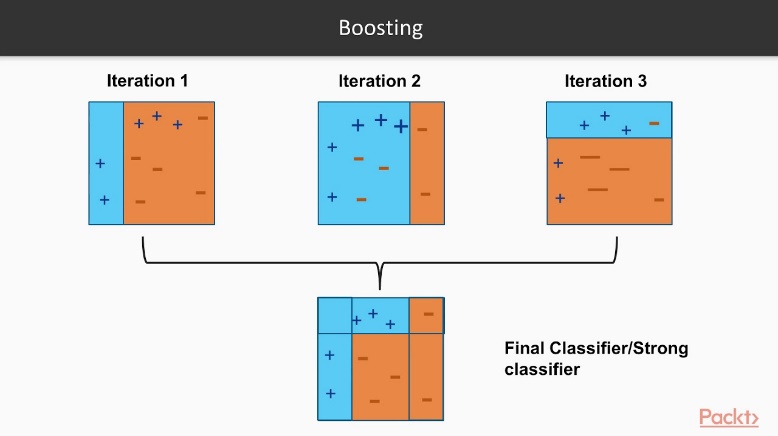
**AdaBoost:-**

AdaBoost, short for Adaptive Boosting, is a machine learning ensemble algorithm that combines multiple **weak classifiers** to create a **strong classifier**. It works by iteratively training weak classifiers on **different subsets** of the training data, with each subsequent classifier giving more weight to the misclassified examples from previous iterations. In this way, AdaBoost focuses on the challenging examples and improves the overall performance of the ensemble by assigning higher importance to the instances that are difficult to classify.

The final prediction is made by **aggregating the predictions** of all the weak classifiers, weighted by their individual performance. AdaBoost is known for its ability to handle complex classification tasks and achieve high accuracy.



1. **Weak Learner:- one of the machine learning models from a set of boosting models whose accuracy is just above 50%**
2. **Decision Stump:- whose max depth is one.**

**Steps**

1. Identify the feature with the maximum information gain as the "root feature" for the decision tree.
2. Perform the Decision tree and find out the predicted value.
3. Assign a weight to each row using the formula 1/n, where n is the total number of features.
4. Calculate the “performance of Stump”() using the below formula.
5. Recalculate the weight using the Exponential Formula and assign it to each row with the formula given below. Please note for correct and incorrectly classified points there is a different formula.
6. Then Normalize the weight by dividing it by the “sum of weight”.
7. Create bins for each row. Note that a row with the maximum bin size will have the highest chance of being selected for the next stump.
8. Repeat the above steps for the desired number of rotations or iterations to create the decision tree.
9. Finally obtain the result by using the below formula.

**Perform the Boosting on the below dataset.**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No | Salary | Credit | Approval |
| 1 | 30K | Bad | NO |
| 2 | 50k | Good | YES |
| 3 | 50k | Neutral | YES |
| 4 | 30K | Neutral | NO |
| **5** | 50k | Good | YES |
| **6** | 50k | Neutral | YES |
| 7 | 30K | Bad | NO |
| 8 | 50k | Good | YES |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr No | Salary | Credit | Approval | Predicted  Value |
| 1 | 30K | Bad | NO | NO |
| 2 | 50k | Good | YES | YES |
| 3 | 50k | Neutral | YES | YES |
| 4 | 30K | Neutral | NO | NO |
| **5** | 50k | Good | YES | NO |
| **6** | 50k | Neutral | YES | NO |
| 7 | 30K | Bad | NO | NO |
| 8 | 50k | Good | YES | YES |

**Step01:** Identify the feature with the maximum information gain as the "root feature" for the decision tree. In the above example, Salary is the feature that has maximum information gain.

**Step02:** Perform the Decision tree and find out the predicted value.

Below is the dataset with the predicted value where rows 5 and 6 were wrongly predicted.

