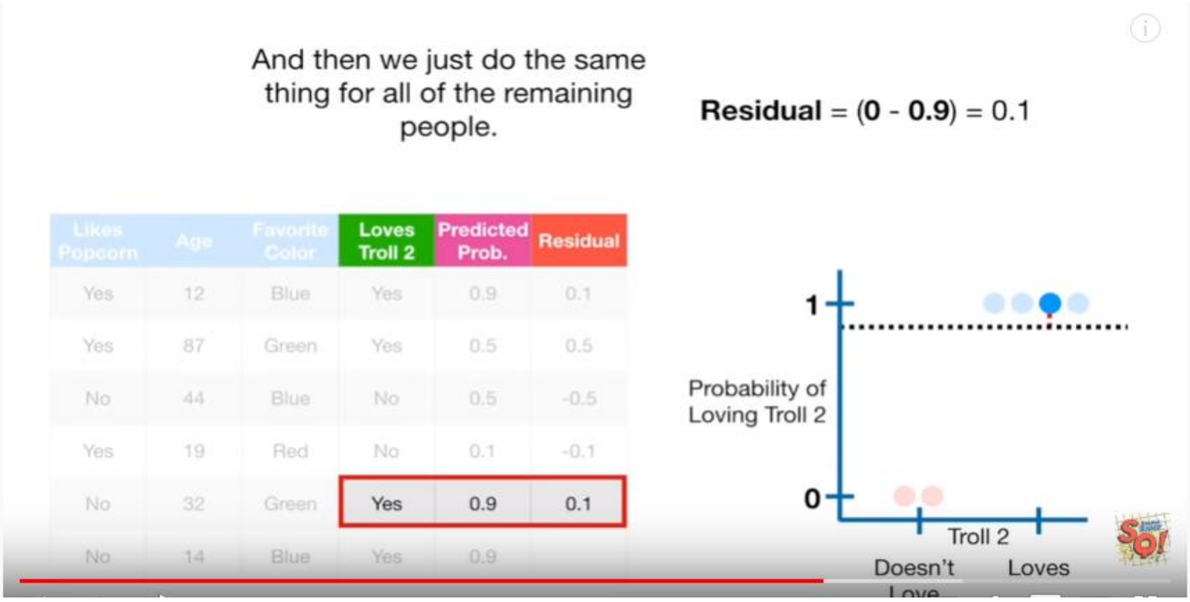

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | 0.5 |
| No | 44 | Blue | No | 0.5 | |
| Yes | 19 | Red | No | 0.1 | |
| No | 32 | Green | Yes | 0.9 | |
| No | 14 | Blue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



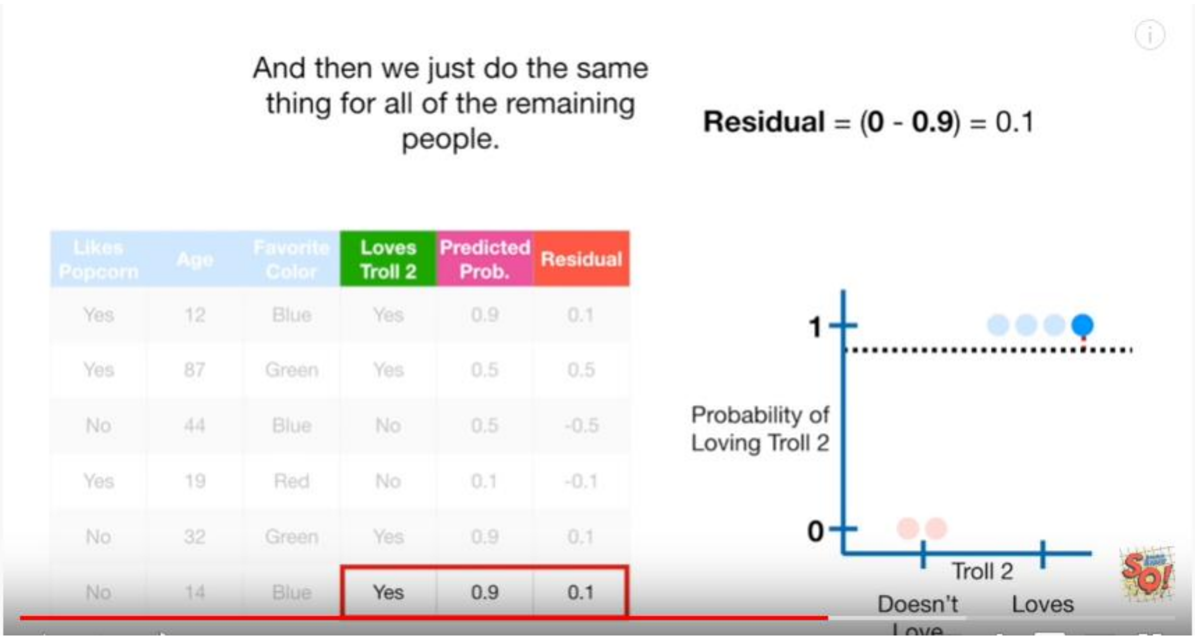

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | 0.5 |
| No | 44 | Blue | No | 0.5 | -0.5 |
| Yes | 19 | Red | No | 0.1 | |
| No | 32 | Green | Yes | 0.9 | |
| No | 14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

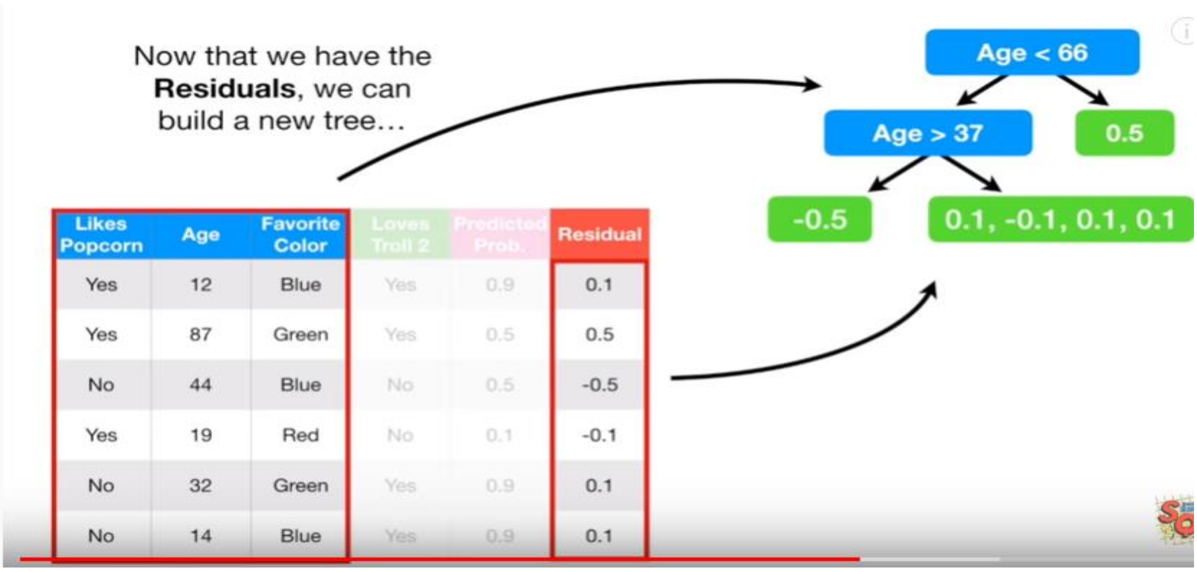


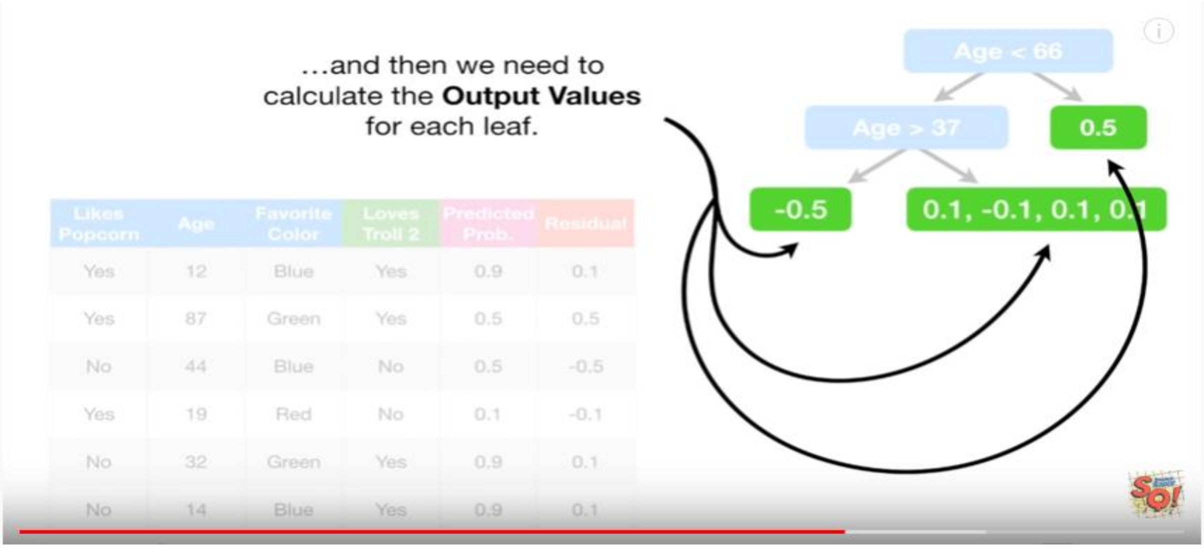

| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | 0.5 |
| No | 44 | Blue | No | 0.5 | -0.5 |
| Yes | 19 | Red | No | 0.1 | -0.1 |
| No | 32 | Green | Yes | 0.9 | 0.1 |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




| 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 |

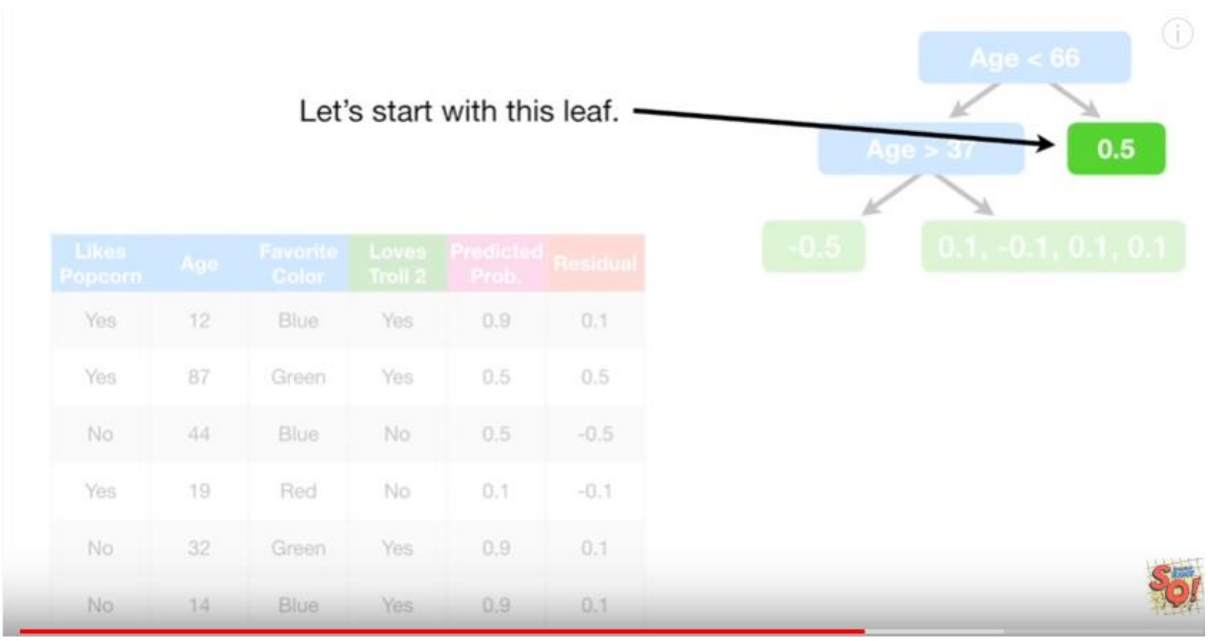

| Initial_prediction 1[Log(odds)=log (4/2=0.7)] | Predicted_Probability(1st) | Residuals(Actual_Prob - Predicted Prob) | Updated_prediction 1 [New log(odds) prediction] | New Predicted_Probability | New_Residuals |
|---|---|---|---|---|---|
| 0.7 | 0.7 | 1-0.7 = 0.3 | 1.8 | 0.9 | 1-0.9 = 0.1 |
| 0.7 | 0.7 | 1-0.7 = 0.3 | -0.1 | 0.5 | 1-0.5 = 0.5 |
| 0.7 | 0.7 | 0-0.7 = -0.7 | -0.1 | 0.5 | 0-0.5 = -0.5 |
| 0.7 | 0.7 | 0-0.7 = -0.7 | -1.94 | 0.1 | 0-0.1 = -0.1 |
| 0.7 | 0.7 | 1-0.7 = 0.3 | 1.8 | 0.9 | 1-0.9 = 0.1 |
| 0.7 | 0.7 | 1-0.7 = 0.3 | 1.8 | 0.9 | 1-0.9 = 0.1 |





