

Decision Tree by using gini function:

 step1:calculate gini index of dataset
 step2:calculate gini index of each feature
 step3:calculate gini gain of each feature
 step4:pick the highest gain feature as root
 step5:repeat this for further subsets to complete tree

Formula of gini index:

$$1 - \sum \text{of } \text{sqr}(P(t))$$

step1:calculate gini index of dataset

$$p(\text{yes}) = 9/14 = 0.64$$

$$p(\text{no}) = 5/14 = 0.35$$

$$\text{sqr}(p(\text{yes})) = .413$$

$$\text{sqr}(p(\text{no})) = .127$$

$$\text{sum of } \text{sqr}(p(\text{yes}), p(\text{no})) = .413 + .127 = .54$$

$$\begin{aligned} \text{gini index of dataset} &= \\ &1 - \text{sum of } \text{sqr}(p(\text{yes}), p(\text{no})) \\ &1 - .54 \\ &.46 \end{aligned}$$

step2:

2.1:calculate gini index of outlook:

$$\begin{aligned} \text{Gini}(\text{outlook}, \text{rainy}) &= 1 - [(2/5)^2 + (3/5)^2] \\ &= 1 - [.16 + .36] = 1 - .52 = .48 \end{aligned}$$

$$\begin{aligned} \text{Gini}(\text{outlook}, \text{overcast}) &= 1 - [(4/4)^2] \\ &= 1 - 1 = .0 \end{aligned}$$

$$\begin{aligned} \text{Gini}(\text{outlook}, \text{sunny}) &= 1 - [(3/5)^2 + (2/5)^2] \\ &= 1 - [.36 + .16] = 1 - .52 = .48 \end{aligned}$$

Gini index of

$$\text{Outlook} = G(o, \text{rainy}) * P(\text{rainy}) + G(o, \text{overcast}) * P(\text{overcast}) + G(o, \text{sunny}) * P(\text{sunny})$$

$$= .48 * (5/14) + 0 + .48 * (5/14)$$

$$= .3428$$

2.2:calculate gini index of temp:

$$\text{Gini}(\text{temp}, \text{hot}) = 1 - [(2/4)^2 + (2/4)^2]$$



Decision Tree by using gini function:

 step1:calculate gini index of dataset
 step2:calculate gini index of each feature
 step3:calculate gini gain of each feature
 step4:pick the highest gain feature as root
 step5:repeat this for further subsets to complete tree

Formula of gini index:

$$1 - \text{Sum of } \text{sqr}(P(t))$$

step1:calculate gini index of dataset

$$p(\text{yes}) = 9/14 = 0.64$$

$$p(\text{no}) = 5/14 = 0.35$$

$$\text{sqr}(p(\text{yes})) = .413$$

$$\text{sqr}(p(\text{no})) = .127$$

$$\text{sum of } \text{sqr}(p(\text{yes}), p(\text{no})) = .413 + .127 = .54$$

$$\begin{aligned} \text{gini index of dataset} &= \\ &1 - \text{sum of } \text{sqr}(p(\text{yes}), p(\text{no})) \\ &1 - .54 \\ &.46 \end{aligned}$$

step2:

2.1:calculate gini index of outlook:

$$\begin{aligned} \text{Gini}(\text{outlook}, \text{rainy}) &= 1 - [(2/5)^2 + (3/5)^2] \\ &= 1 - [.16 + .36] = 1 - .52 = .48 \end{aligned}$$

$$\begin{aligned} \text{Gini}(\text{outlook}, \text{overcast}) &= 1 - [(4/4)^2] \\ &= 1 - 1 = .0 \end{aligned}$$

$$\begin{aligned} \text{Gini}(\text{outlook}, \text{sunny}) &= 1 - [(3/5)^2 + (2/5)^2] \\ &= 1 - [.36 + .16] = 1 - .52 = .48 \end{aligned}$$

Gini index of

$$\text{Outlook} = G(o, \text{rainy}) * P(\text{rainy}) + G(o, \text{overcast}) * P(\text{overcast}) + G(o, \text{sunny}) * P(\text{sunny})$$

$$= .48 * (5/14) + 0 + .48 * (5/14)$$

$$= .3428$$

2.2:calculate gini index of temp:

$$\text{Gini}(\text{temp}, \text{hot}) = 1 - [(2/4)^2 + (2/4)^2]$$