-----

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sr No | Salary | Credit | Approval | Predicted  Value | Weight |
| 1 | 30K | Bad | NO | NO | 0.125 |
| 2 | 50k | Good | YES | YES | 0.125 |
| 3 | 50k | Neutral | YES | YES | 0.125 |
| 4 | 30K | Neutral | NO | NO | 0.125 |
| **5** | 50k | Good | YES | NO | 0.125 |
| **6** | 50k | Neutral | YES | NO | 0.125 |
| 7 | 30K | Bad | NO | NO | 0.125 |
| 8 | 50k | Good | YES | YES | 0.125 |
| Total | | | | |  |

**Step03:** Assign a weight to each row using the formula 1/n, where n is the total number of features.

Here weight is calculated using the below formula and assigned to each row.

**Step04:** Calculate the & α using the below formula.

1st calculate the TE by using the below formula.

Now calculate the by using the below formula.

=

=

=

=

=

=

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sr No | Salary | Credit | Approval | Predicted  Value | Weight | Weight After  Calculation |
| 1 | 30K | Bad | NO | NO | 0.125 | **0.072** |
| 2 | 50k | Good | YES | YES | 0.125 | **0.072** |
| 3 | 50k | Neutral | YES | YES | 0.125 | **0.072** |
| 4 | 30K | Neutral | NO | NO | 0.125 | **0.072** |
| **5** | 50k | Good | YES | NO | 0.125 | 0.217 |
| **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 |
| 7 | 30K | Bad | NO | NO | 0.125 | **0.072** |
| 8 | 50k | Good | YES | YES | 0.125 | **0.072** |
| Total | | | | | 1.000 |  |

**Step05:** Recalculate the weight using the below formula. (formula given below)

|  |  |
| --- | --- |
| exponential( e ) | 2.71828 |
| Performance of Stump | 0.5493 |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sr No | Salary | Credit | Approval | Predicted  Value | Weight | Weight After  Calculation | Normalize the  weight |
| 1 | 30K | Bad | NO | NO | 0.125 | **0.072** | **0.083** |
| 2 | 50k | Good | YES | YES | 0.125 | **0.072** | **0.083** |
| 3 | 50k | Neutral | YES | YES | 0.125 | **0.072** | **0.083** |
| 4 | 30K | Neutral | NO | NO | 0.125 | **0.072** | **0.083** |
| **5** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.250 |
| **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.250 |
| 7 | 30K | Bad | NO | NO | 0.125 | **0.072** | **0.083** |
| 8 | 50k | Good | YES | YES | 0.125 | **0.072** | **0.083** |
| Total | | | | |  | 0.866 |  |

**Step06:** Normalize the weight by dividing it with “**”**. (formula given below)

For this scenario

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sr No | Salary | Credit | Approval | Predicted Value | Weight | Weight  After Calculation | Normalize  the  weight | Creating bin size |
| 1 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.083 | 0 - 0.08 |
| 2 | 50k | Good | YES | YES | 0.125 | 0.072 | 0.083 | 0.08 - 0.17 |
| 3 | 50k | Neutral | YES | YES | 0.125 | 0.072 | 0.083 | 0.17 - 0.25 |
| 4 | 30K | Neutral | NO | NO | 0.125 | 0.072 | 0.083 | 0.25 - 0.33 |
| **5** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.250 | 0.50 - 0.58 |
| **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.250 | 0.75 - 0.83 |
| 7 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.083 | 0.83 - 0.92 |
| 8 | 50k | Good | YES | YES | 0.125 | 0.072 | 0.083 | 0.92 - 1.00 |
| Total | | | | | 1.000 | 0.866 | 1.00 |  |

**Step07:** Create bins for each row. Note that a row with the maximum bin size will have the highest chance of being selected for the next stump.