…and then we need to calculate the **Output Values** for each leaf.
Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1

| 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 |

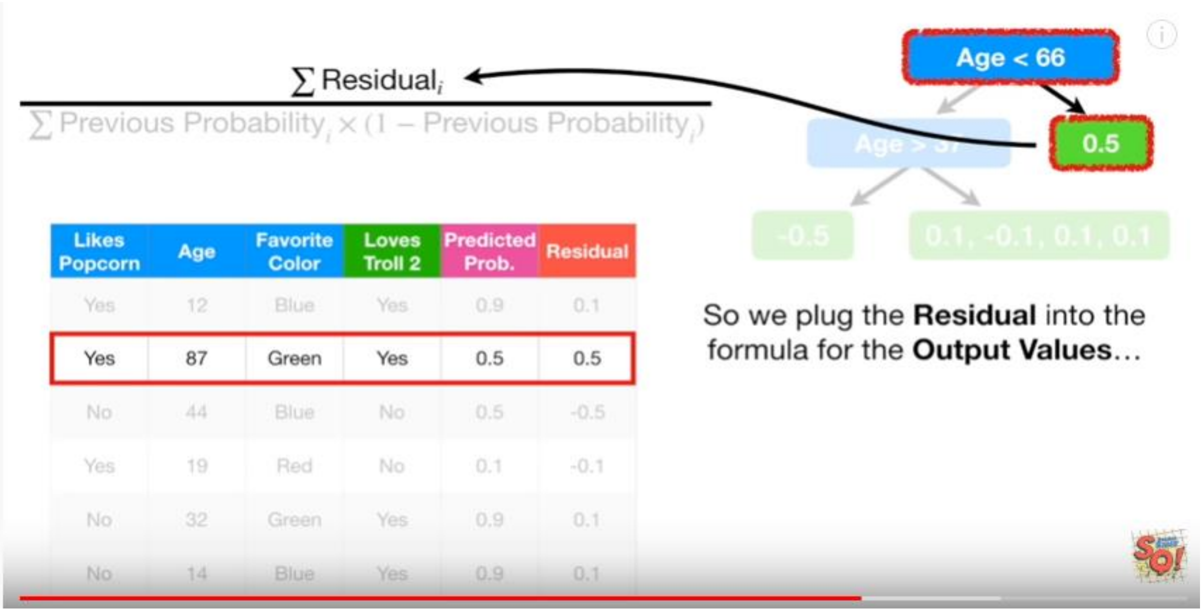



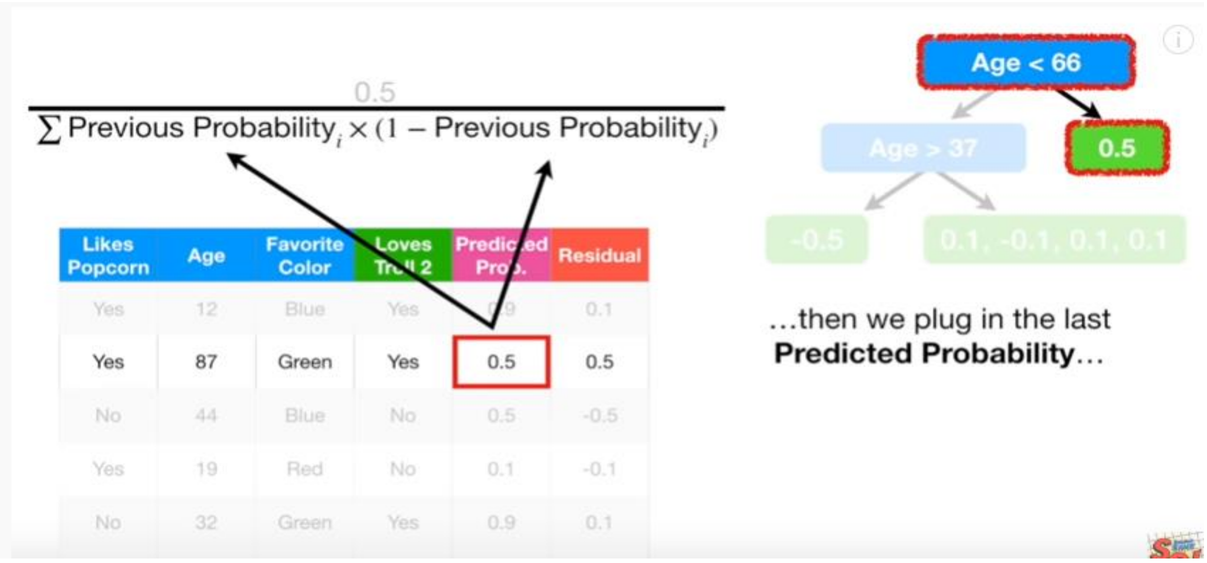
Let's start with this leaf.
Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1

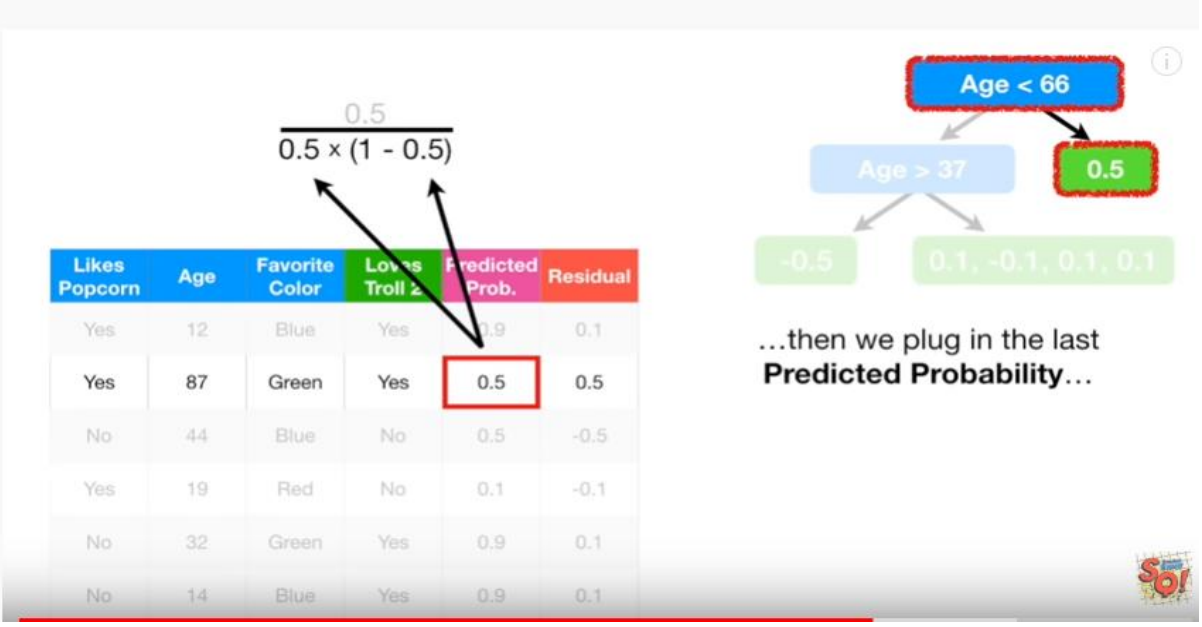
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | 0.5 |
| No | 44 | Blue | No | 0.5 | -0.5 |
| Yes | 19 | Red | No | 0.1 | -0.1 |
| No | 32 | Green | Yes | 0.9 | 0.1 |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

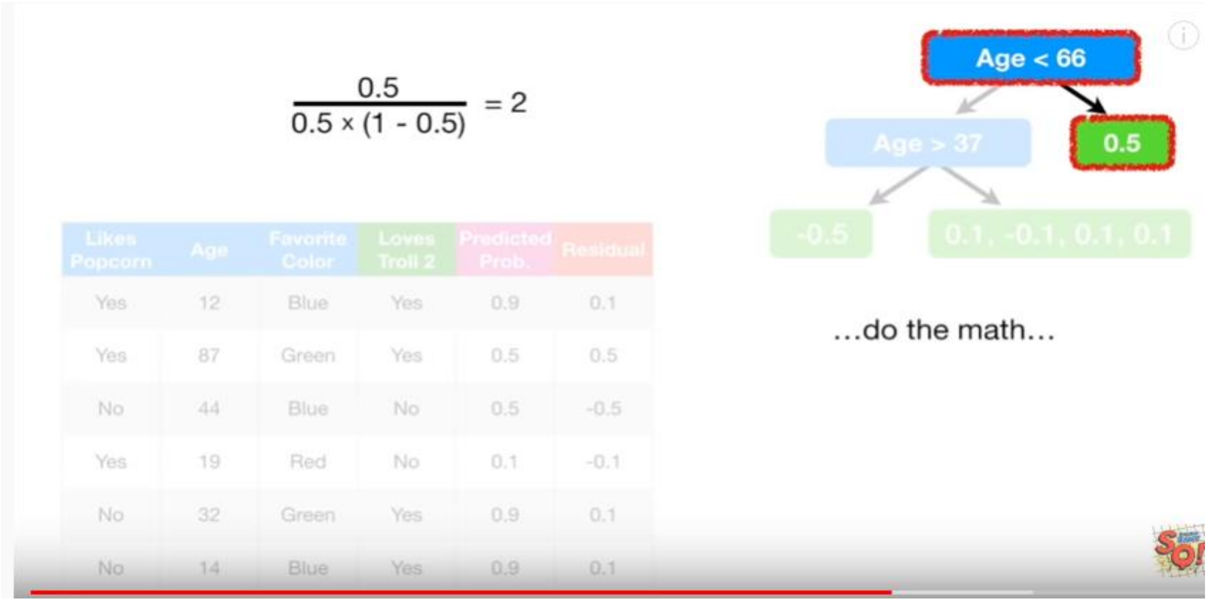

Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1
**NOTE:** Only the second person…
…goes to this leaf.

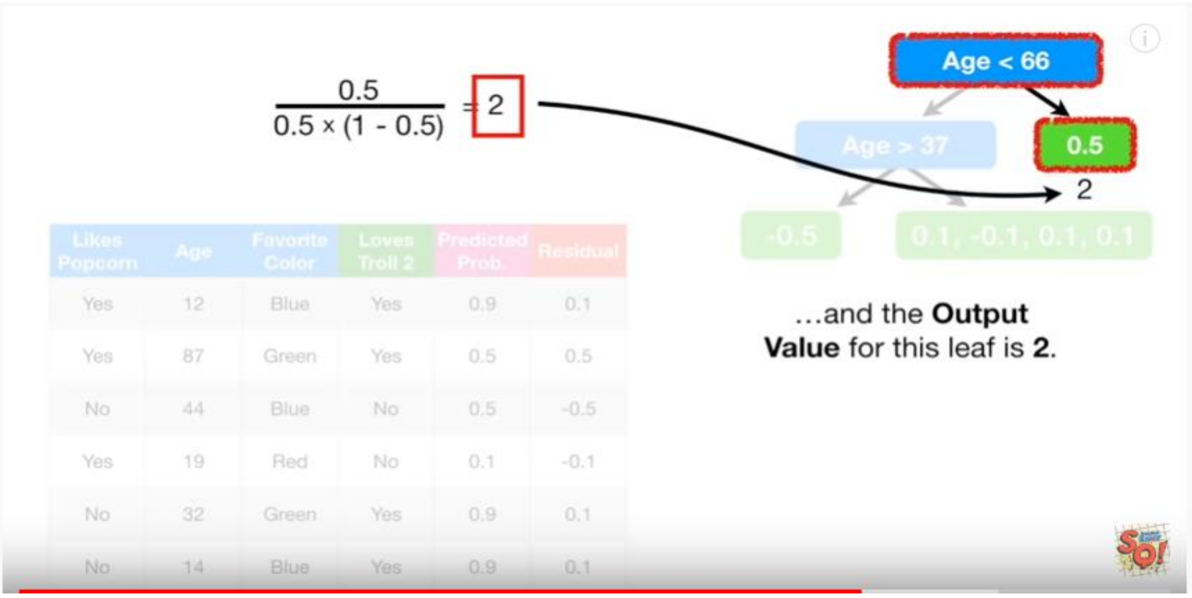
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | 0.5 |
| No | 44 | Blue | No | 0.5 | -0.5 |
| Yes | 19 | Red | No | 0.1 | -0.1 |
| No | 32 | Green | Yes | 0.9 | 0.1 |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

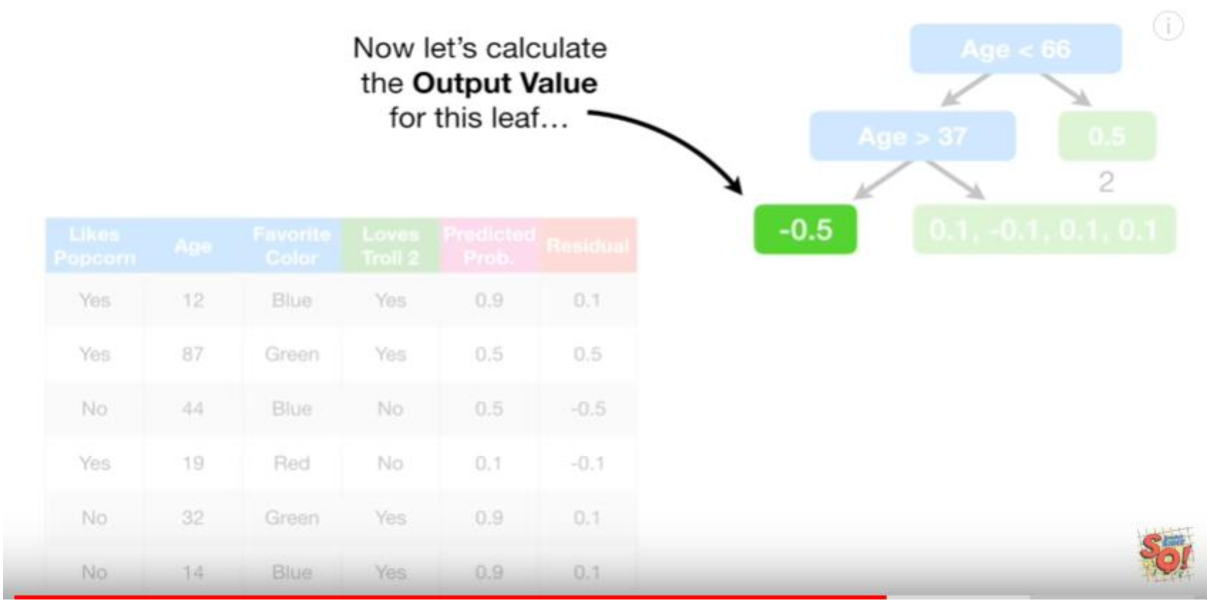

$\dfrac{\sum \text{Residual}_i}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$
Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 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 **Residual** into the formula for the **Output Values**…

