Homework assignment in Week 7

Q1: For a classification problem with 4 classes (A, B, C,D), calculate the Entropy of a set which has 4 instances of A class, 2 instances of B class, 4 instances of C class, and 3 instances of D class.

The calculation process must be included in the answer.

$$H(x) = -\sum_i^c p_i \log_2 p_i = \sum_{i=1}^c p_i \log_2 \frac{1}{p_i}$$

$$S_{total} = 4 + 2 + 4 + 3 = 13$$

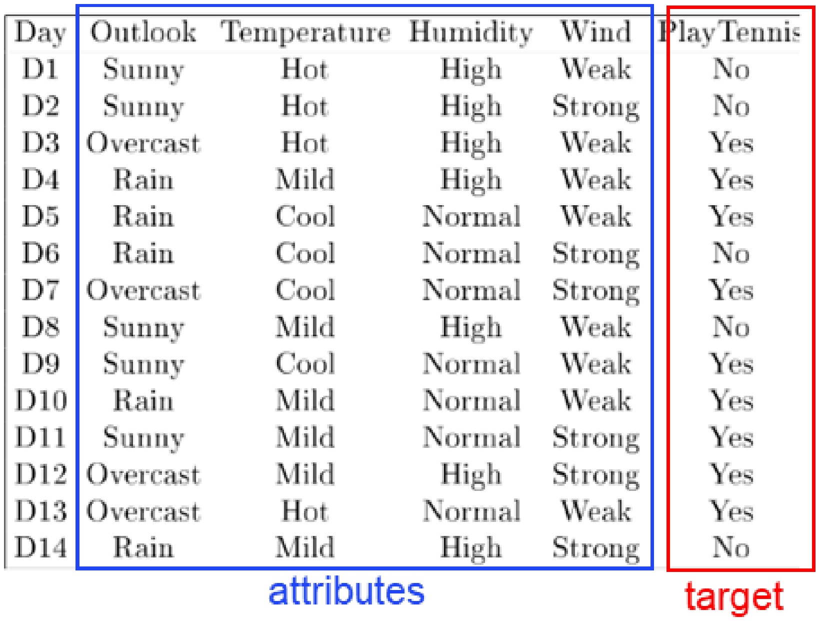
$$entropy(S) = -\sum_i^c p_i \log_2 p_i = -\frac{4}{13} \times \log_2 \frac{4}{13} - \frac{2}{13} \times \log_2 \frac{2}{13} - \frac{4}{13} \times \log_2 \frac{4}{13} - \frac{3}{13} \times \log_2 \frac{3}{13}$$

$$= 0.5232 + 0.4155 + 0.5232 + 0.4882$$

$$= 1.9501$$

Q2: Imagine you play tennis, and you invite your friend. Your friend sometimes comes to join but sometimes not. For your friend, it depends on a number of factors, for example, weather, temperature, humidity, and wind. Please use the right dataset to build a decision tree which can predict whether or not your friend will join you to play tennis.

You must list the calculation process (to build the decision tree) and plot the decision tree.



Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$entropy(total) = -\sum_i^c p_i \log_2 p_i = -\frac{9}{14} \times \log_2 \frac{9}{14} - \frac{5}{14} \times \log_2 \frac{5}{14} = 0.9403$

Outlook:

Sunny= $\frac{5}{14}$

Overcast = $\frac{4}{14}$

Rain = $\frac{5}{14}$

$entropy(sunny) = -\frac{3}{5} \times \log_2 \frac{3}{5} - \frac{2}{5} \times \log_2 \frac{2}{5} = 0.9709$

$entropy(overcast) = -\frac{4}{4} \times \log_2 \frac{4}{4} - \frac{0}{4} \times \log_2 \frac{0}{4} = 0$

$entropy(rain) = -\frac{3}{5} \times \log_2 \frac{3}{5} - \frac{2}{5} \times \log_2 \frac{2}{5} = 0.9709$

$\frac{5}{14} \times 0.9709 = 0.34675$

$\frac{4}{14} \times 0 = 0$

$\frac{5}{14} \times 0.9709 = 0.34675$

$IG(T, outlook) = entropy(total) - entropy(outlook) = 0.9403 - 0.69350 = 0.2468$

Temperature:

