Gradient Boosting Classification

22 March 2022 10:38

Classification Main Ideas

- When gradient Boost is used for classification, it has lot in common with logistic Regression

Training 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

→ Walk through Step-By-Step, the most common way the gradient Boost fits the training Data.

1. We start with a leaf that represents an initial Prediction for every individual.

   - When we use Gradient Boost for classification, the initial Prediction for every individual is "log (Odds)".

   - Calculate log(Odds) that someone loves "Troll-2".
   - Since 4 people in the training dataset loves Troll-2, & 2 people do not.

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

- Which will be our Initial leaf : $\boxed{\log(4/2) = 0.7}$ Leaf

- Just like logistic regression, the easiest way to use the log(Odds) for classification is to convert it to probability.

⇒ Prob of loving Troll 2 = $e^{\log(4/2)}$ ... 0.7.

$\Rightarrow$ Prob of loving Troll 2 $= \dfrac{e^{\log(4/2)}}{1+e^{\log(4/2)}} = 0.7.$

- Since the probability of "loving Troll 2" is greater than 0.5, we can classify everyone in the training dataset as someone who loves Troll-2.

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.

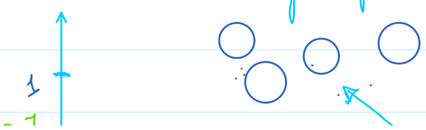
- Now classifying everyone in the training data as someone who loves troll 2 is pretty lame, because two of the people do not love the movie.

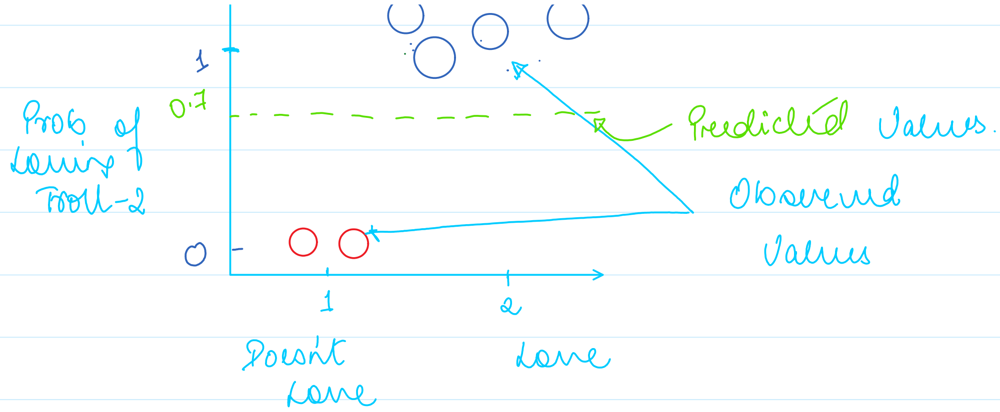
2. We can measure how bad the initial Prediction is by calculating Pseudo Residuals, the difference between the Observed and the Predicted values

$$\text{Residuals} = (\text{Observed} - \text{Predicted})$$

Likes Popcorn	Age	Favorite color	Loves Troll 2	Residuals
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

- Although the math is easy, I think it's easier to grasp what's going on if we draw the Residuals on a graph.

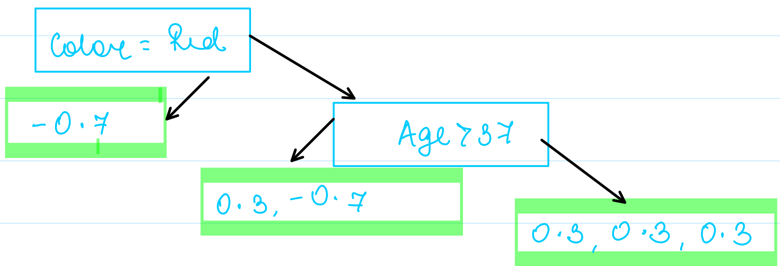




- The red dots, with Prob of Loving Troll-2 = 0, 2 people
- Blue dots, Prob of Loving Troll 2 = 1, represent the 4 people

3. Now we build a tree, using Likes Popcorn, Age, fav color to predict residuals.

Tree-1.



Likes Popcorn	Age	Favorite color	Loves Troll 2	Residuals	
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 practice, people often set the maximum no of leaves to be between 8 & 32.

4. Now let's calculate the output values of the leaves

[- When we used Gradient Boost for Regression, a leaf with single Residual had an Output value equal to the Residual]

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

- We need to transform the leaf values.

$$\frac{\sum \text{Residuals}_i \ (\text{All residuals of the leaf})}{\sum [\text{Previous Probability}_i \times (1 - \text{Previous Prob}_i)]}$$

3 Leaves

1. Leaf 1 - (-0.7) $= \dfrac{-0.7}{0.7 \times (1-0.7)} = -3.3$

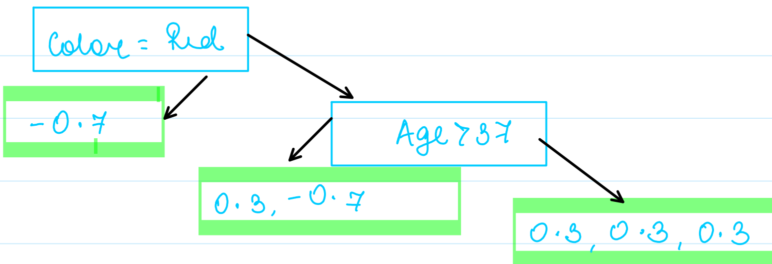
2. Leaf-2 $= \dfrac{0.3 + (-0.7)}{2 \cdot 0.7 \cdot (1-0.7)} = -0.95 \approx -1$

3. Leaf-3 $= \dfrac{0.3 + 0.3 + 0.3}{3 \times 0.7\,(1-0.7)} = 1.42$

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

Learning rate

$\boxed{\log(4/2) = 0.7} + 0.8 \times$



° Now let's calculate the log (odds) Prediction for this person.