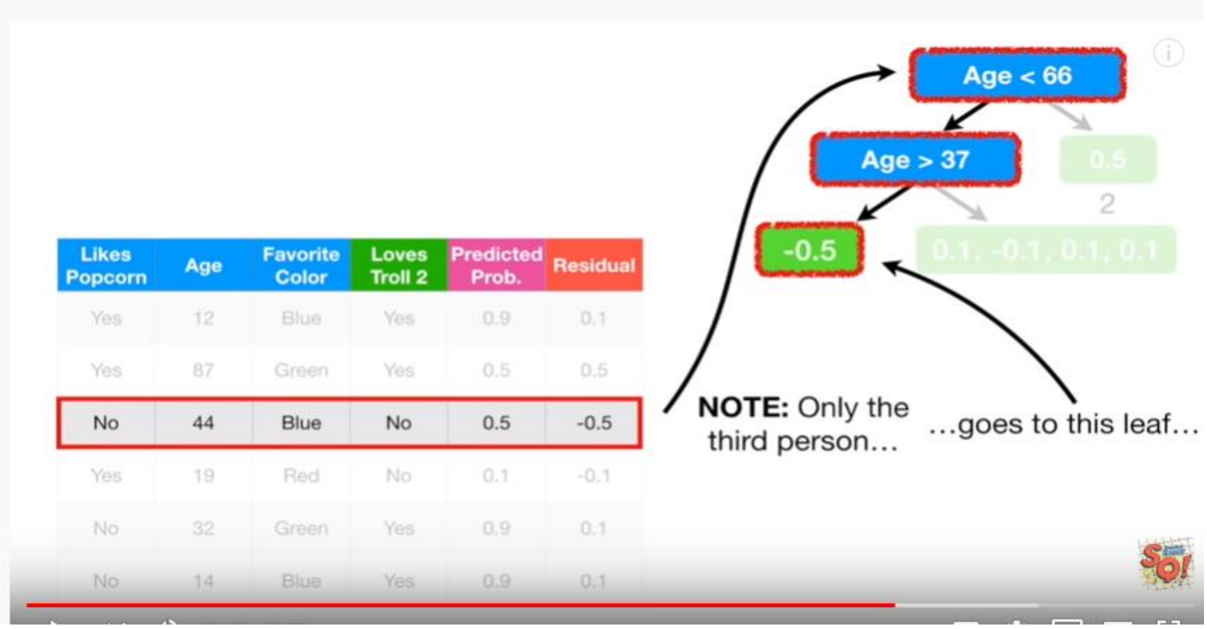

$\dfrac{0.5}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$
Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
…then we plug in the last **Predicted Probability**…

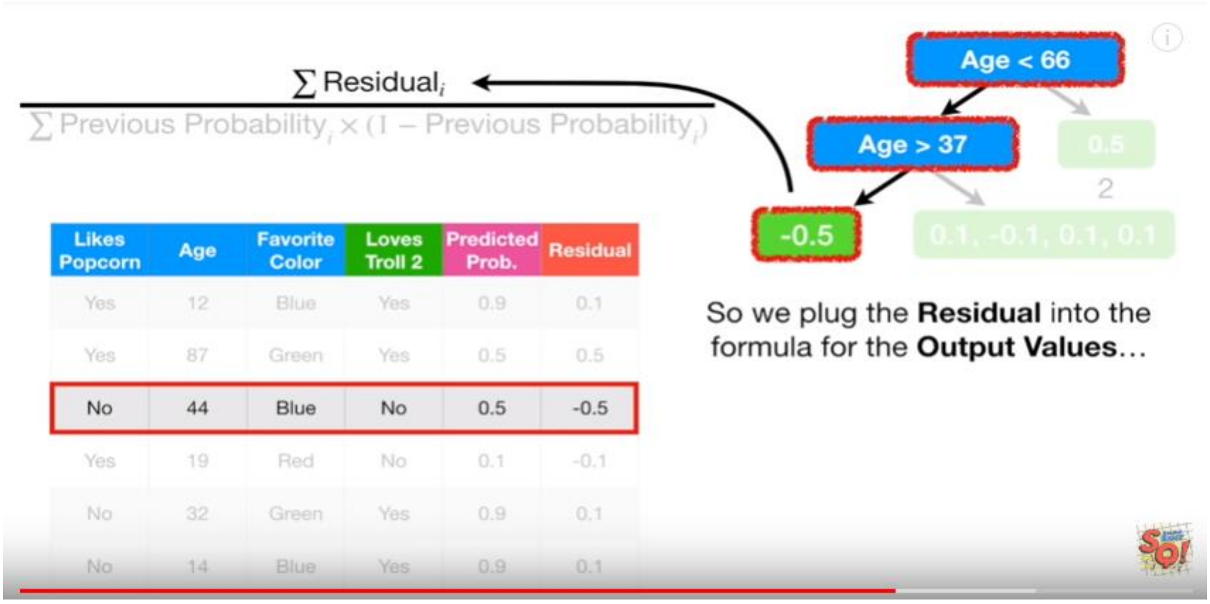

$\dfrac{0.5}{0.5 \times (1 - 0.5)}$
Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
No | 14 | Blue | Yes | 0.9 | 0.1
…then we plug in the last **Predicted Probability**…

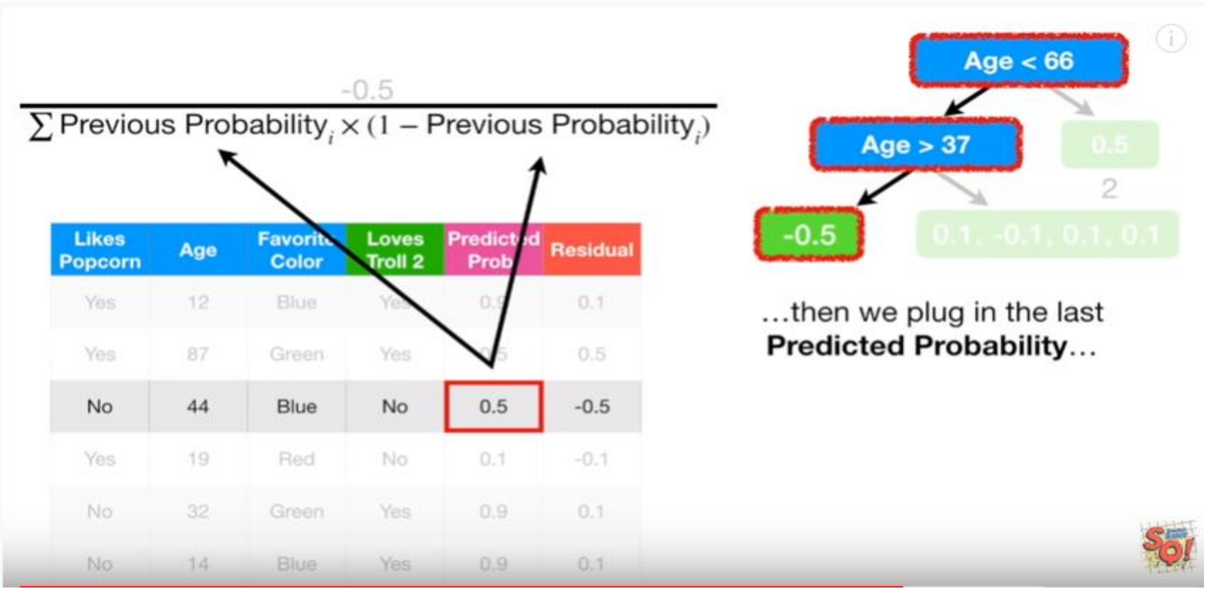

0.5 / (0.5 × (1 - 0.5)) = 2
Age < 66
Age > 37
0.5
-0.5
0.1, -0.1, 0.1, 0.1
…do the math…
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
No | 14 | Blue | Yes | 0.9 | 0.1


0.5 / (0.5 × (1 - 0.5)) = 2
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
…and the Output Value for this leaf is 2.
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

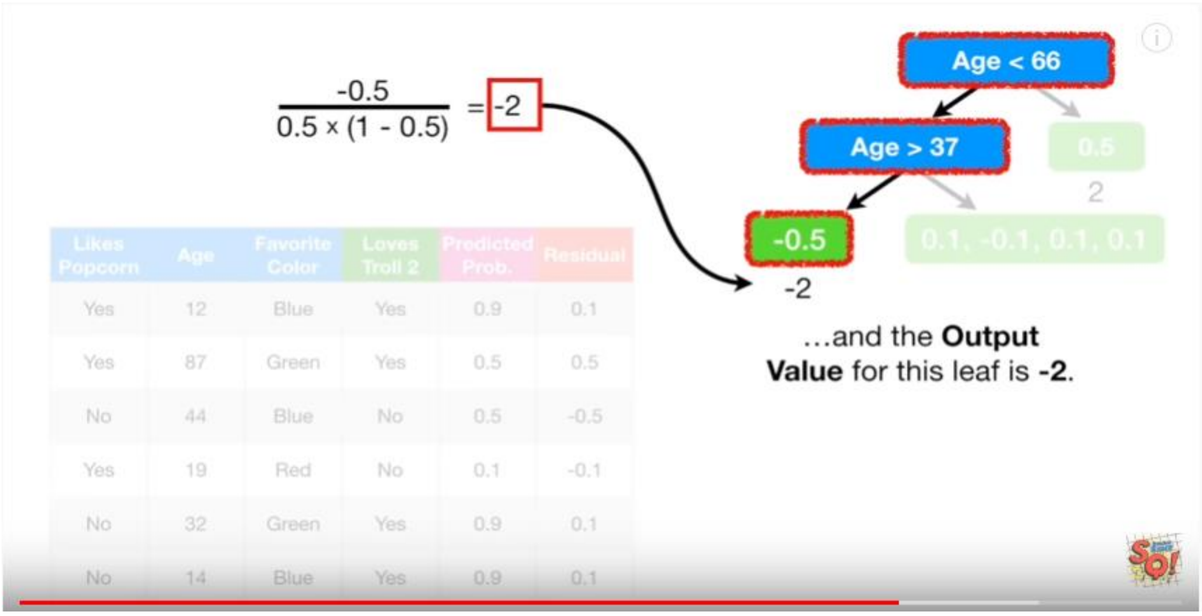

Now let's calculate the Output Value for this leaf…
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
No | 14 | Blue | Yes | 0.9 | 0.1

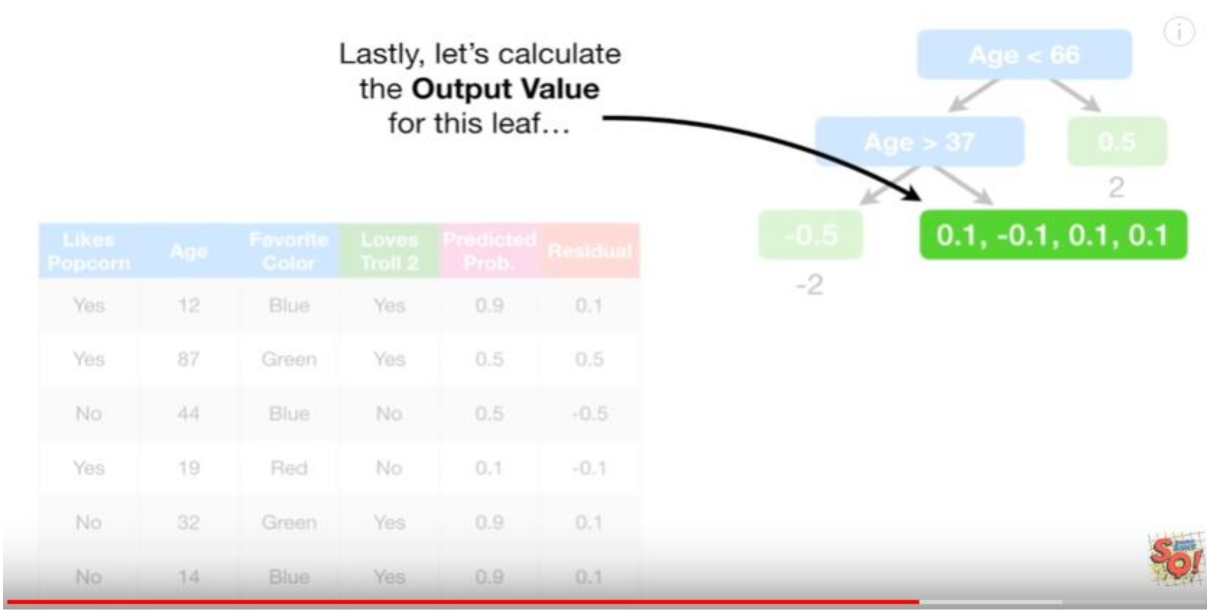

Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
No | 14 | Blue | Yes | 0.9 | 0.1
NOTE: Only the third person…
…goes to this leaf…


$\frac{\sum \text{Residual}_i}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
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 **Residual** into the formula for the **Output Values**…

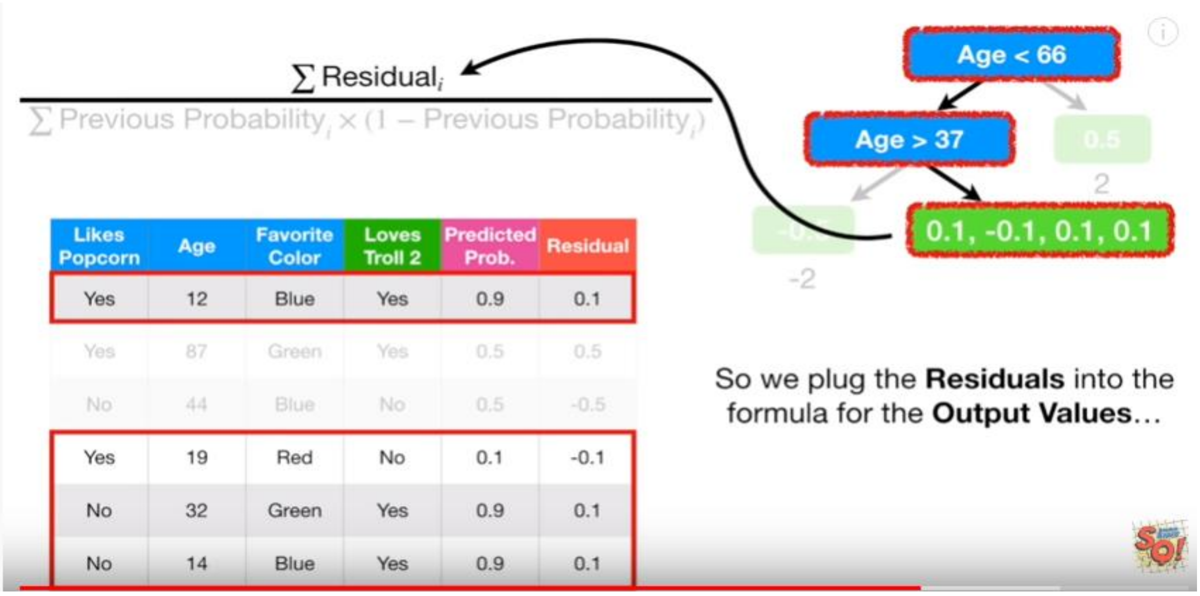

$\frac{-0.5}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | 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
…then we plug in the last **Predicted Probability**…


-0.5 / 0.5 x (1 - 0.5) = -2
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
...do the math...
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