Cool = $\frac{4}{14}$

Hot = $\frac{4}{14}$

Mild = $\frac{7}{14}$

$entropy(cool) = -\frac{3}{4} \times \log_2 \frac{3}{4} - \frac{1}{4} \times \log_2 \frac{1}{4} = 0.8113$

$entropy(hot) = -\frac{2}{4} \times \log_2 \frac{2}{4} - \frac{2}{4} \times \log_2 \frac{2}{4} = 1$

$entropy(cool) = -\frac{5}{7} \times \log_2 \frac{5}{7} - \frac{2}{7} \times \log_2 \frac{2}{7} = 0.8631$

$\frac{4}{14} \times 0.8113 = 0.2318$

$\frac{4}{14} \times 1 = \frac{4}{14}$

$\frac{7}{14} \times 0.8631 = 0.43155$

$IG(T, temp) = entropy(total) - entropy(temp) = 0.9403 - 0.9491 = -0.00786$

Humidity:

Normal = $\frac{7}{14}$

High = $\frac{7}{14}$

$entropy(high) = -\frac{3}{7} \times \log_2 \frac{3}{7} - \frac{4}{7} \times \log_2 \frac{4}{7} = 0.9852$

$entropy(normal) = -\frac{6}{7} \times \log_2 \frac{6}{7} - \frac{1}{7} \times \log_2 \frac{1}{7} = 0.5917$

$\frac{7}{14} \times 0.9852 = 0.4926$

$\frac{7}{14} \times 0.5917 = 0.29585$

$IG(T, temp) = entropy(total) - entropy(humidity) = 0.9403 - 0.78845 = 0.15185$

Wind:

Weak = $\frac{8}{14}$

Strong = $\frac{6}{14}$

$entropy(weak) = -\frac{2}{8} \times \log_2 \frac{2}{8} - \frac{6}{8} \times \log_2 \frac{6}{8} = 0.8113$

$entropy(strong) = -\frac{3}{6} \times \log_2 \frac{3}{6} - \frac{3}{6} \times \log_2 \frac{3}{6} = 1$

