SMAI - ASSI-3 Decision trees and random forests 20191010 46

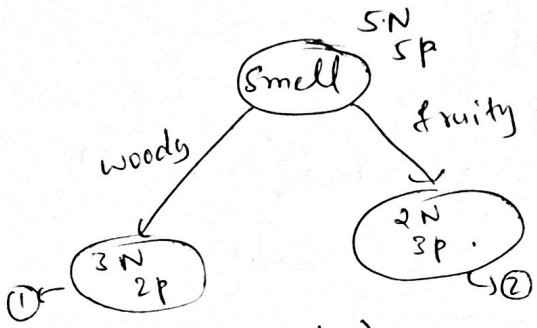
N.V. Sneeshlal
Reddy

(5)

(a) Now we have to compute the entropy of each rule. in first stage.

(i) we will classify with Smell with woody and not woody
↳ {fruity}
↳ from table.

N → Negatives
p → positives



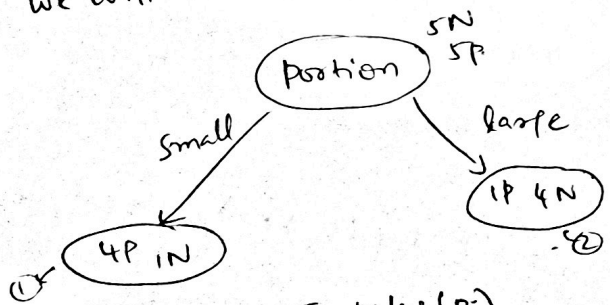
Entropy = $\Sigma - p_i \log(p_i)$

$E_1 = -\frac{2}{5}\log(2/5) - \frac{3}{5}\log(3/5) = 0.9674$

$E_2 = -\frac{2}{5}\log(2/5) - 3/5\log(3/5) = 0.9674$

$E = \frac{1}{2}E_1 + \frac{1}{2}E_2 = 0.9674$

(ii) we will classify with portion with small and not small
↳ {large}
↳ from table



Entropy = $\Sigma - p_i \log(p_i)$

$E_1 = -\frac{1}{5}\log(1/5) - \frac{4}{5}\log(\frac{4}{5}) = 0.72192$

$E_2 = -\frac{1}{5}\log(1/5) - \frac{4}{5}\log(\frac{4}{5}) = 0.72192$

$E = \frac{1}{2}E_1 + \frac{1}{2}E_2 = 0.72192.$

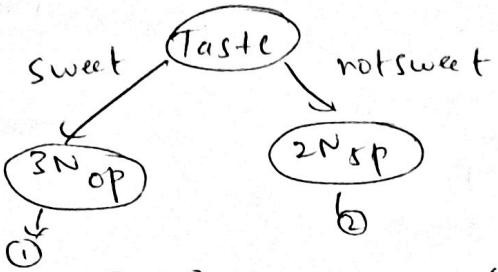
(iii) We will classify with Taste.

Now we will get 3 cases

sour, not sour

sweet, not sweet

salty, not salty.

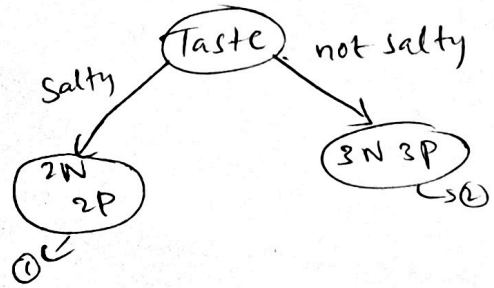


$E_1 = 0$

$E_2 = -\frac{2}{7}\log\left(\frac{2}{7}\right) - \frac{5}{7}\log\left(\frac{5}{7}\right) = 0.86311$

$E = \frac{7}{10} \times (0.86311) = 0.604177$

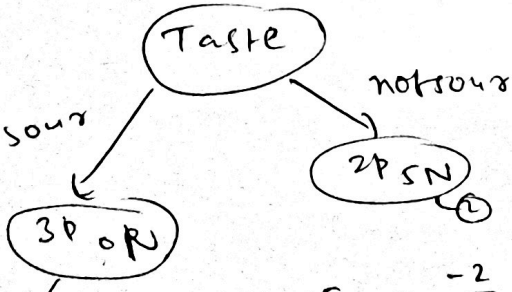
$\Rightarrow$ $E = 0.604177$



$E_1 = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$

$E_2 = -\frac{1}{2}\log(1/2) - \frac{1}{2}\log(1/2) = 1$

$\Rightarrow$ $E = 1$

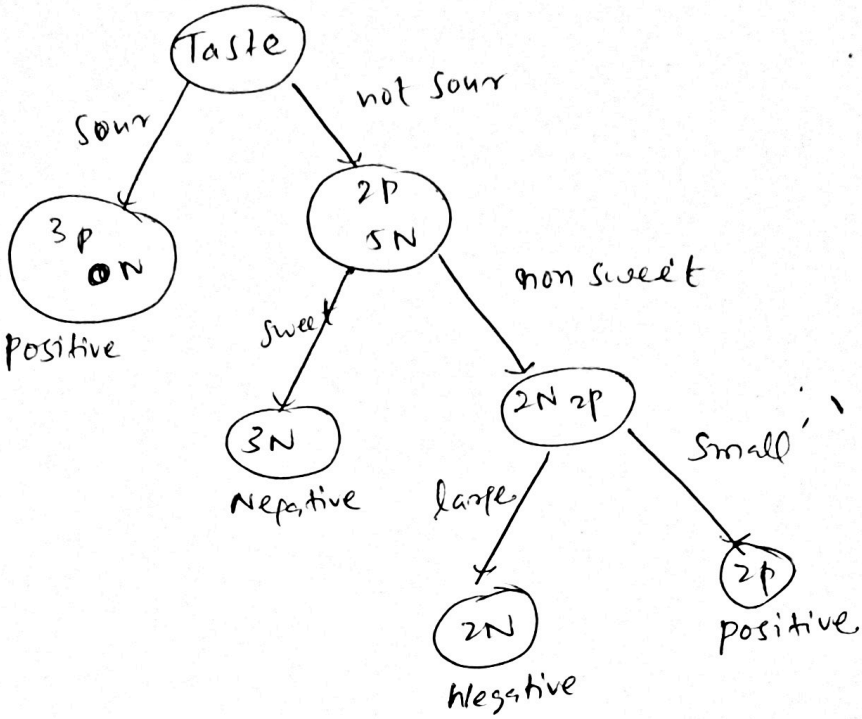


$E_1 = 0$, $E_2 = -\frac{2}{7}\log(2/7) - 5/7\log(5/7) = 0.86311$

$E = \frac{7}{10} \times (0.86311) = 0.604177$

$\Rightarrow$ $E = 0.604177$

(b) From above we can see that the entropy is less for classification of taste with sour and not sour.



So, the next split is sweet and non sweet and the next split will be large and small as we can see each split gets a pure node. So, this is best splits of the tree.