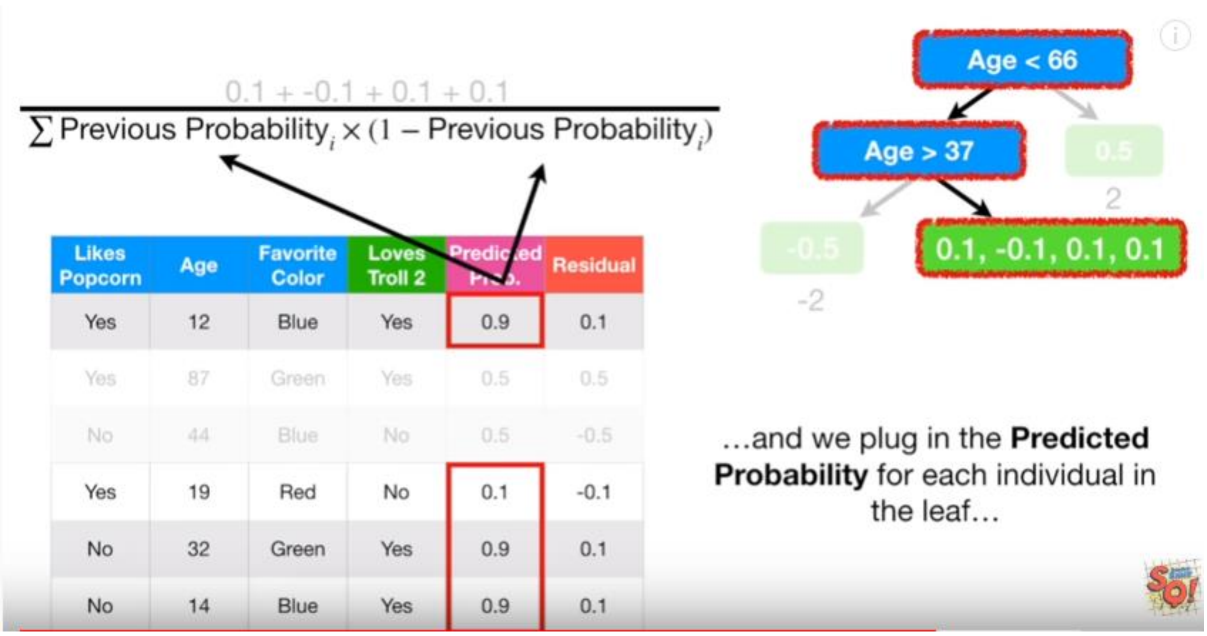

-0.5 / 0.5 x (1 - 0.5) = -2
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
...and the Output Value for this leaf is -2.
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -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


Lastly, let's calculate the Output Value for this leaf...
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
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

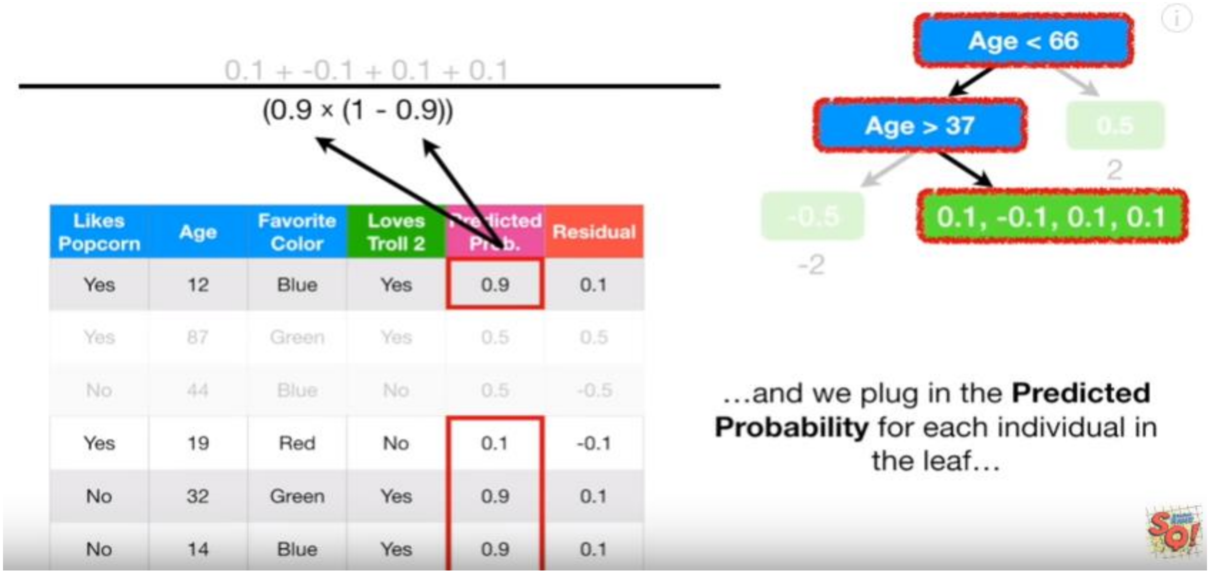
Age < 66
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 0.1 |
| Yes | 87 | Green | Yes | 0.5 | 0.5 |
| No | 44 | Blue | No | 0.5 | -0.5 |
| Yes | 19 | Red | No | 0.1 | -0.1 |
| No | 32 | Green | Yes | 0.9 | 0.1 |
| No | 14 | Blue | Yes | 0.9 | 0.1 |
**NOTE:** A bunch of people…
…go to this leaf.

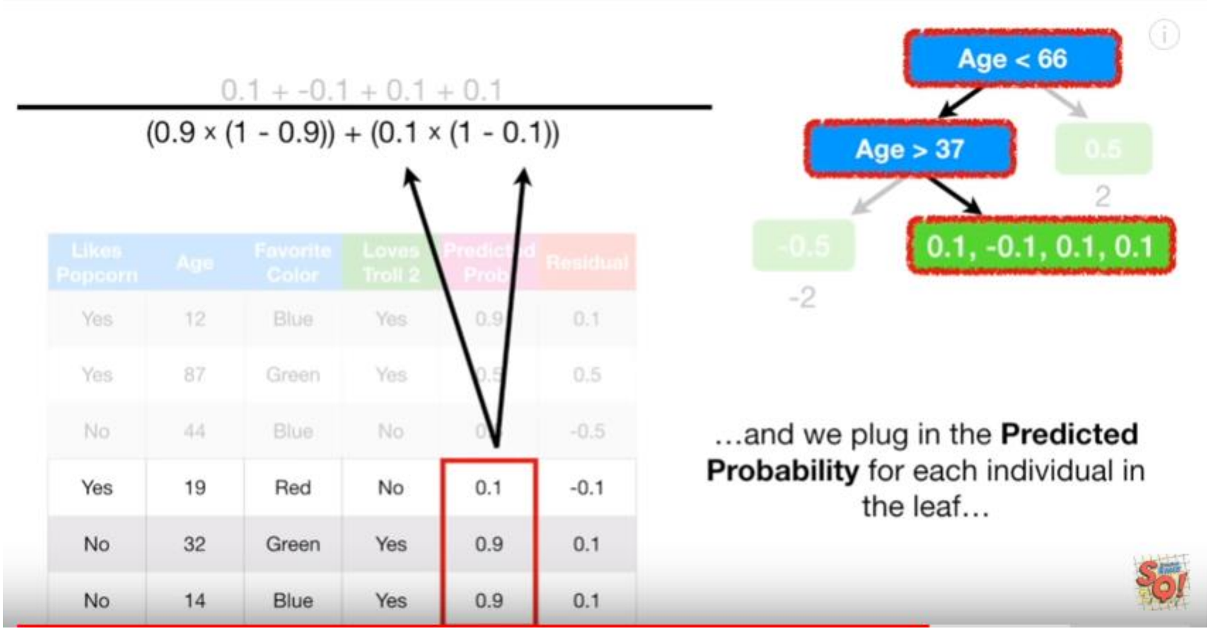



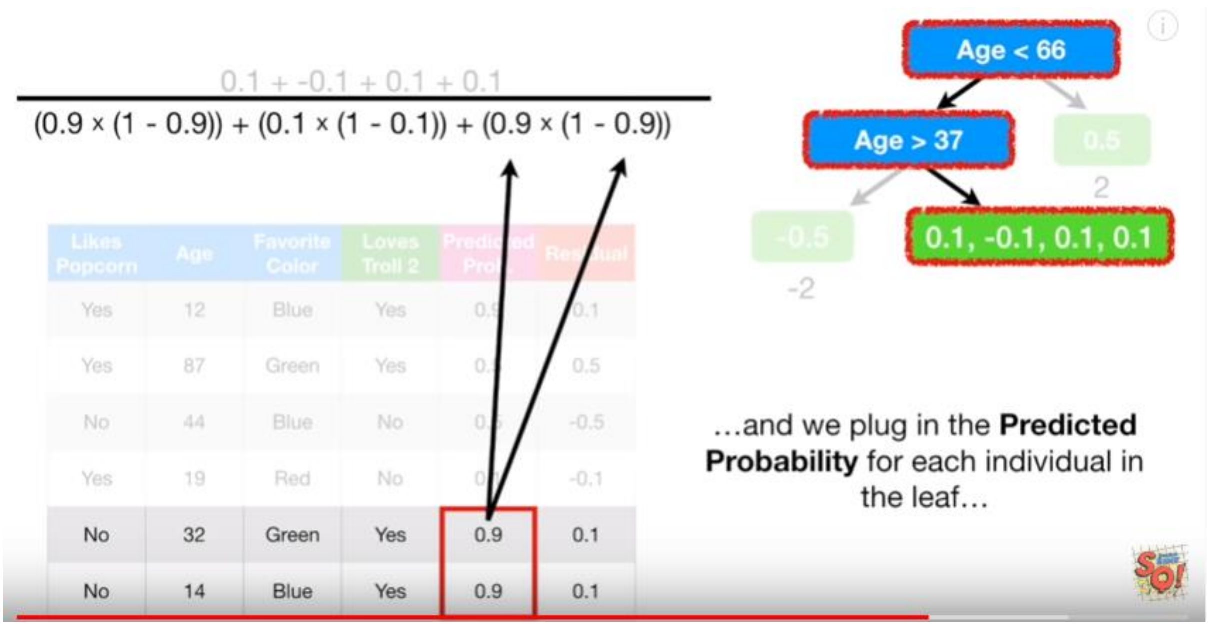
$$\frac{\sum \text{Residual}_i}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$$
Age < 66
Age > 37
0.5
2
-2
0.1, -0.1, 0.1, 0.1
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 0.9 | 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**…

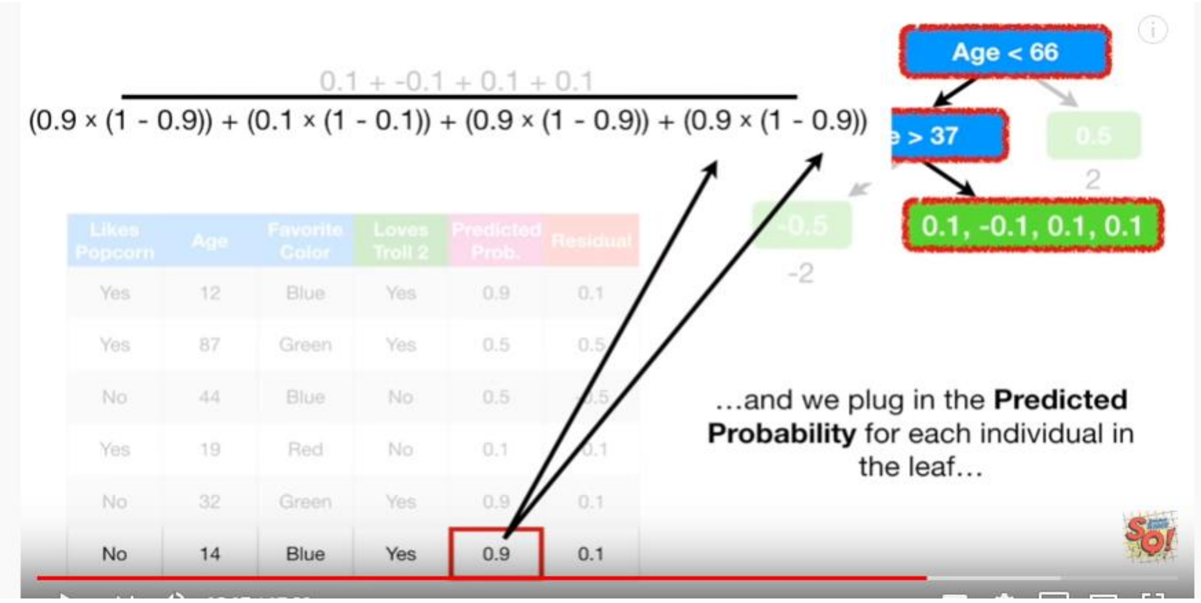



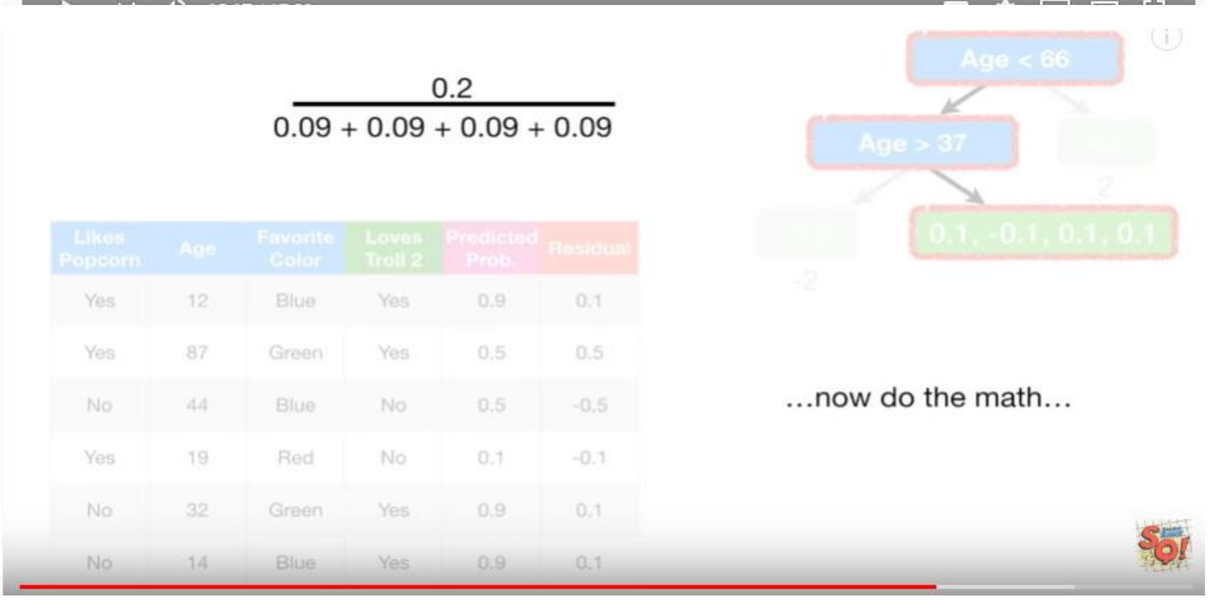
$$\frac{0.1 + -0.1 + 0.1 + 0.1}{\sum \text{Previous Probability}_i \times (1 - \text{Previous Probability}_i)}$$
Age < 66
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
| Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual |
|---|---|---|---|---|---|
| Yes | 12 | Blue | Yes | 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…