**Repeat the above steps for the desired number of rotations or iterations to create the decision tree. For this case, we will repeat the same four more 4 times If you observe the data you will come to know that rows no 5 and 6 were wrongly predicted by the earlier stump hence in the below stump 5,6,8 got wrongly predicted again**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rotation Two**   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | exponential( e ) | | | 2.71828 |  |  |  |  | |  | Performance of Stump | | | 0.2554 |  |  |  |  | |  |  |  |  |  |  |  |  |  | | Sr No | Salary | Credit | Approval | Predicted Value | Weight | Weight After Calculation | Normalize the weith | Creating bin size | | 1 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.071 | 0 - 0.07 | | 2 | 50k | Good | YES | YES | 0.125 | 0.072 | 0.071 | 0.07 - 0.14 | | 3 | 50k | Neutral | YES | YES | 0.125 | 0.072 | 0.071 | 0.14 - 0.21 | | 4 | 30K | Neutral | NO | NO | 0.125 | 0.072 | 0.071 | 0.21 - 0.29 | | **5** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.214 | 0.43 - 0.50 | | **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.214 | 0.64 - 0.71 | | 7 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.071 | 0.71 - 0.79 | | **8** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.214 | 0.93 - 1.00 | | Total | | | | | 1.000 | 1.010 | 1.00 |  |   .. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rotation Two**   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | exponential(e ) | | | 2.71828 |  |  |  |  | |  | Performance of Stump | | | 0.5493 |  |  |  |  | |  |  |  |  |  |  |  |  |  | | Sr No | Salary | Credit | Approval | Predicted  Value | Weight | Weight After Calculation | Normalize the weith | Creating bin size | | 1 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.083 | 0 - 0.08 | | 2 | 50k | Good | YES | YES | 0.125 | 0.072 | 0.083 | 0.08 - 0.17 | | 3 | 50k | Neutral | YES | YES | 0.125 | 0.072 | 0.083 | 0.17 - 0.25 | | 4 | 30K | Neutral | NO | NO | 0.125 | 0.072 | 0.083 | 0.25 - 0.33 | | **5** | 50k | Good | YES | YES | 0.125 | 0.072 | 0.083 | 0.33 - 0.42 | | **6** | 50k | Neutral | YES | YES | 0.125 | 0.072 | 0.083 | 0.42 - 0.50 | | **5** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.250 | 0.67 - 0.75 | | **8** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.250 | 0.92 - 1.00 | | Total | | | | | 1.000 | 0.866 | 1.00 |  |   .. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rotation Three**   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | exponential(e ) | | | 2.71828 |  |  |  |  | |  | Performance of Stump | | | 0.9730 |  |  |  |  | |  |  |  |  |  |  |  |  |  | | Sr No | Salary | Credit | Approval | Predicted Value | Weight | Weight After Calculation | Normalize the weith | Creating bin size | | 1 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.100 | 0 - 0.10 | | 2 | 50k | Good | YES | YES | 0.125 | 0.072 | 0.100 | 0.10 - 0.20 | | 3 | 50k | Neutral | YES | YES | 0.125 | 0.072 | 0.100 | 0.20 - 0.30 | | 4 | 30K | Neutral | NO | NO | 0.125 | 0.072 | 0.100 | 0.30 - 0.40 | | **5** | 50k | Good | YES | YES | 0.125 | 0.072 | 0.100 | 0.40 - 0.50 | | **6** | 50k | Neutral | YES | YES | 0.125 | 0.072 | 0.100 | 0.50 - 0.60 | | **5** | 50k | Good | YES | YES | 0.125 | 0.072 | 0.100 | 0.60 - 0.70 | | **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.300 | 0.90 - 1.00 | | Total | | | | | 1.000 | 0.722 | 1.00 |  |   .. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rotation Four**   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | exponential(e ) | | | 2.71828 |  |  |  |  | |  | Performance of Stump | | | -0.2554 |  |  |  |  | |  |  |  |  |  |  |  |  |  | | Sr No | Salary | Credit | Approval | Predicted  Value | Weight | Weight After Calculation | Normalize the weith | Creating bin size | | 1 | 30K | Bad | NO | NO | 0.125 | 0.072 | 0.056 | 0 - 0.06 | | 2 | 50k | Good | YES | YES | 0.125 | 0.072 | 0.056 | 0.06 - 0.11 | | 3 | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.167 | 0.22 - 0.28 | | 4 | 30K | Neutral | NO | NO | 0.125 | 0.072 | 0.056 | 0.28 - 0.33 | | **5** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.167 | 0.44 - 0.50 | | **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.167 | 0.61 - 0.67 | | **5** | 50k | Good | YES | NO | 0.125 | 0.217 | 0.167 | 0.78 - 0.83 | | **6** | 50k | Neutral | YES | NO | 0.125 | 0.217 | 0.167 | 0.94 - 1.00 | | Total | | | | | 1.000 | 1.299 | 1.00 |  |   .. |

|  |
| --- |
| Finally, obtain the result by using the below formula. |