prajith
2320030484

Illustrate Naive Bayes on the dataset to predict whether we can pet an animal or not. Find $P(x_i \mid y_j)$ for each x_i in X each y_j in Y. All these calculations must be demonstrated.
Illustrate decision tree on the dataset to predict whether we can pet an animal or not and all the entropy calculations must be demonstrated in the assignment.

S NO	Animals	Size of Animal	Body Color	Can we pet
0	Dog	Medium	Black	Yes
1	Dog	Big	White	No
2	Rat	Small	White	Yes
3	Cow	Big	White	Yes
4	Cow	Small	~~Black~~ Brown	No
5	Cow	Big	Black	Yes
6	Rat	Big	Brown	No
7	Dog	Small	Brown	Yes
8	Dog	Medium	Brown	Yes
9	~~Dog~~ Cow	Medium	White	No
10	Dog	Small	Black	Yes
11	Rat	Medium	Black	No
12	Rat	Small	Brown	No
13	Cow	Big	White	Yes

Naive Bayes Prediction

Total example 14

Count of yes 8 (P(yes))
Count of no 6 (P(no))

$P(\text{yes}) = \frac{8}{14} = \frac{4}{7} = 0.57$

$P(\text{no}) = \frac{6}{14} = \frac{3}{7} = 0.43$

Decision Tree classification

Step 1. Calculate overall entropy

Total examples = 14 (8 Yes , 6 No)

Entropy (S) = $-\left(\frac{8}{14}\log_2\frac{8}{14}+\frac{6}{14}\log_2\frac{6}{14}\right)$

≈ 0.98

Step 2: For Size of animal.

Small (5 total) - 3 Yes 2 No

$= -\left(\frac{3}{5}\log_2\frac{3}{5}+\frac{2}{5}\log_2\frac{2}{5}\right)$

≈ 0.970

For Medium (4 total) - (3 Yes 1 No)

$= -\left(\frac{3}{4}\log_2\frac{3}{4}+\frac{1}{4}\log_2\frac{1}{4}\right)$

≈ 0.81

For Big (5 total) - (2 Yes 3 No)

Entropy (S_{Big}) $= -\left(\frac{2}{5}\log_2\frac{2}{5}+\frac{3}{5}\log_2\frac{3}{5}\right)$

$= -(0.4 \times -1.322 + 0.6 \times -0.736)$

≈ 0.970

Calculate Weighted Avg. entropy for Size of Animal

Weighted entropy (Size) $= \frac{5}{14} \times$ Entropy(S_{small}) $+ \frac{4}{14} \times$ Entropy(S_{med})

$+ \frac{5}{14}$ Entropy(S_{Big})

• Calculation process for the Body colour feature:

Decision Tree classification for Body colour:

Step 1) Calculate Entropy for Body colour

Assuming the dataset distribution for Body colours:

- Black: 2 yes, 2 No (Total: 4)
- White: 3 yes, 1 No (Total: 4)
- Brown: 3 yes, 3 No (Total: 6)

1) Black:

$$\text{Entropy}(S_{Black}) = -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right)$$

$$= -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1.0$$

2) White:

$$\text{Entropy}(S_{White}) = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = 0.81$$

3) Brown:

$$\text{Entropy}(S_{Brown}) = -\left(\frac{3}{6}\log_2\frac{3}{6} + \frac{3}{6}\log_2\frac{3}{6}\right) = 1.0$$

Step 2) Calculate Weighted Entropy for Body colour

The total counts for each Body colour:

1) Total Black: 4
2) Total White: 4
3) Total Brown: 6

Compute the Weighted Entropy:

$$\text{Weighted Entropy (colour)} : \frac{4}{14}\times 1.0 + \frac{4}{14} + \frac{4}{14}\times 0.81 + \frac{6}{14}\times 1.0$$

$= \frac{4}{14} + \frac{3.24}{14} + \frac{6}{14}$

$= \frac{13.24}{14} \approx 0.95$

step 3) Calculate information Gain for Body colour using the overall entropy calculated previously (≈ 0.95)

Gain (s, Colour) = Entropy (s) - weighted Entropy (colour)

$= 0.98 - 0.95 \approx 0.03$

Final Step: Determine Best split

- Size of Animal = information Gain ≈ 0.06
- Body Colour = Information Gain ≈ 0.03

Best Split:

Since the feature with the highest information gain is size of Animal (0.06). this will be selected as the first split in the decision tree.

For Yes

Size of animal:

Small = 3
Medium = 3
Big = 2

Body color:

Black = 2
White = 3
Brown = 3

For No

Size of animal:

Small = 2
Medium = 1
Big = 3

Body Color:

Black = 2
White = 1
Brown = 3

- For Size of animal

$P(\text{Small} \mid \text{Yes}) = \frac{3}{8}$

$P(\text{Medium} \mid \text{Yes}) = \frac{3}{8}$

$P(\text{Big} \mid \text{Yes}) = \frac{2}{8}$

$P(\text{Small} \mid \text{No}) = \frac{2}{6}$

$P(\text{Medium} \mid \text{No}) = \frac{1}{6}$

$P(\text{Big} \mid \text{No}) = \frac{3}{6}$

- For Body color

$P(\text{Black} \mid \text{Yes}) = \frac{2}{8}$

$P(\text{White} \mid \text{Yes}) = \frac{3}{8}$

$P(\text{Brown} \mid \text{Yes}) = \frac{3}{8}$

$P(\text{Black} \mid \text{No}) = \frac{2}{6}$

$P(\text{White} \mid \text{No}) = \frac{1}{6}$

$P(\text{Brown} \mid \text{No}) = \frac{3}{6}$

For a new animal describes as Big & White, calculate the posterior probabilities

$P(\text{Yes} \mid \text{Big, White}) = P(\text{Big} \mid \text{Yes}) \times P(\text{White} \mid \text{Yes}) \times P(\text{Yes})$

$P(\text{No} \mid \text{Big, White}) = P(\text{Big} \mid \text{No}) \times P(\text{White} \mid \text{No}) \times P(\text{No})$

$P(\text{Yes} \mid \text{Big, White}) = \left(\frac{2}{8}\right) \times \left(\frac{3}{8}\right) \times \left(\frac{4}{7}\right)$

$P(\text{No} \mid \text{Big, White}) = \left(\frac{3}{6}\right) \times \left(\frac{1}{6}\right) \times \left(\frac{3}{7}\right)$

$= \frac{5}{14} \times 0.970 + \frac{4}{14} \times 0.811 + \frac{5}{14} \times 0.970$

$= 0.346 + 0.231 + 0.346$

$= 0.923$

Step 4. Calculate Information gain for Size of Animal

Gain (S, Size) = Entropy (S) - Weighted Entropy (Size)

$= 0.98 - 0.923$

$= 0.057$