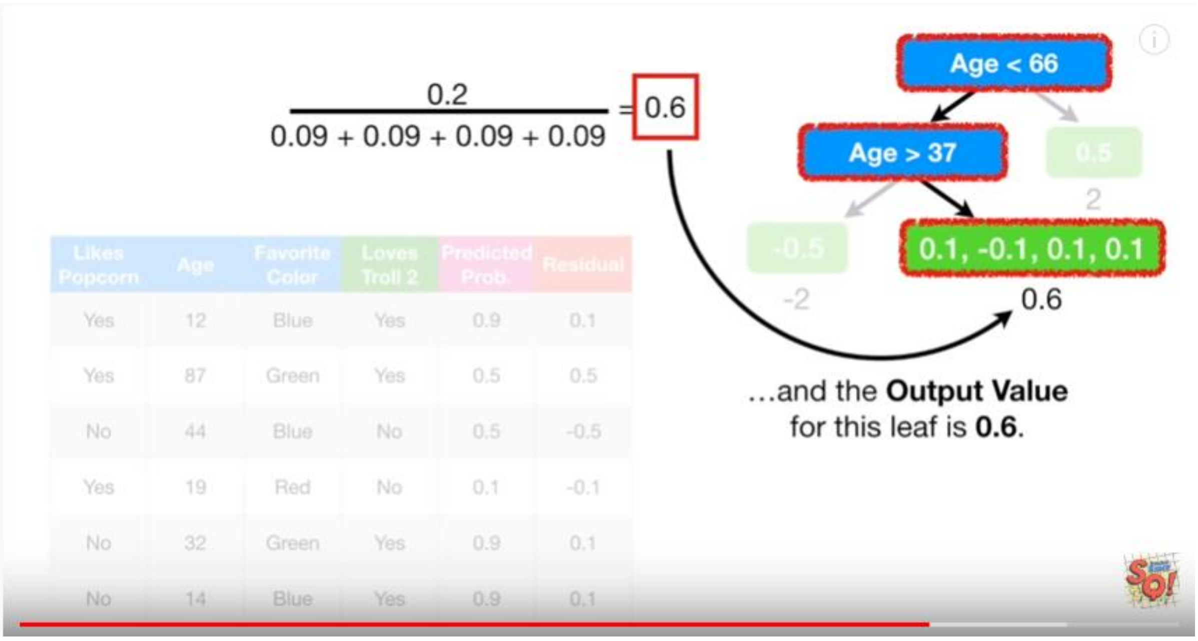

0.1 + -0.1 + 0.1 + 0.1
(0.9 × (1 - 0.9))
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
No | 14 | Blue | Yes | 0.9 | 0.1
Age < 66
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
...and we plug in the **Predicted Probability** for each individual in the leaf...

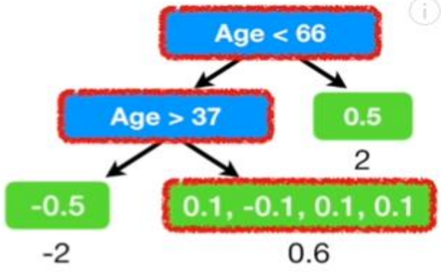

0.1 + -0.1 + 0.1 + 0.1
(0.9 × (1 - 0.9)) + (0.1 × (1 - 0.1))
Likes Popcorn | Age | Favorite Color | Loves Troll 2 | Predicted Prob. | Residual
Yes | 12 | Blue | Yes | 0.9 | 0.1
Yes | 87 | Green | Yes | 0.5 | 0.5
No | 44 | Blue | No | 0.5 | -0.5
Yes | 19 | Red | No | 0.1 | -0.1
No | 32 | Green | Yes | 0.9 | 0.1
No | 14 | Blue | Yes | 0.9 | 0.1
Age < 66
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
...and we plug in the **Predicted Probability** for each individual in the leaf...


0.1 + -0.1 + 0.1 + 0.1
(0.9 × (1 - 0.9)) + (0.1 × (1 - 0.1)) + (0.9 × (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
Age < 66
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
...and we plug in the **Predicted Probability** for each individual in the leaf...


Age < 66
> 37
0.5
2
0.1, -0.1, 0.1, 0.1
-0.5
-2
0.1 + -0.1 + 0.1 + 0.1
(0.9 × (1 - 0.9)) + (0.1 × (1 - 0.1)) + (0.9 × (1 - 0.9)) + (0.9 × (1 - 0.9))
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Predicted Prob.
Residual
No
14
Blue
Yes
0.9
0.1
...and we plug in the **Predicted Probability** for each individual in the leaf...

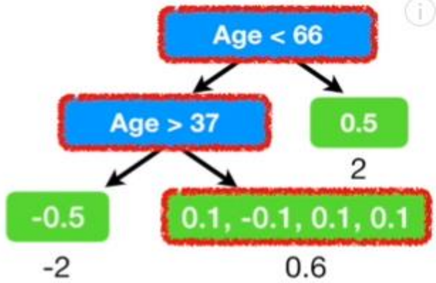

0.2
0.09 + 0.09 + 0.09 + 0.09
...now do the math...


Age < 66
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
0.6
0.2
0.09 + 0.09 + 0.09 + 0.09
= 0.6
...and the **Output Value** for this leaf is **0.6**.



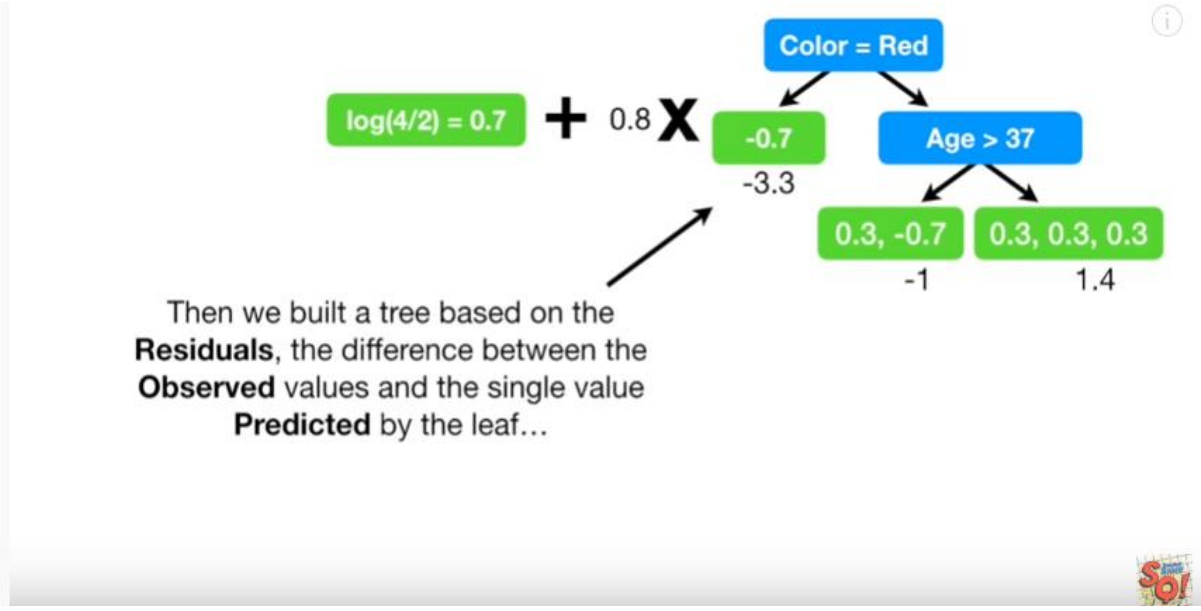

**BAM!**

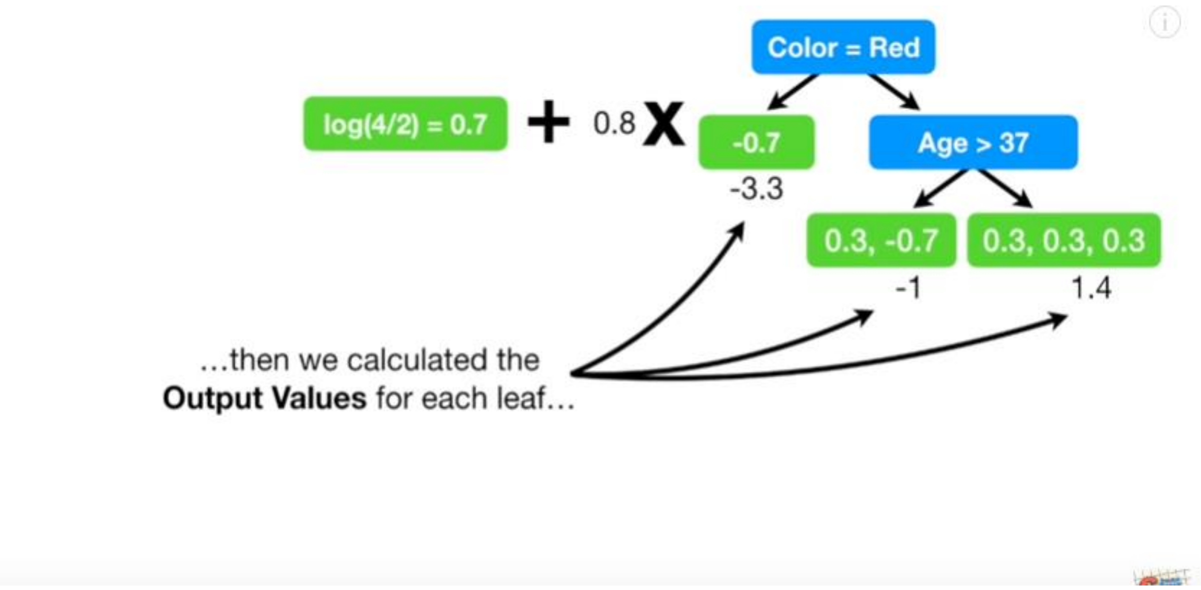
Now that we've calculated all of the **Output Values** for this tree, we can combine it with everything else we've done so far.

