lab 3

Data Mining CPS844

Michael Francki 500554567

————1————

a) Instances: 14

b) ~ 63% class = yes ; ~38% = no given no evidence

c) There are 2 instance of outlook = sunny & class = yes
   there are 3 instances of outlook = sunny & class = no
         need to add one two each to deal with attribute class
combinations
         with zero instances in training data to avoid 0 prob

        2+1 = 3
        3+1 = 4

d) there are 9 instances of class = yes
   there are 5 instances of class = no

there are 3 possible values for the attribute outlook

        9+3 = 12
        5+3 = 8

They are not 14 and 5 because that makes no sense

e) Half of the the instance have class = yes
        given no evidence there is 50% class = yes (also 50% class =
no)

          outlook
  sunny   3.0   4.0
  overcast   1.0   1.0
  rainy   4.0   3.0
  [total]   8.0   8.0

there are still 2 instances of outlook = sunny class = yes ; 3 of
outlook = sunny class = no

but overcast now has 0 instances of class = yes OR class = no
0+1 = 1
0+1 = 1

There are now 5 instances of class = yes
and 5 instances of class = no

5+3 = 8
5+3 = 8

————2————–
J48
training set
Correctly Classified Instances   14   100   %

CVx10
Correctly Classified Instances   7   50   %

oneR
Training Set
Correctly Classified Instances   10   71.4286 %

outlook:
        sunny   –> no
        overcast –> yes
        rainy   –> yes

CVx10
Correctly Classified Instances   6   42.8571 %

outlook:
        sunny   –> no
        overcast –> yes
        rainy   –> yes

They create the same rule set they are equally accurate

NB:
Training Set
Correctly Classified Instances   13   92.8571 %

CVx10
Correctly Classified Instances   8   57.1429 %

Although cross validation seems to give worse accuracy given new test
data will probably perform better because just using training data
without CV can lead to over fitting to the training data.

————3————–

weather.arff data, for which humidity and temperature are numeric

J48
training data
Correctly Classified Instances   14   100   %

outlook = sunny
| humidity <= 75: yes (2.0)
| humidity > 75: no (3.0)

Branches on 75

10x-cv
Correctly Classified Instances   9   64.2857 %

outlook = sunny
| humidity <= 75: yes (2.0)
| humidity > 75: no (3.0)

Branches on 75

oneR
training data
Correctly Classified Instances   10   71.4286 %

10x-cv
Correctly Classified Instances   6   42.8571 %

NB
training data
Correctly Classified Instances   13   92.8571 %

10x-cv
Correctly Classified Instances   9   64.2857 %

*J48 seems to be the best

————4————–
Using 10x-cv:

—myWeather1.arff

J48
Correctly Classified Instances   91   91   %

windy = TRUE: yes (50.0/4.0)

```
windy = FALSE: no (50.0/5.0)


oneR
Correctly Classified Instances          91              91      %

windy:
        TRUE    -> yes
        FALSE   -> no


NB
Correctly Classified Instances          91              91      %

*Play when it is windy


—myWeather2.arff

J48
Correctly Classified Instances          80              80      %

outlook = sunny: yes (43.0/4.0)
outlook = overcast
|   temperature = hot: yes (17.0/4.0)
|   temperature = mild: yes (2.0/1.0)
|   temperature = cool: no (3.0)
outlook = rainy
|   temperature = hot: yes (9.0/3.0)
|   temperature = mild: no (8.0/1.0)
|   temperature = cool: no (18.0)


oneR
Correctly Classified Instances          77              77      %


outlook:
        sunny    -> yes
        overcast -> yes
        rainy    -> no



NB
Correctly Classified Instances          88              88      %

*Play if it is sunny or hot out side (mild temp is okay if its not
raining)

—myWeather3.arff
```

```
J48
Correctly Classified Instances          100              100       %
windy = TRUE: no (56.0)
windy = FALSE
|   outlook = sunny: yes (21.0)
|   outlook = overcast: no (12.0)
|   outlook = rainy: no (11.0)

oneR
Correctly Classified Instances           86               86       %
outlook:
        sunny    -> yes
        overcast -> no
        rainy    -> no

NB
Correctly Classified Instances          100              100       %

*Play if it is sunny but don't play if it is windy
```