Nishant Sachdeva
2018111040

MDL Assignment 4 Report

Question Statement

Given table is as follows:

Wind(mph)	Angle(degree)	Spin(rotations)	Shot
90	35	7	Y
75	40	12	Y
93	48	3	Y
85	38	12	N
77	42	10	N
97	50	7	N
105	47	6	Y
110	48	10	N
90	55	7	N
85	42	11	N

GOALS

1. Create A Decision Tree using the Greedy Algorithm Paradigm.
2. List out the steps in complete detail in a pdf format

SPECIFICATIONS

1. We notice that in our table, there are three features that affect the final tree.
2. And we also see that all three of them are real valued . So, we'll follow the following method to find our final tree.
3. Start by sorting the values of the attribute, and then consider only split points that are between two examples in sorted order that have different classifications, while keeping track of the running totals of positive and negative examples on each side of the split point.

Our final working is as follows:

The formula for getting information gain is

$I(p,n) = -\{ \ p/(p+n) \log_2 (p/(p+n)) \ + n/(p+n) \log_2 (n/(p+n)) \ \}$

And finally for all these , we get the entropy

$$E(attribute) = \sum_{all\ subsets} \frac{no\_of\_elements\_in\_subset}{total\_number\_of\_elements} \times I(n,y)\ for\ subset$$

And finally our , information gain is calculated through ,

4. **Information gain**

$$Gain(attribute) = I(n,y)of\ set - E(attribute)$$

So, once we get this information gain for every attribute, we take the attribute that gives us the maximum information gain ! at every step
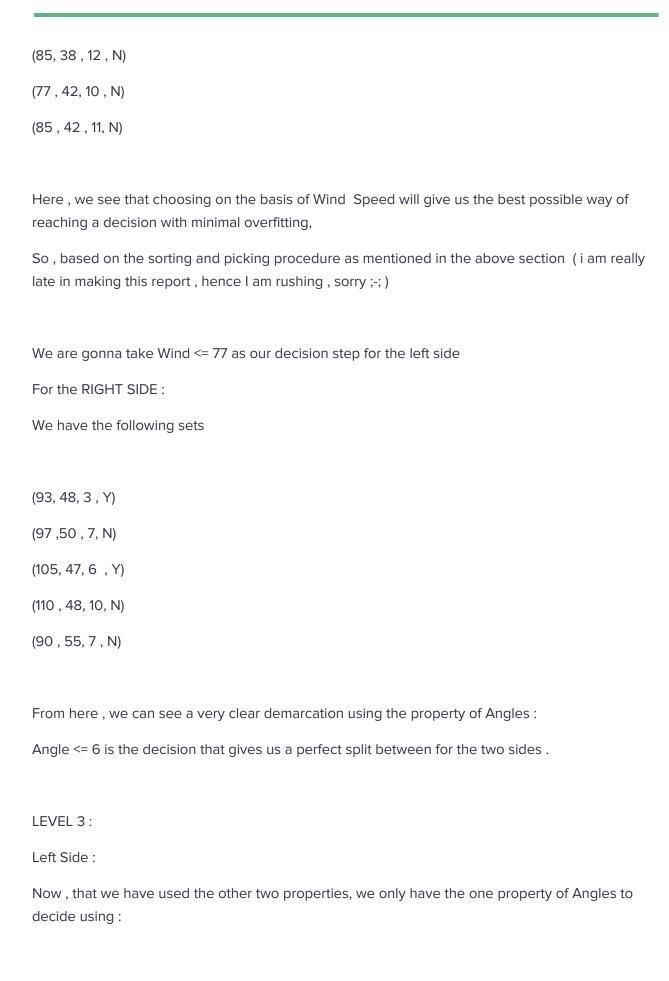
This procedure is then repeated for every single stage , until the tree draws to and end

Feature 1 :

Wind

In sorted Order :

75 : y, 77 : n , 85 : n , 85 : n, 90 : n , 90 : y, 93 : y , 97 : n , 105 : n , 110 : n

Possible Divisions at:

 90 = 1, 3 and 2, 4

 93 = 2, 4 and 3 ,1 :

Angle:

In sorted order

35 : y, 38: n , 40 y, 42 y, 42 n, 47 y, 48 y , 48 n , 50 n, 55 n

Possible Divisions at:

38, 42, 48

Spin:

 In sorted order :

3 y , 6 y, 7 n , 7 n , 7 y , 10 n , 10 n , 11 n , 12 n , 12 y

Possible Divisions at

 7 : 3 ,2 and 4 , 1

12 : 1 ,1 and 3, 5

Based on the calculations , For the root node, we are choosing Angle as the attribute for our thing.

STAGE 2:

Left :

We have

(90 , 35, 7 , Y)

(75 , 40 , 12 , Y)

(85, 38 , 12 , N)

(77 , 42, 10 , N)

(85 , 42 , 11, N)

Here , we see that choosing on the basis of Wind Speed will give us the best possible way of reaching a decision with minimal overfitting,

So , based on the sorting and picking procedure as mentioned in the above section (i am really late in making this report , hence I am rushing , sorry ;-;)

We are gonna take Wind <= 77 as our decision step for the left side

For the RIGHT SIDE :

We have the following sets

(93, 48, 3 , Y)

(97 ,50 , 7, N)

(105, 47, 6 , Y)

(110 , 48, 10, N)

(90 , 55, 7 , N)

From here , we can see a very clear demarcation using the property of Angles :

Angle <= 6 is the decision that gives us a perfect split between for the two sides .

LEVEL 3 :

Left Side :

Now , that we have used the other two properties, we only have the one property of Angles to decide using :

Left Side :

- We have the following sets :

  - (75, 40 , 12 , Y)

  - (77 , 42, 10 , N)

Here we can very clearly make the demarcation between the Y and the N .

We will take Spin < 11 as the demarcation parameter. This will give us two very clear sides and thus we will reach the final decision.

Right Side :

- We have the following sets :
  - (93, 48 , 3 , Y)
  - (97, 50, 7 , N)
  - (105, 47, 6 , Y)
  - (110, 48, 10, N)
  - (90, 55 , 7 , N)

Here we can very clearly observe the demarcation created by the presence of the spin feature. Here, we notice that we do not need to spend a level in the tree dealing with the Wind Speed feature. Hence we will simply end the demarcation with this level with the separation parameter being [Spin < 7] . However, Please note that , here, we will definitely have options for the possibilities for other trees too.

MILESTONES

The Decision Tree generated is as follows :