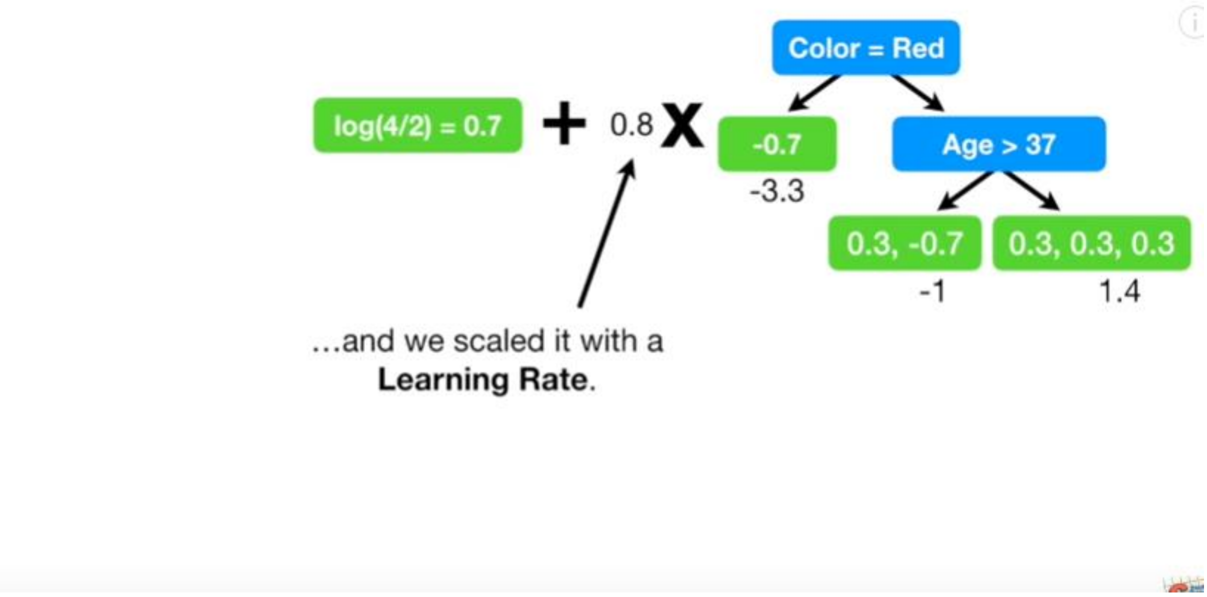


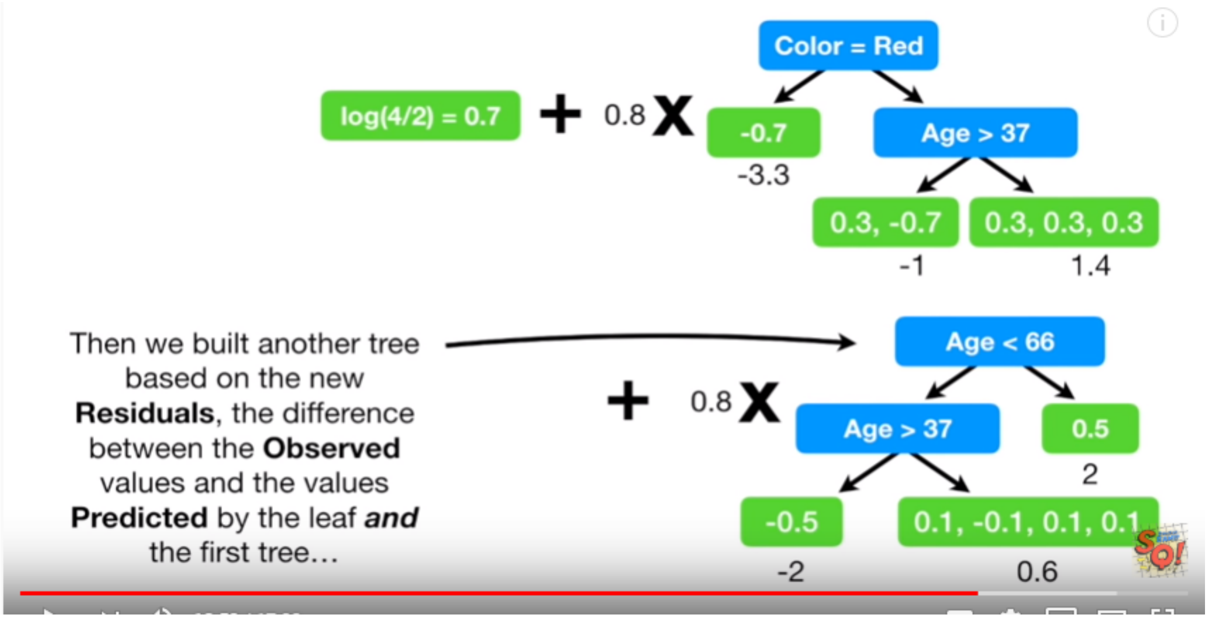


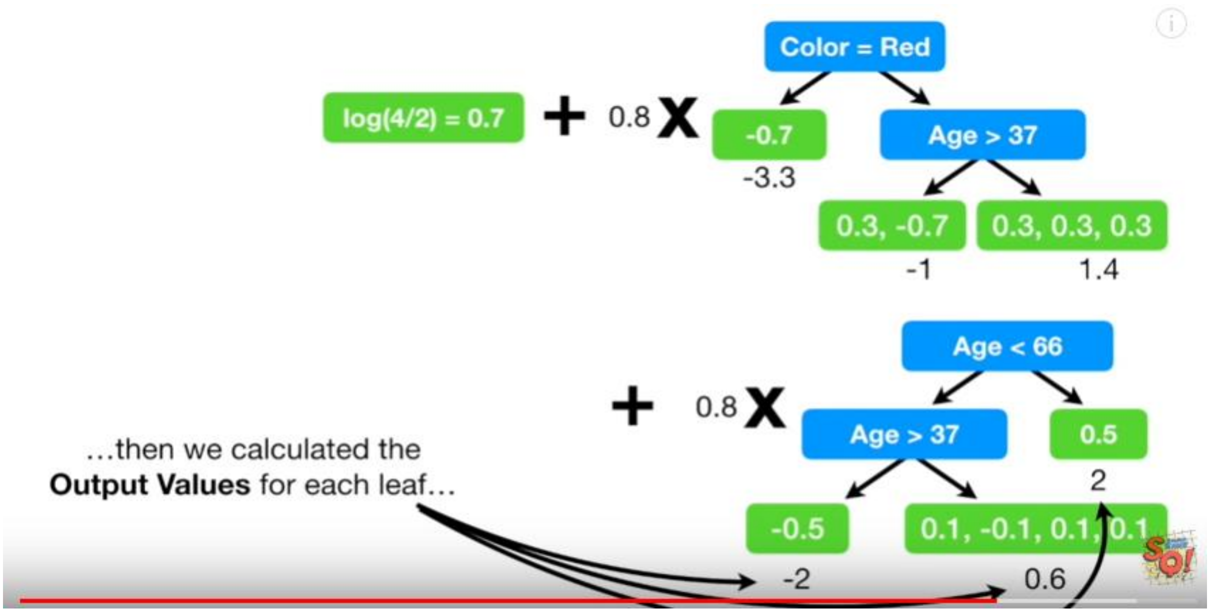
We started with just a leaf, which made one **Prediction** for every individual.

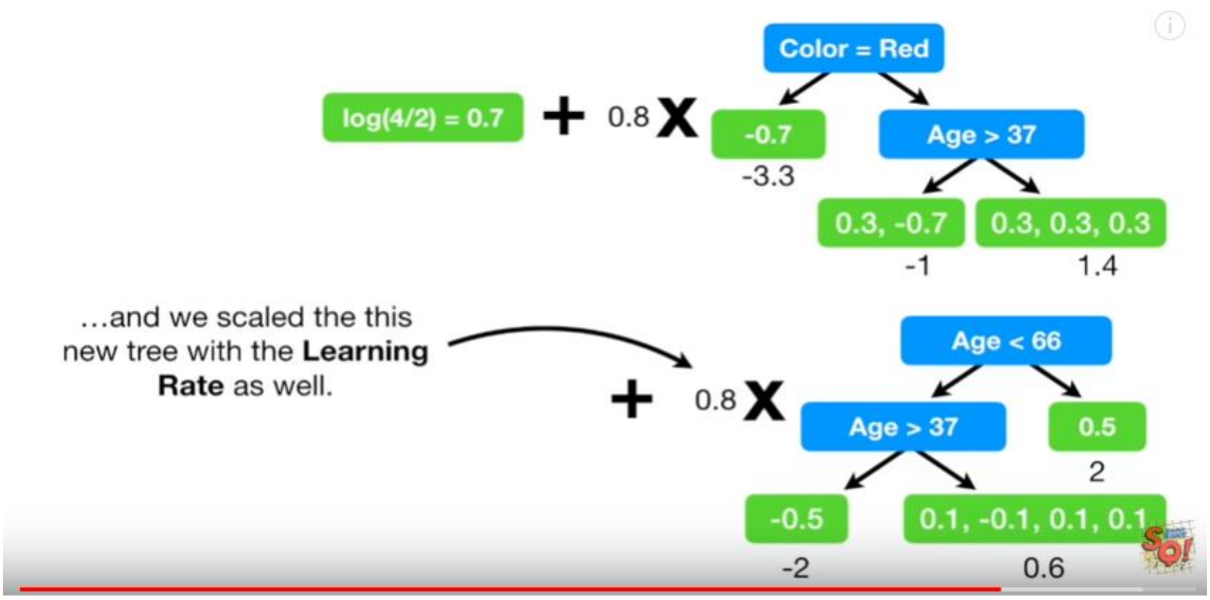

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Then we built a tree based on the **Residuals**, the difference between the **Observed** values and the single value **Predicted** by the leaf…

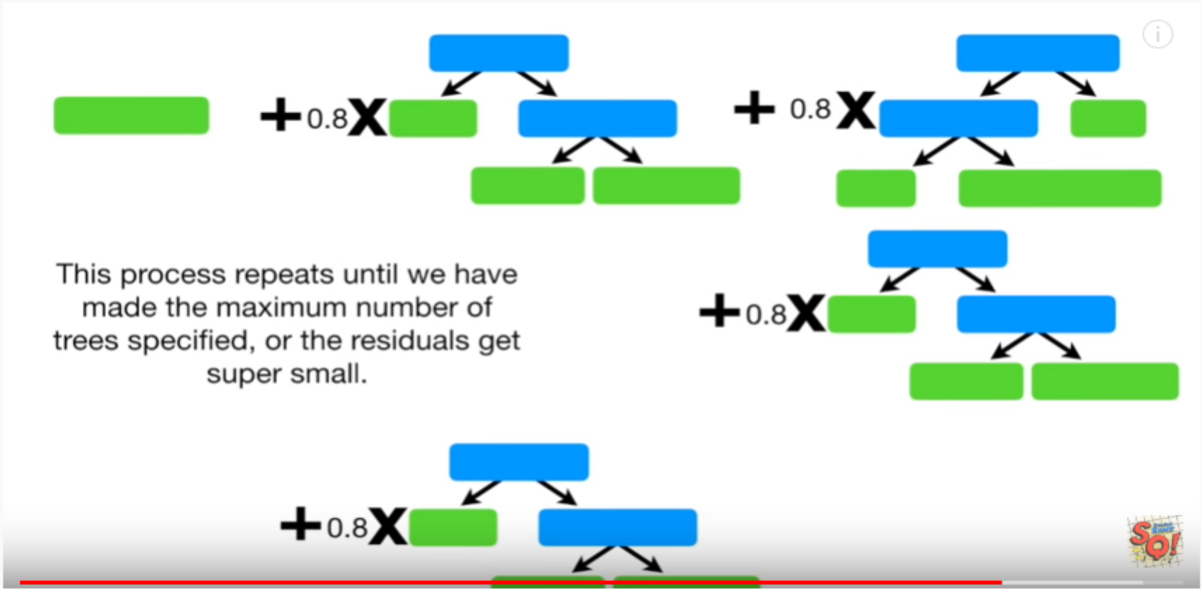

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
…then we calculated the **Output Values** for each leaf…

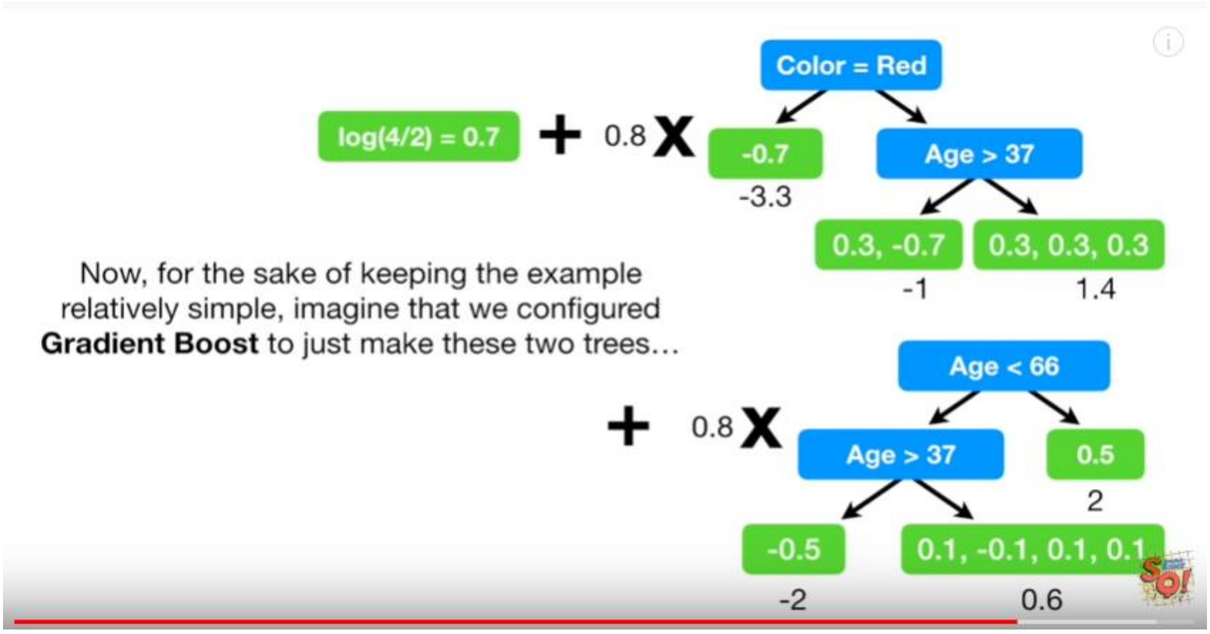

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
…and we scaled it with a **Learning Rate**.

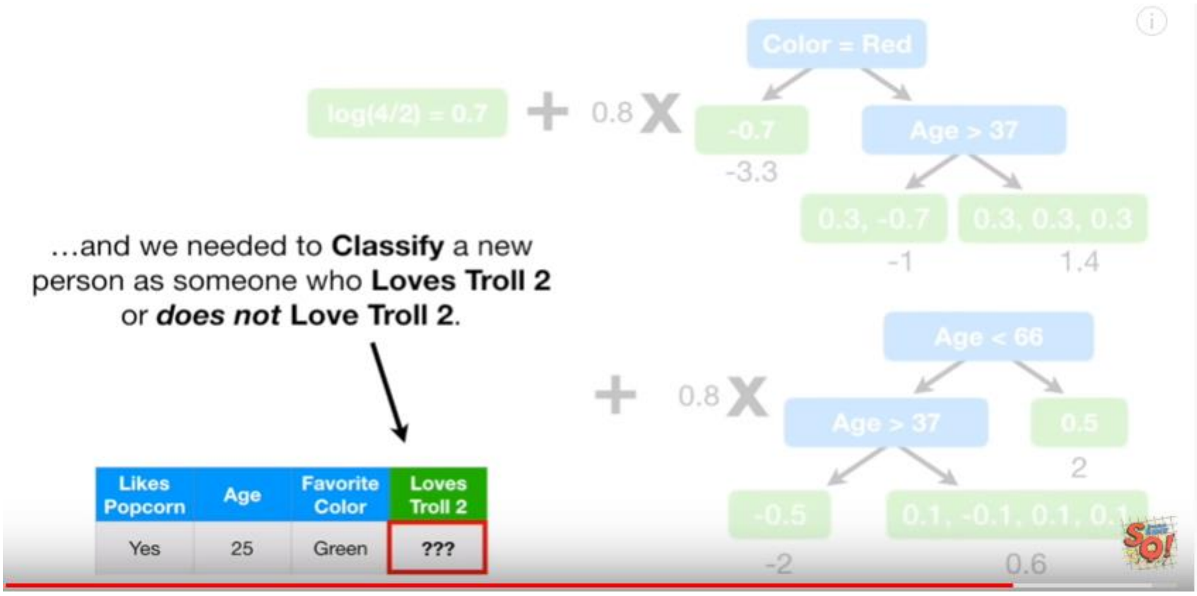

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Then we built another tree based on the new **Residuals**, the difference between the **Observed** values and the values **Predicted** by the leaf ***and*** the first tree…
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
0.6

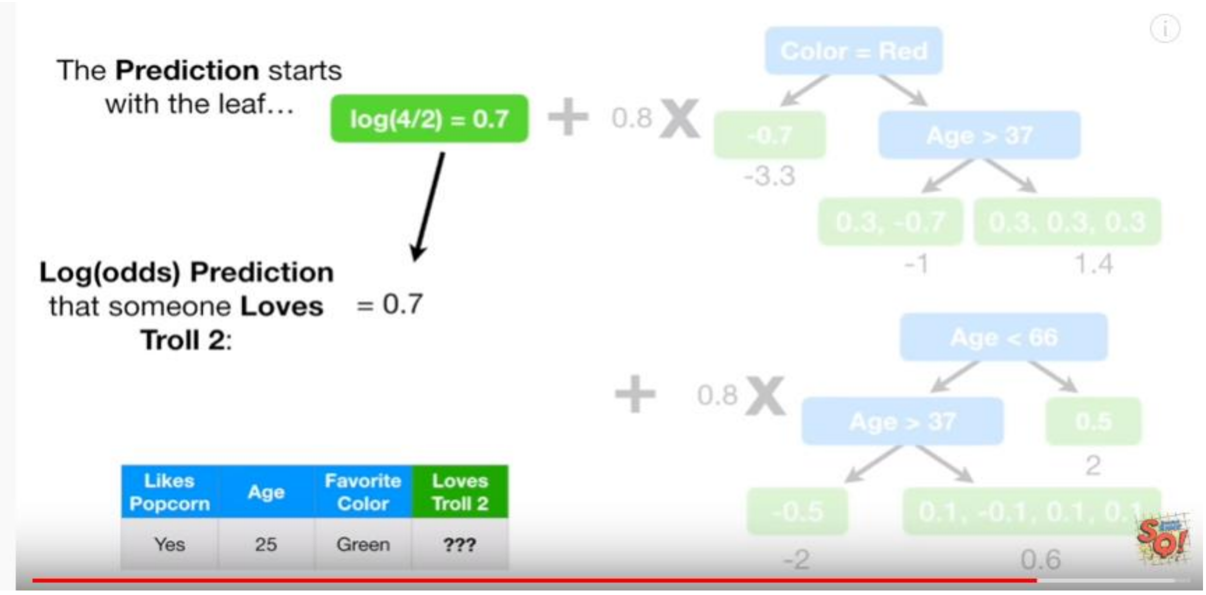

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Age < 66
+ 0.8 X
Age > 37
0.5
2
…then we calculated the **Output Values** for each leaf…
-0.5
-2
0.1, -0.1, 0.1, 0.1
0.6

