CSCE 421: Machine Learning

HW1

1 [10 pts]

P(B = r) = 0.2, P(B = b) = 0.2, P(B = g) = 0.6.

P(F = a) = P(F = a|B = r)P(B = r) + P(F = a|B = b)P(B = b) + P(F = a|B = g)P(B = g)

= 3/10*0.2 + 1/2*0.2 + 3/10*0.6 = 0.06 + 0.1 + 0.18 = 0.34

P(B = g | F = o) = P(F = o | B = g) * P(B = g) / P(F = o) = 3/10 * 0.6 / P(F = o)

P(F = o) = P(F = o|B = r)P(B = r) + P(F = o|B = b)P(B = b) + P(F = o|B = g)P(B = g)

= 4/10*0.2 + 1/2*0.2 + 3/10*0.6 = 0.08 + 0.1 + 0.18 = 0.36

So we can get P(B = g | F = o) = 0.18/0.36 = 0.5

2 [20 pts]

P(red | yes) = 3/5 P(red | no) = 2/5

P(Domestic | yes) = 2/5 P(Domestic | no) = 3/5

P(SUV | yes) = 1/5 P(SUV | no) = 3/5

P(yes | Red Domestic SUV) = P(red|yes)*P(Domestic|yes)*P(SUV|yes)*P(yes) / P(red Domestic SUV)

= 3/5* 2/5 * 1/5 * 0.5 / x = 3/125 /x

P(no | Red Domestic SUV) = P(red|no)*P(Domestic|no)*P(SUV|no)*P(no) / P(red Domestic SUV)

= 2/5* 3/5 * 3/5 * 0.5 / x = 9/125 /x

So P(yes | Red Domestic SUV) < P(no | Red Domestic SUV)

3 [20 pts]

(a) The attributes are conditionally independent of each other given the class label.
(b) If we use the simplifying assumption, then we only need 2*d*(k-1) parameters.

If we don't use the assumption, we need to calculate (k^d -1)*2 parameters, which is much more than the first one.

4 [25 pts]

From the Bayes' theorem P(Y|X) ∝ P(X|Y)P(Y)

For the texts A,

P(Physics | A) ∝ P(carbon | Physics) * P(atom | Physics) * P(life | Physics) * P(earth | Physics) * P(Physics)

= 0.005 * 0.1 * 0.001 * 0.005 * 0.35 = 8.75e-10

P(Biology | A) ∝ P(carbon | Biology) * P(atom| Biology) * P(life | Biology) * P(earth |Biology) * P(Biology)

= 0.03 * 0.01 * 0.1 * 0.006 * 0.4 = 7.2e-08

P(Chemistry | A) ∝ P(carbon | Chemistry) * P(atom| Chemistry) * P(life | Chemistry) * P(earth | Chemistry) * P(Chemistry)

= 0.05 * 0.2 * 0.008 * 0.003 * 0.25 = 6.0e-08

So we classify texts A into Biology.

For the texts B,

P(Physics | B) ∝ P(carbon | Physics) * P(atom | Physics) * P(protons | Physics) * P(Physics)

= 0.005*0.1*0.05*0.35 = 8.75e-06

P(Biology | B) ∝ P(carbon | Biology) * P(atom | Biology) * P(protons | Biology) * P(Biology)

= 0.01*0.03*0.001*0.4 = 1.2e-07

P(Chemistry | B) ∝ P(carbon |Chemistry) * P(atom |Chemistry) * P(protons |Chemistry) * P(Chemistry)

= 0.2*0.05*0.05*0.25 = 1.25e-04

So we classify texts B into Chemistry.