(-3.3, -1, 1.42)

→ log(odds) prediction + Learning-rate × Output value.

→ $0.7 + (0.8 \times 1.4) = 1.8$.

° Now we convert the new log(odds) Prediction into

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

$$\text{Probability} = \frac{e^{1.8}}{1+e^{1.8}} = 0.858 \approx 0.9$$

Likes Popcorn	Age	Favorite color	Loves Troll 2	Residuals	Predicted Probability
Yes	12	Blue	Yes	0.3	0.9
Yes	87	Green	Yes	0.3	0.5
No	44	Blue	No	-0.7	0.5
Yes	19	Red	No	-0.7	0.1
No	32	Green	Yes	0.3	0.9
No	14	Blue	Yes	0.3	0.9

- Now we calculate the new log(Odds) Prediction for the second person.

$$0.7 + 0.8 \times (-1) = -0.1.$$

$$\text{Prob} = \frac{e^{-0.1}}{1+e^{-0.1}} = 0.5.$$

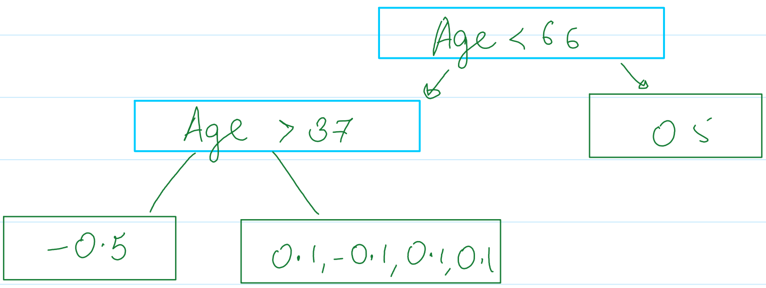
- Third Person - $0.7 + 0.8\,(-3.3) = -1.94$

$$\text{Prob} = \frac{e^{-1.94}}{1+e^{-1.94}} = 0.12 \approx 0.1$$

→ Now just like before, we calculate the new residuals

Likes Popcorn	Age	Favorite color	Loves Troll 2	Residuals	Predicted Probability	New Residuals
Yes	12	Blue	Yes	0.3	0.9	0.1
Yes	87	Green	Yes	0.3	0.5	0.5
No	44	Blue	No	-0.7	0.5	-0.5
Yes	19	Red	No	-0.7	0.1	-0.1
No	32	Green	Yes	0.3	0.9	0.1
No	14	Blue	Yes	0.3	0.9	0.1

→ Now, that we have the residuals, we built a new Tree



1. Leaf : $\frac{0.5}{0.5(1-0.5)} = \frac{1}{0.5} \Rightarrow 2$, $0.7 + 0.5 + 0.8 \times 0.5$

$= 0.12 + 0.4 = 0.52$

$Prob = \frac{e^{0.52}}{1+e^{0.52}} = 0.62.$

2 Leaf . $\frac{0.1 - 0.1 + 0.1 + 0.1}{0.9(1-0.9) + (-0.1)(1+0.1) + 0.9(1-0.9) + 0.9(1-0.9)}$

$= \frac{0.2}{0.9 \cdot 0.1 - 0.1 \cdot 0.9 + 0.9 \times 0.1 + 0.9 \times 0.1} = 0.6$

$Prob_1 = 0.7 + 0.9 + 0.8 \times 0.6 = 2.08$ [3 Rows]

$0.7 + 0.1 + 0.8 \times 0.6 = 1.28$ [1 Row.

$Prob = \frac{e^{2.08}}{1+e^{2.08}} = 0.88$

$Prob = \frac{e^{1.28}}{1+e^{1.28}} = 0.78$

3. Leaf 3 : $\frac{-0.5}{0.5(1-0.5)} \Rightarrow -2$

$\Rightarrow 0.7 + 0.5 + 0.8 \times 2 = 2.8$

$Prob = \frac{e^{2.8}}{1+e^{2.8}} = 0.94$

Likes Popcorn	Age	Favorite color	Loves Troll 2	Residuals	Predicted Probability	New Residuals	Predicted Probability
Yes	12	Blue	Yes	0.3	0.9	0.1	0.88
Yes	87	Green	Yes	0.3	0.5	0.5	0.62
No	44	Blue	No	-0.7	0.5	-0.5	0.94
Yes	19	Red	No	-0.7	0.1	-0.1	0.78
No	32	Green	Yes	0.3	0.9	0.1	0.88
No	14	Blue	Yes	0.3	0.9	0.1	0.88