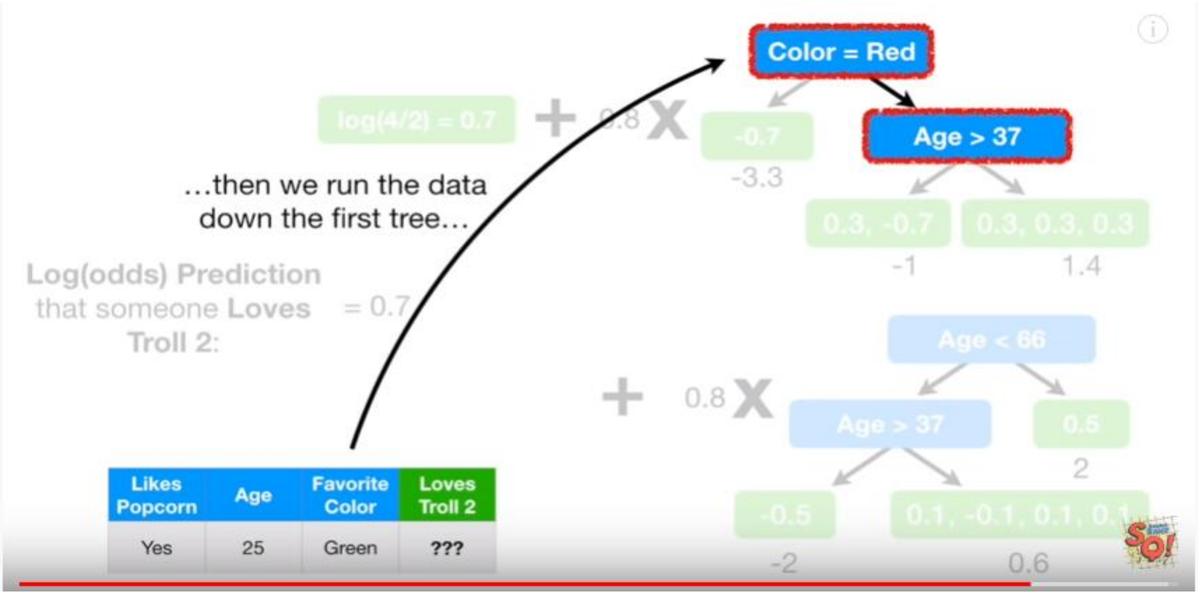

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
…and we scaled the this new tree with the **Learning Rate** as well.
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
0.6

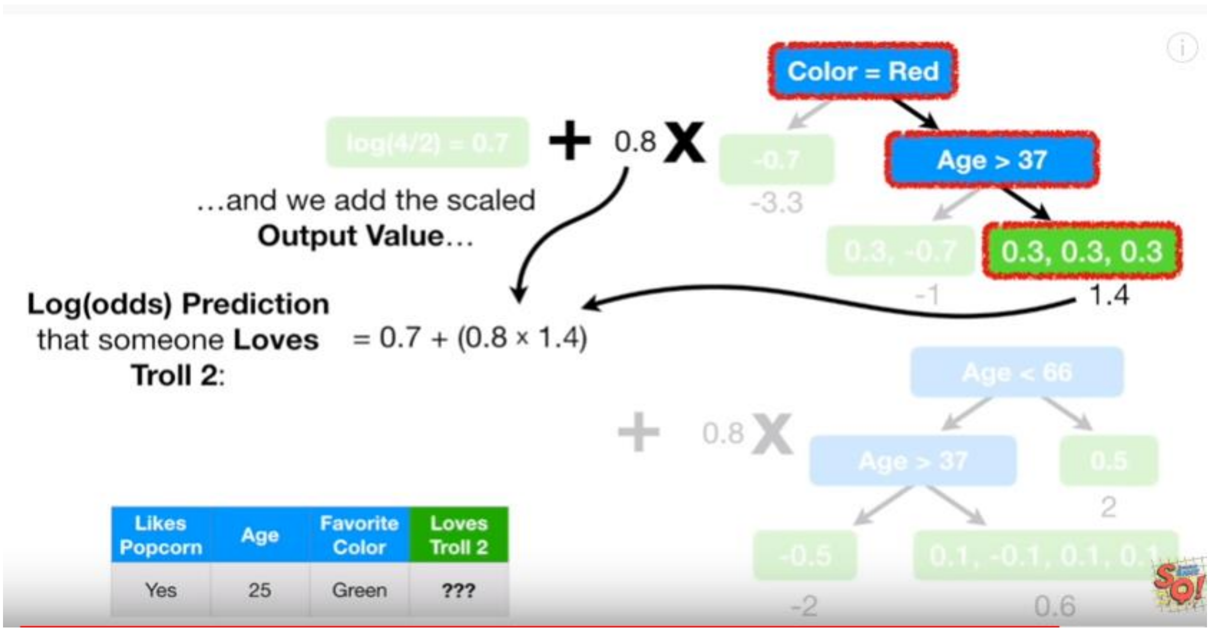

+0.8X
+ 0.8X
This process repeats until we have made the maximum number of trees specified, or the residuals get super small.
+0.8X
+0.8X

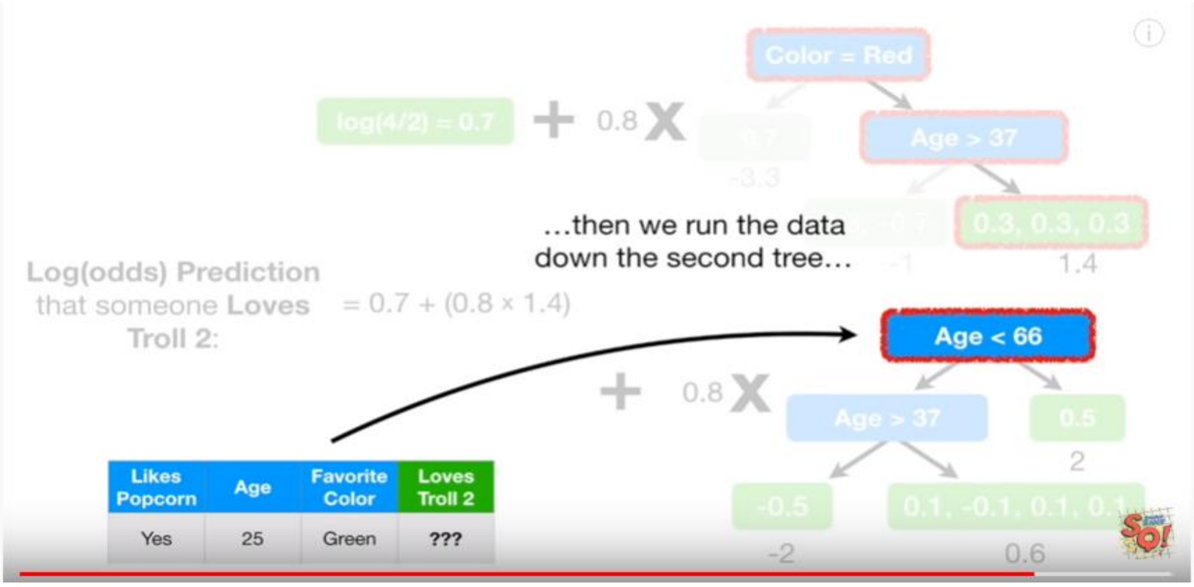

Color = Red
log(4/2) = 0.7
+ 0.8X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
Now, for the sake of keeping the example relatively simple, imagine that we configured **Gradient Boost** to just make these two trees…
Age < 66
+ 0.8X
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
0.6


Color = Red
log(4/2) = 0.7
+ 0.8X
-0.7
-3.3
Age > 37
0.3, -0.7
-1
0.3, 0.3, 0.3
1.4
…and we needed to **Classify** a new person as someone who **Loves Troll 2** or ***does not*** **Love Troll 2**.
Age < 66
+ 0.8X
Age > 37
0.5
2
-0.5
-2
0.1, -0.1, 0.1, 0.1
0.6
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Yes
25
Green
???

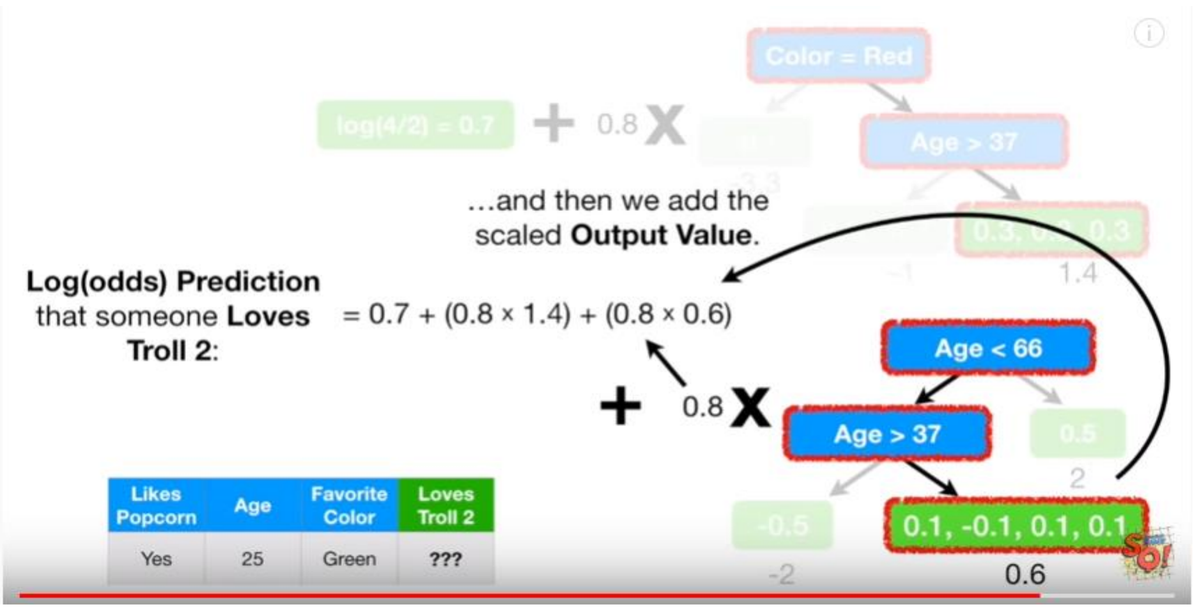

The Prediction starts with the leaf...
log(4/2) = 0.7
+ 0.8 X
Color = Red
-0.7
-3.3
Age > 37
0.3, -0.7
0.3, 0.3, 0.3
-1
1.4
Log(odds) Prediction that someone Loves Troll 2: = 0.7
+ 0.8 X
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???

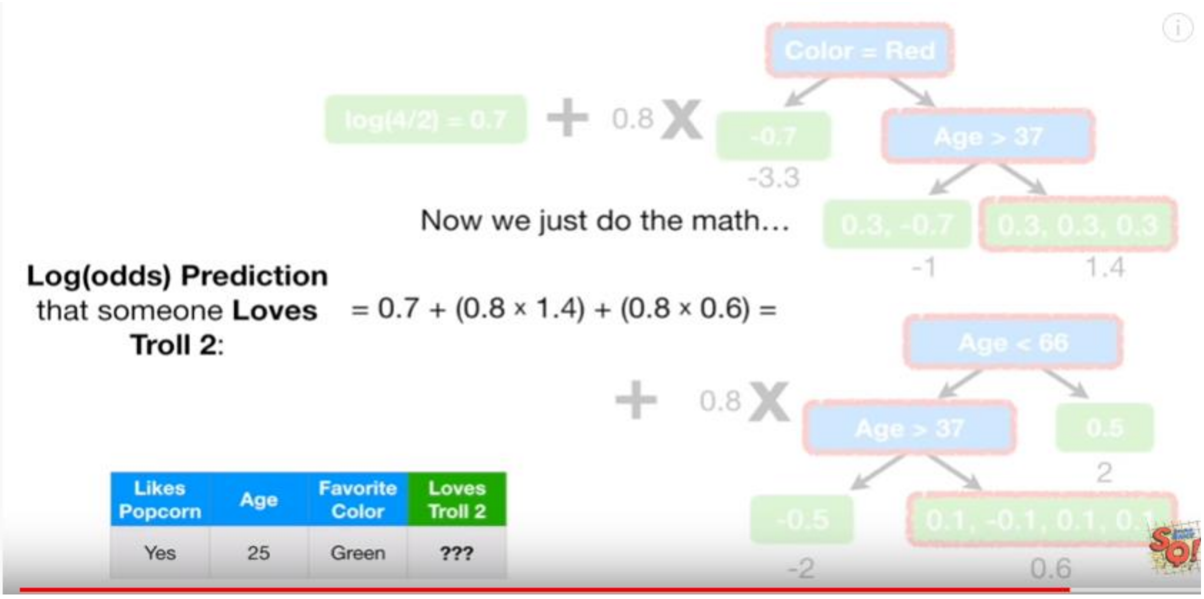

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
Age > 37
...then we run the data down the first tree...
-3.3
0.3, -0.7
0.3, 0.3, 0.3
-1
1.4
Log(odds) Prediction that someone Loves Troll 2: = 0.7
+ 0.8 X
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???

