Q3:

See the details in code.

Gain\_weather = 0.246749819774

Gain\_temperature = 0.029222565659

Gain\_humidity = 0.151835501362

Gain\_wind = 0.0481270304083

We can find weather is the first split point.

For weather:

We can find overcast is 0

If it is overcast, it must be yes.

Set weather’s attributes sunny and rain as root:

Gain\_sunny\_and\_temparature = 0.203910926591

Gain\_ sunny\_and\_humidity = 0.346768069448

Gain\_ sunny\_and\_wind = 0.00713324786499

Gain\_rain\_and\_temparature = 0.00713324786499

Gain\_ rain\_and\_humidity = 0.00713324786499

Gain\_ rain\_and\_wind = 0.346768069448

For weather is sunny, we need consider humidity first and then temperature.

We can prune out data under 0.01:

For rain we only need consider rain and wind condition:

If rain and wind is strong: no

If rain and wind is weak: yes

Set weather’s attributes sunny and humidity’s attributes as root:

Gain\_sunny\_and\_high\_and\_temperature = 0

Gain\_sunny\_and\_normal\_and\_temperature = 0

In this case, we don’t need to consider temperature difference in different humidity.

If weather is sunny and humidity is normal: Yes

If weather is sunny and humidity is high: No

no

high

3.a

yes

normal

humidity

sunny

no

yes

weak

strong

wind

yes

yes

rain

overcast

weather

3.b

For test data 1, predict yes but No in fact;

For test data 2, predict yes and Yes in fact;

For test data 3, predict no and No in fact;

For test data 4, predict yes and Yes in fact;

Confusion matrix:

|  |  |  |
| --- | --- | --- |
| Actual class/ predicted class | Play(predicted) | Not Play(predicted) |
| Play(actual) | 2 | 0 |
| Not Play(actual) | 1 | 1 |

Precision = 2/ (2+1) = 0.67

Recall = 2/ (2+0) = 1

3.c

Gain\_ratio\_weather = 0.156427562421

Gain\_ratio\_temperature = 0.0187726462224

Gain\_ratio\_humidity = 0.151835501362

Gain\_ratio\_wind = 0.0488486155115

Weather’s gain ratio is still the highest, it is still the split point.