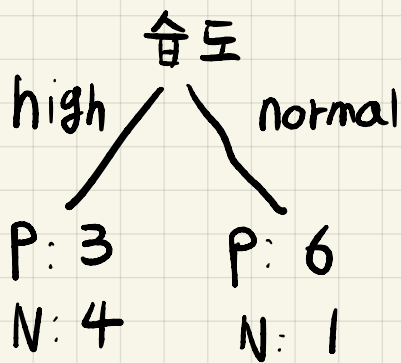


1. (1) k 개의 attributes를 기준으로 최대 한번만 나눌 수 있기 때문에 최대 depth는 k 이다.
 이에 따라서 최대 leaf는 2^k 이다. 다만 2^k 보다 N 이 작다면 최대 leaf는 모든 관찰값이 분리된 N 이다.
 즉 최대 leaf수는 $\min\{2^k, N\}$ 이다.

- (2) 모든 attributes가 연속인 경우 모든 관찰값이 구분될 수 있으므로 최대 leaf는 N 이다.

2. (1) 습도로 구분한 경우

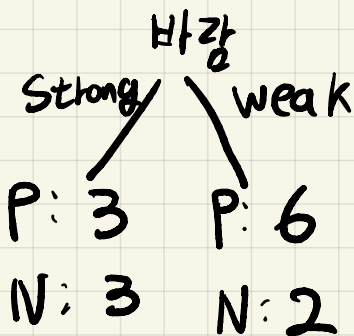


$$\text{high entropy} = -\left(\frac{3}{7} \log_2 \frac{3}{7} + \frac{4}{7} \log_2 \frac{4}{7}\right) = 0.985$$

$$\text{normal entropy} = -\left(\frac{6}{7} \log_2 \frac{6}{7} + \frac{1}{7} \log_2 \frac{1}{7}\right) = 0.591$$

$$\text{습도 entropy} = \frac{7}{14} \times 0.985 + \frac{7}{14} \times 0.591 = 0.788$$

바람으로 구분한 경우



$$\text{strong entropy} = -\left(\frac{3}{6} \log_2 \frac{3}{6} + \frac{3}{6} \log_2 \frac{3}{6}\right) = 1$$

$$\text{weak entropy} = -\left(\frac{6}{8} \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8}\right) = 0.811$$

$$\text{바람 entropy} = \frac{6}{14} \times 1 + \frac{8}{14} \times 0.811 = 0.892$$

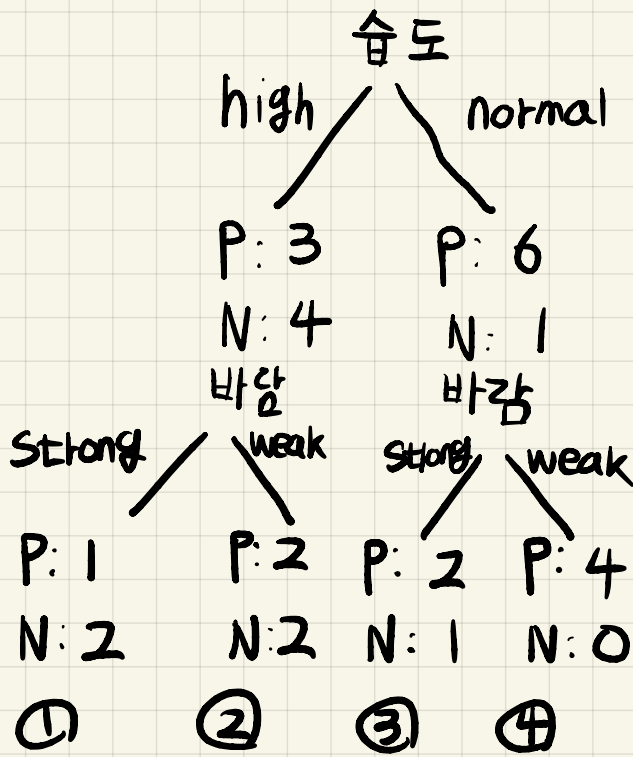
$$\text{원래 entropy} = -\left(\frac{9}{14} \log_2 \frac{9}{14} + \frac{5}{14} \log_2 \frac{5}{14}\right) = 0.94$$

$$\text{습도로 분기시 entropy gain} = 0.94 - 0.788 = 0.152$$

$$\text{바람으로 분기시 entropy gain} = 0.94 - 0.892 = 0.048$$

entropy gain이 큰 습도를 root에서 분기시 사용한다.

(2)



각각의 entropy

$$① -\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right) = 0.918$$

$$② -\left(\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\frac{2}{4}\right) = 1$$

$$③ -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right) = 0.918$$

$$④ -\left(\frac{4}{4}\log_2\frac{4}{4}\right) = 0$$

총 information entropy = $\frac{3}{14} \times 0.918 + \frac{4}{14} \times 1 + \frac{3}{14} \times 0.918 + \frac{4}{14} \times 0 = 0.619$

두번째 분계를 통한 information gain = $0.988 - 0.619 = 0.109$ 이다.

3. λ 값이 커질수록 하나의 데이터가 영향을 미치는 범위가

감소하게 되고 서포트 벡터의 수가 증가하여

decision boundary가 더 많은 서포트 벡터에

의해서 결정지게 된다. 데이터 하나하나에 의해

boundary가 더 많이 바뀔 수 있으므로 overfitting
의 가능성이 커지게 된다.