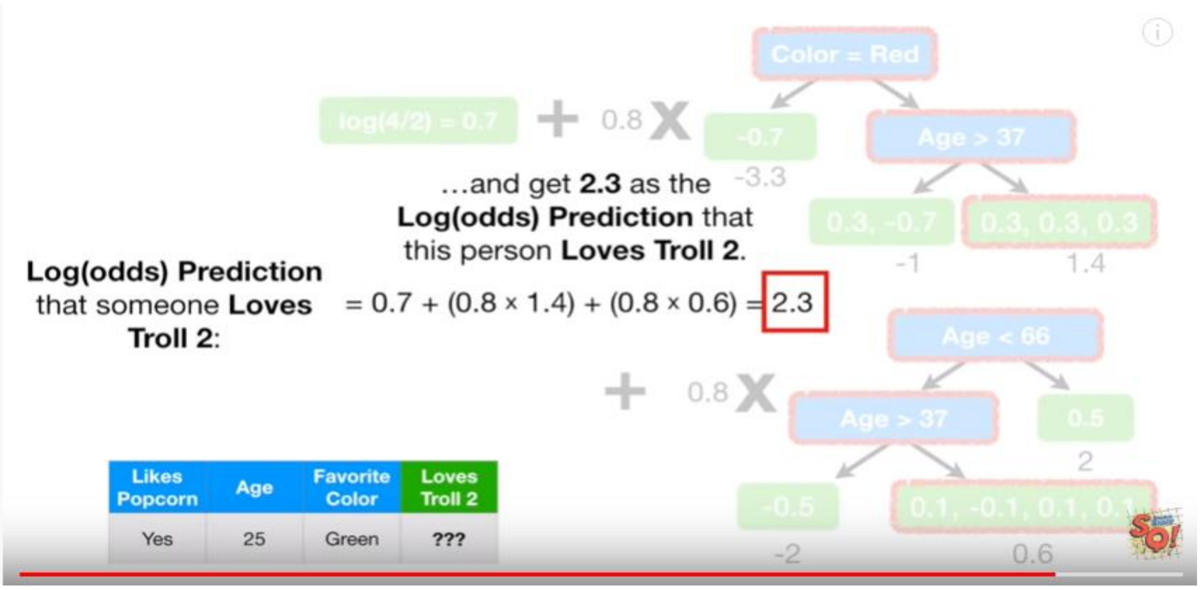

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
Age > 37
...and we add the scaled Output Value...
-3.3
0.3, -0.7
0.3, 0.3, 0.3
-1
1.4
Log(odds) Prediction that someone Loves Troll 2: = 0.7 + (0.8 × 1.4)
+ 0.8 X
Age < 66
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???

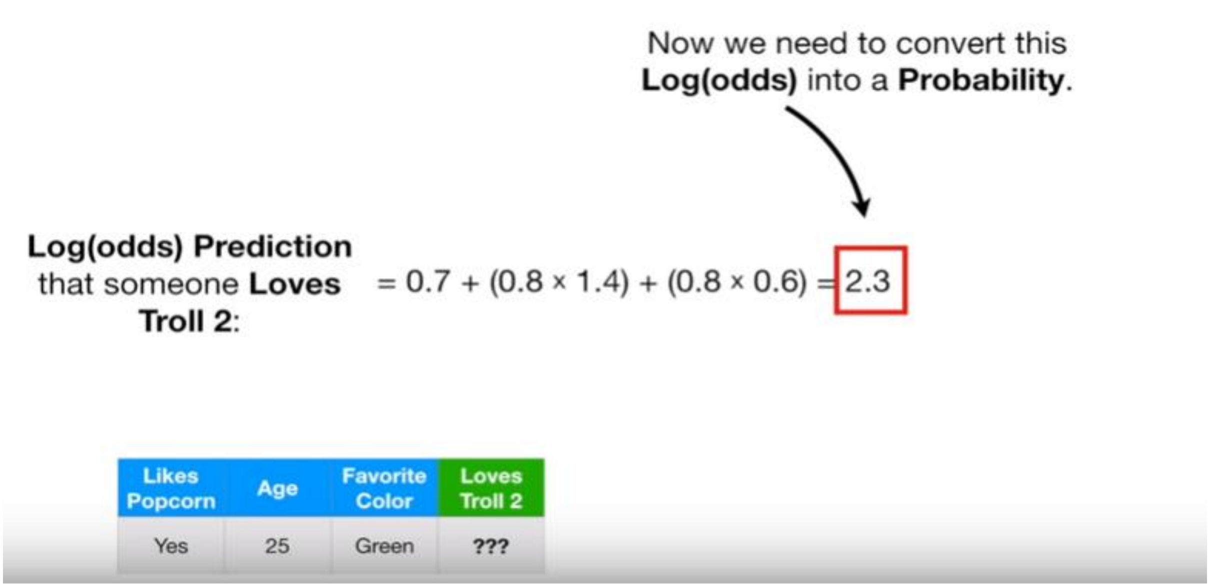

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
-0.7
0.3, 0.3, 0.3
-1
1.4
…then we run the data
down the second tree…
Log(odds) Prediction
that someone Loves
Troll 2:
= 0.7 + (0.8 × 1.4)
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Yes
25
Green
???

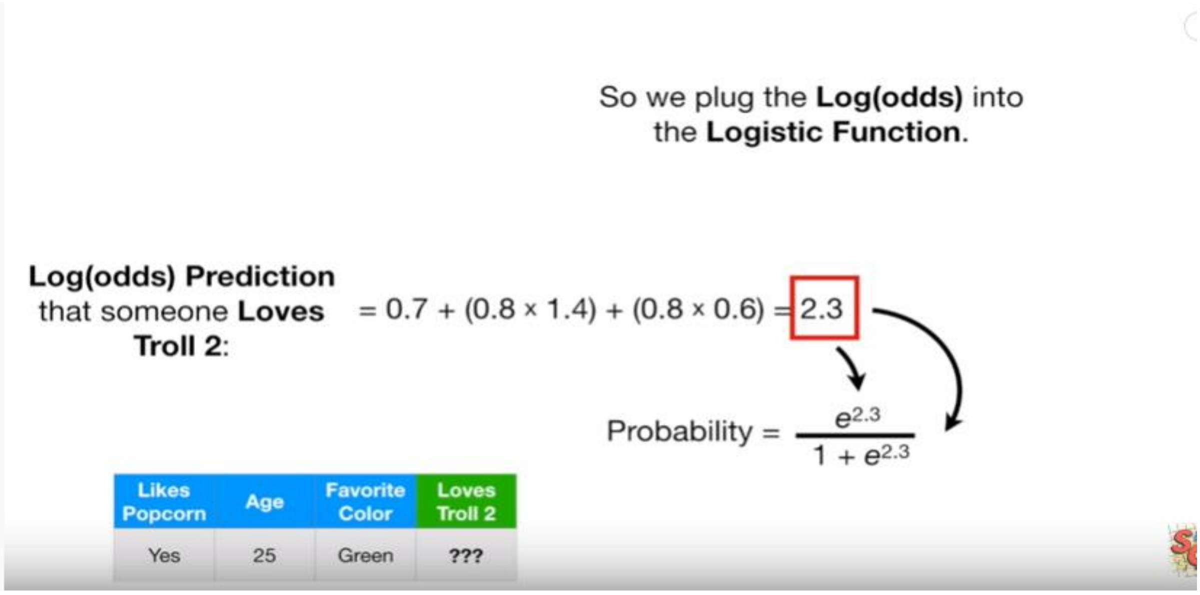

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
-0.7
0.3, 0.3, 0.3
-1
1.4
…then we run the data
down the second tree…
Log(odds) Prediction
that someone Loves
Troll 2:
= 0.7 + (0.8 × 1.4)
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Yes
25
Green
???

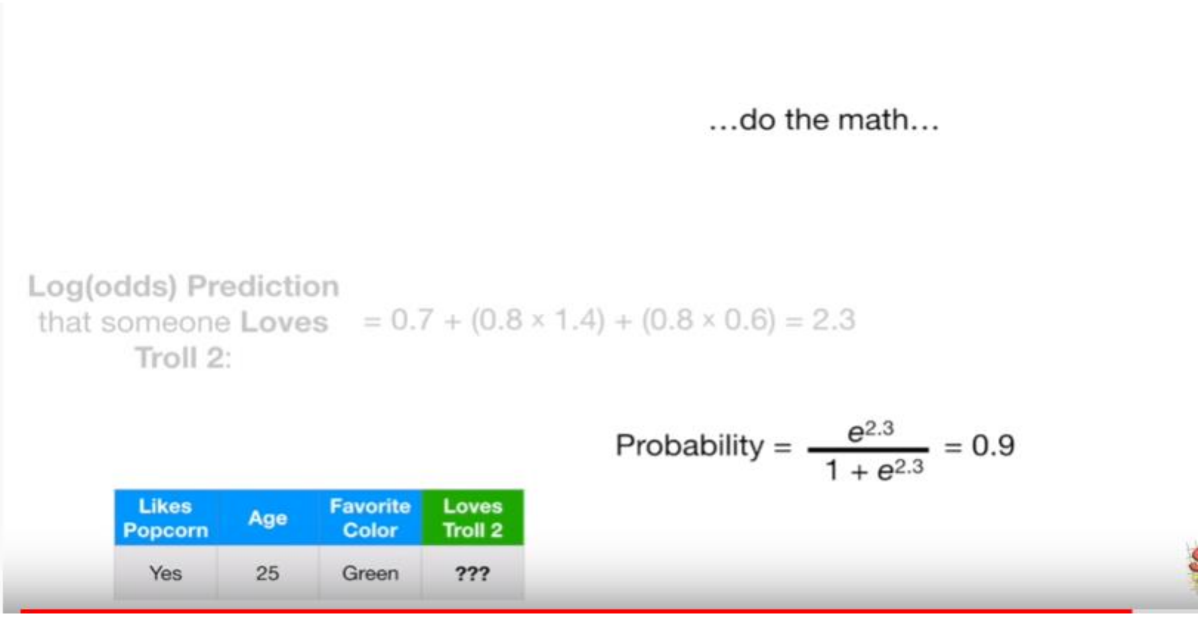

Color = Red
log(4/2) = 0.7
+ 0.8 X
Age > 37
-3.3
…and then we add the
scaled Output Value.
0.3, 0.3, 0.3
-1
1.4
Log(odds) Prediction
that someone Loves
Troll 2:
= 0.7 + (0.8 × 1.4) + (0.8 × 0.6)
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn
Age
Favorite Color
Loves Troll 2
Yes
25
Green
???

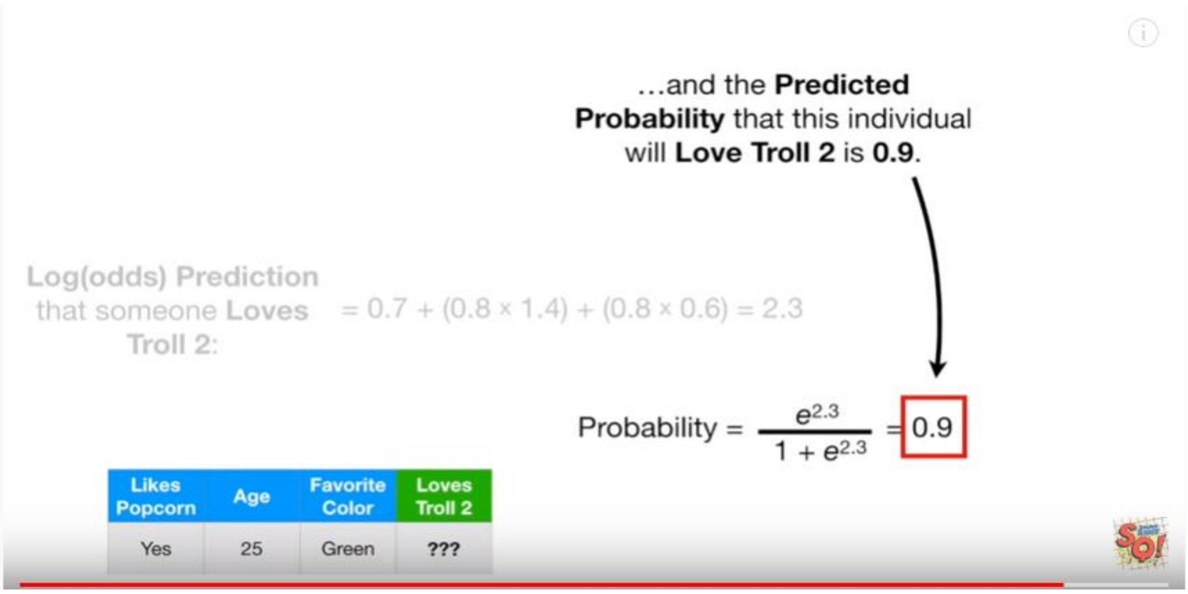

Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
Now we just do the math...
0.3, -0.7
0.3, 0.3, 0.3
-1
1.4
Log(odds) Prediction that someone Loves Troll 2:
= 0.7 + (0.8 × 1.4) + (0.8 × 0.6) =
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???


Color = Red
log(4/2) = 0.7
+ 0.8 X
-0.7
-3.3
Age > 37
...and get 2.3 as the Log(odds) Prediction that this person Loves Troll 2.
0.3, -0.7
0.3, 0.3, 0.3
-1
1.4
Log(odds) Prediction that someone Loves Troll 2:
= 0.7 + (0.8 × 1.4) + (0.8 × 0.6) = 2.3
Age < 66
+ 0.8 X
Age > 37
0.5
2
-0.5
0.1, -0.1, 0.1, 0.1
-2
0.6
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???


Now we need to convert this Log(odds) into a Probability.
Log(odds) Prediction that someone Loves Troll 2:
= 0.7 + (0.8 × 1.4) + (0.8 × 0.6) = 2.3
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???


So we plug the Log(odds) into the Logistic Function.
Log(odds) Prediction that someone Loves Troll 2:
= 0.7 + (0.8 × 1.4) + (0.8 × 0.6) = 2.3
Probability = $\frac{e^{2.3}}{1 + e^{2.3}}$
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???


…do the math…
Log(odds) Prediction that someone Loves Troll 2:
= 0.7 + (0.8 × 1.4) + (0.8 × 0.6) = 2.3
Probability = $\frac{e^{2.3}}{1 + e^{2.3}}$ = 0.9
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 × 1.4) + (0.8 × 0.6) = 2.3
Probability = $\frac{e^{2.3}}{1 + e^{2.3}}$ = 0.9
Likes Popcorn | Age | Favorite Color | Loves Troll 2
Yes | 25 | Green | ???

Since we are using **0.5** as our threshold for deciding how to **Classify** people, and **0.9 > 0.5**...

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!!! |

**NOTE:** Before we go, I want to remind you that **Gradient Boost** usually uses trees with between **8** to **32** to leaves.

We used small trees in this **StatQuest** because our **Training Dataset** was super small.