$\frac{8}{14} \times 0.8113 = 0.4636$

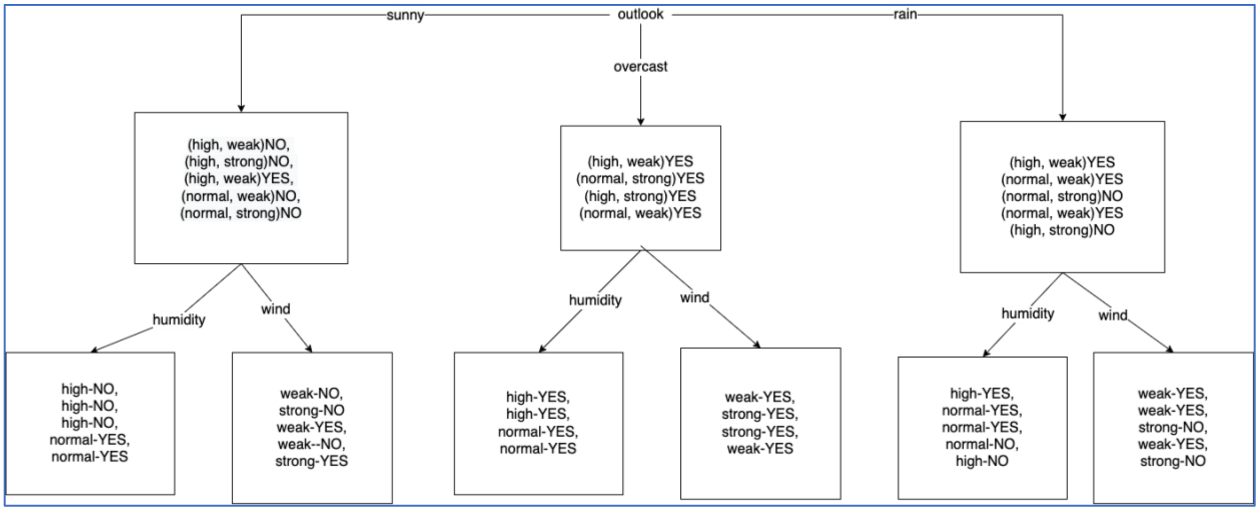
$\frac{6}{14} \times 1 = \frac{6}{14}$

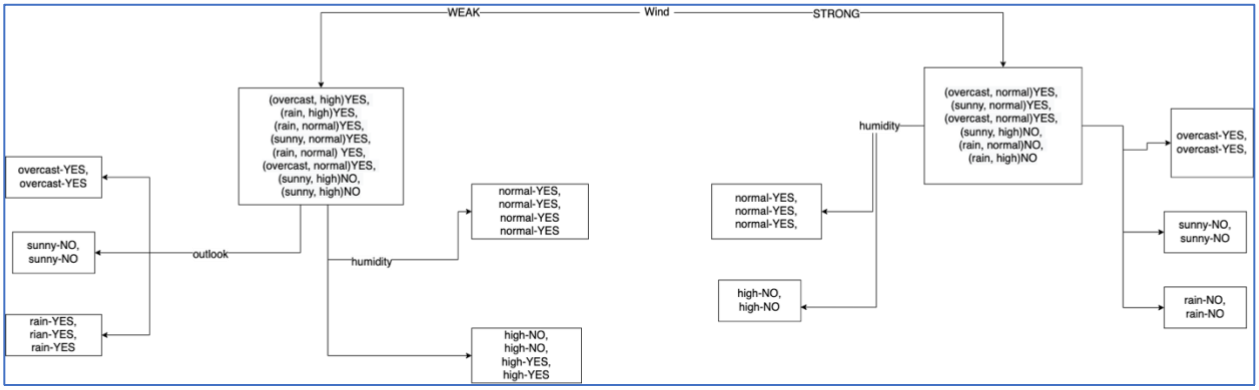
$IG(T, wind) = entropy(total) - entropy(wind) = 0.9403 - 0.8922 = 0.04813$

IG of outlook, humidity and wind are positive, IG of temperature is negative.

Outlook, humidity and wind will be used in the decision tree.

Decision tree


outlook
sunny
overcast
rain
(high, weak)NO,
(high, strong)NO,
(high, weak)YES,
(normal, weak)NO,
(normal, strong)NO
(high, weak)YES
(normal, strong)YES
(high, strong)YES
(normal, weak)YES
(high, weak)YES
(normal, weak)YES
(normal, strong)NO
(normal, weak)YES
(high, strong)NO
humidity
wind
humidity
wind
humidity
wind
high-NO,
high-NO,
high-NO,
normal-YES,
normal-YES
weak-NO,
strong-NO
weak-YES,
weak--NO,
strong-YES
high-YES,
high-YES,
normal-YES,
normal-YES
weak-YES,
strong-YES,
strong-YES,
weak-YES
high-YES,
normal-YES,
normal-YES,
normal-NO,
high-NO
weak-YES,
weak-YES,
strong-NO,
weak-YES,
strong-NO


WEAK
Wind
STRONG
(overcast, high)YES,
(rain, high)YES,
(rain, normal)YES,
(sunny, normal)YES,
(rain, normal) YES,
(overcast, normal)YES,
(sunny, high)NO,
(sunny, high)NO
overcast-YES,
overcast-YES
sunny-NO,
sunny-NO
rain-YES,
rian-YES,
rain-YES
outlook
humidity
normal-YES,
normal-YES,
normal-YES
normal-YES
high-NO,
high-NO,
high-YES,
high-YES
humidity
(overcast, normal)YES,
(sunny, normal)YES,
(overcast, normal)YES,
(sunny, high)NO,
(rain, normal)NO,
(rain, high)NO
normal-YES,
normal-YES,
normal-YES,
high-NO,
high-NO
overcast-YES,
overcast-YES,
sunny-NO,
sunny-NO
rain-NO,
rain-NO