

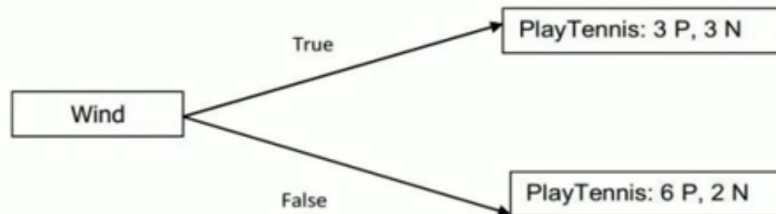
- If a data set  $D$  contains examples from  $n$  classes, gini index,  $gini(D)$  is defined as:

$$gini(D) = 1 - \sum_{j=1}^n p_j^2$$

### Gini index calculation:

There are 5 Ns and 9 Ps, so the

- Calculate the information gain after the Wind test is applied:



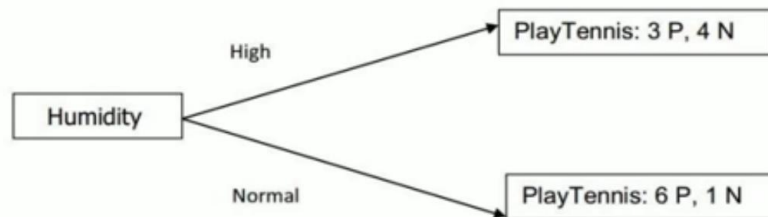
$$Gini(PlayTennis|Wind=True) = 1 - (3/6)^2 - (3/6)^2 = 0.5$$

$$Gini(PlayTennis|Wind=False) = 1 - (6/8)^2 - (2/8)^2 = 0.375$$

Therefore, the Gini index after the Wind test is applied is

$$6/14 \times 0.5 + 8/14 \times 0.375 = 0.4286$$

- Calculate the information gain after the Humidity test is applied:



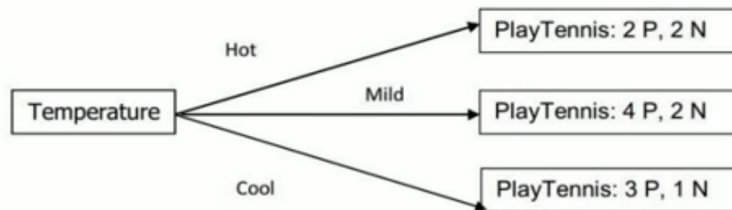
$$Gini(PlayTennis|Humidity=High) = 1 - (3/7)^2 - (4/7)^2 = 0.4898$$

$$Gini(PlayTennis|Humidity=Normal) = 1 - (6/7)^2 - (1/7)^2 = 0.2449$$

Therefore, the Gini index after the Humidity test is applied is

$$7/14 \times 0.4898 + 7/14 \times 0.2449 = 0.3674$$

- Calculate the information gain after the Temperature test is applied:



$$\text{Gini (PlayTennis | Temperature = Hot)} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$$

$$\text{Gini (PlayTennis | Temperature = Mild)} = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = 0.4444$$

$$\text{Gini (PlayTennis | Temperature = Cool)} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$$

Therefore, the Gini index after the Temperature test is applied is

$$\frac{4}{14} \times 0.5 + \frac{6}{14} \times 0.4444 + \frac{4}{14} \times 0.375 = 0.4405$$