5 [25 pts]

P(play = Y) = 9/14

P(play = N) = 5/14

Entropy = -9/14*log2(9/14) – (5/14)*log2(5/14) (all the log is based on 2)

= 0.9403

Entropy(T = hot) = -2/4*log2(2/4) – 2/4*log2(2/4) = 1

Entropy(T = mild) = -4/6*log2(4/6) – 2/6*log2(2/6) = 0.9183

Entropy(T = cool) = -3/4*log2(3/4) – 1/4 *log2(1/4) = 0.8113

IG(Temp) = 0.9403-(4/14*1 + 6/14*0.9183 + 4/14 * 0.8113) =0.0292

Entropy(outlook = sunny) = -2/5*log2(2/5) – 3/5 *log2(3/5) = 0.9710

Entropy(outlook = overcast) = 0

Entropy(outlook = rain) = -2/5*log2(2/5) – 3/5 *log2(3/5) = 0.9710

IG(Outlook) = 0.9403 – (5/14 * 0.9710 + 0 +5/14 * 0.9710)=0.2467 (In the 2nd tree, we choose outlook as root node)

Entropy(Humidity = high) = -3/7*log2(3/7) – 4/7*log2(4/7) = 0.9852

Entropy(Humidity = normal) = -6/7*log2(6/7) – 1/7*log2(1/7) = 0.5917

IG(Humidity) = 0.9403-(7/14*0.9852 + 7/14*0.5917) =0.1519

Entropy(Windy = true) = -3/6*log2(3/6) -3/6*log2(3/6) = 1

Entropy(Windy = false) = -6/8*log2(6/8) – 2/8*log2(2/8) = 0.8113

IG(Windy) = 0.9403 – (6/14 * 1 + 8/14 * 0.8113) =0.0481

For the tree 1:

Entropy(Outlook = sunny | T = hot) = 0

Entropy(Outlook = overcast | T = hot) = 0

IG(outlook, T = hot) = 1

IG(humidity, T = hot) and IG(windy, T = hot) are both smaller than IG(outlook, T = hot), we choose outlook under branch hot.

Entropy(Outlook = sunny| T = mild) = 1

Entropy(Outlook = overcast| T = mild) = 0

Entropy(Outlook = rain| T = mild) = 0.9183

IG(outlook, T = mild) = 0.9183 – (2/6 * 1 + 3/6 * 0.9183) = 0.1258

Entropy(Humidity = high | T = mild) = 1

Entropy(Humidity = normal | T = mild) = 0

IG(humidity, T = mild) =0.9183 – (4/6*1) =0.2516

Entropy(Windy = true | T = mild) = 0.9183

Entropy(Windy = false | T = mild) = 0.9183

IG(windy, T = mild) = 0

IG(outlook, T = mild) and IG(windy, T = mild) are both smaller than IG(humidity, T =mild), we choose humidity under branch mild.

IG(outlook, T = cool) = 0.3113

IG(humidity, T = cool) = 0

IG(windy, T = cool) = 0.3113

we choose outlook or windy under the branch cool

For the branch hot, we already touched the leaves.

For the normal branch of the mild part, we already touched the leave.

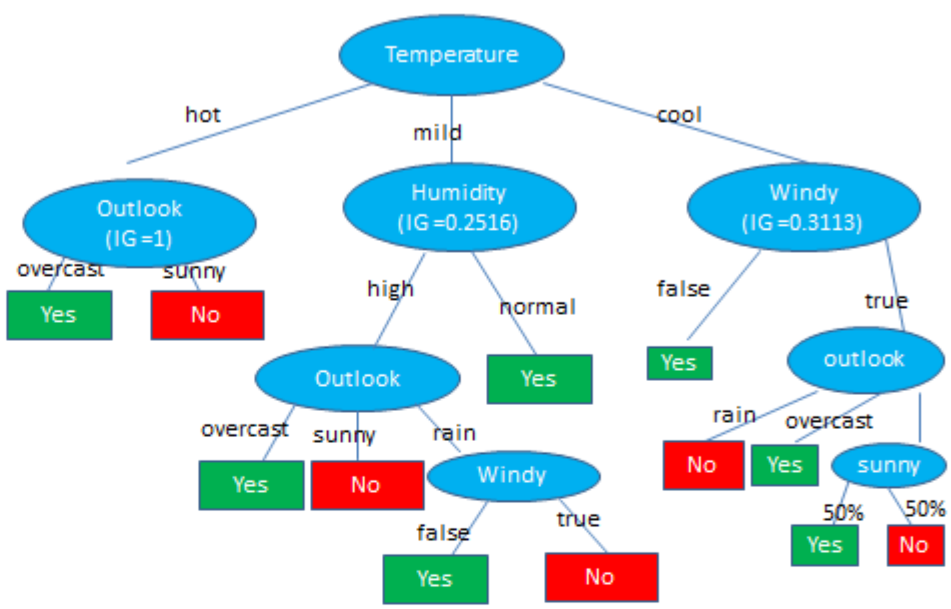
For the high branch of the mild part, we found out that IG(outlook|high) > IG(windy|high)

So we choose outlook under branch high.

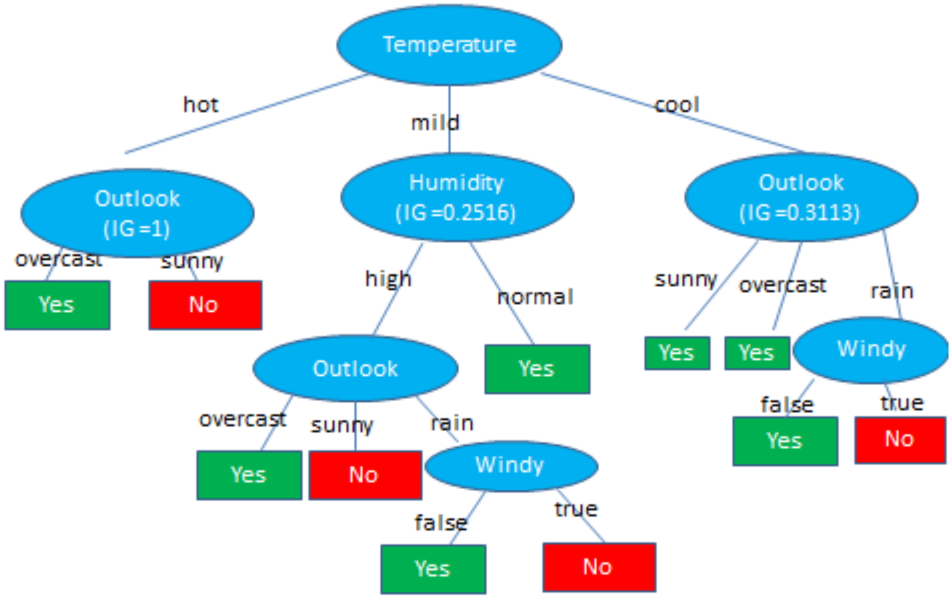
For the branch cool, if we choose outlook under cool, we already touched the leave on the sunny and overcast. In the rain branch, we choose windy, because IG(Humidity|cool,rain) < IG(windy|cool,rain). We can show the final tree marked as (1). If we choose windy under cool, we can show the final tree as the following marked as (2).

Tree 1:

(1)



(2)

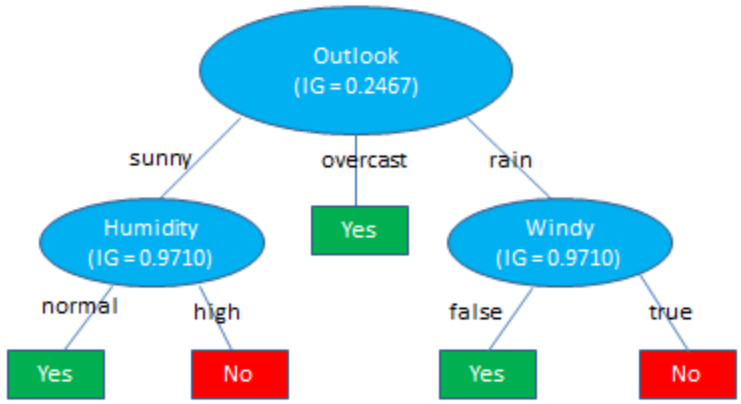


Tree 2:

IG(Outlook) = 0.2467 (In the 2nd tree, we choose outlook as root node)

IG(humidity|sunny) = 0.9710

IG(windy|rain) = 0.9710


Outlook
(IG = 0.2467)
sunny
overcast
rain
Humidity
(IG = 0.9710)
Yes
Windy
(IG = 0.9710)
normal
high
false
true
Yes
No
Yes
No