

CSE5DMI Data Mining Sem 2 2018, Assignment One

Part I

Question 2a)

```
1: X= 'HELLO'
2:
3: def fun():
4:     X='NI'
5:     def nested():
6:         print(X)
7:     nested()
8:
9: fun()
10: print(X)
```

We have defined the variable X to be the string “HELLO”. Then in line 9 we call the function fun. Fun sets the local variable X (defined in the function) to “NI” and then calls nested. Nested prints the value of X which is “NI”.

As the function fun is completed we return to line 10. There it says to print X. Here the value of X will still be “HELLO” as the variable X defined in fun is only visible in that function.

>Output:
NI
HELLO

Question 2b)

As tuples are immutable there is no way to change the tuple itself, but a new one needs to be created in the process. We can for example turn it into a list, change the value, then turn it back into a tuple. I will however use slicing:

```
t = (4, 5, 6)
t = (1,) + t[1:]
```

Part II

Question 3b)

ii) Accuracy and area under (receiver operating characteristic, ROC) curve (AUC)

Accuracy of cross validation: 0.978
AUC: 0.980

Question 3c)

It is possible to limit the number of leaf nodes in the decision tree to 4 such that each leaf node represents a particular class label. Following is the result obtained with the `SklTreeLearner` when setting the parameter `max_leaf_nodes=4`:

Accuracy of cross validation: 0.853

AUC: 0.924

This shows that both the accuracy and AUC have decreased when limiting the number of leaf nodes. It is due to our decision tree being smaller and less complex.

For example, see the difference between figure 1 and 2. Figure 1 shows a decision tree classifier with 4 leaf nodes and figure 2 has 8. For each split in the tree we also increase the number of leaf nodes, thus figure 1 has only 3 attributes whilst figure 2 has 7. If the extra attributes in figure 2 are not redundant i.e. provides more information than price, size and tenants; we will get a more accurate classification. This is also reflected in the result.

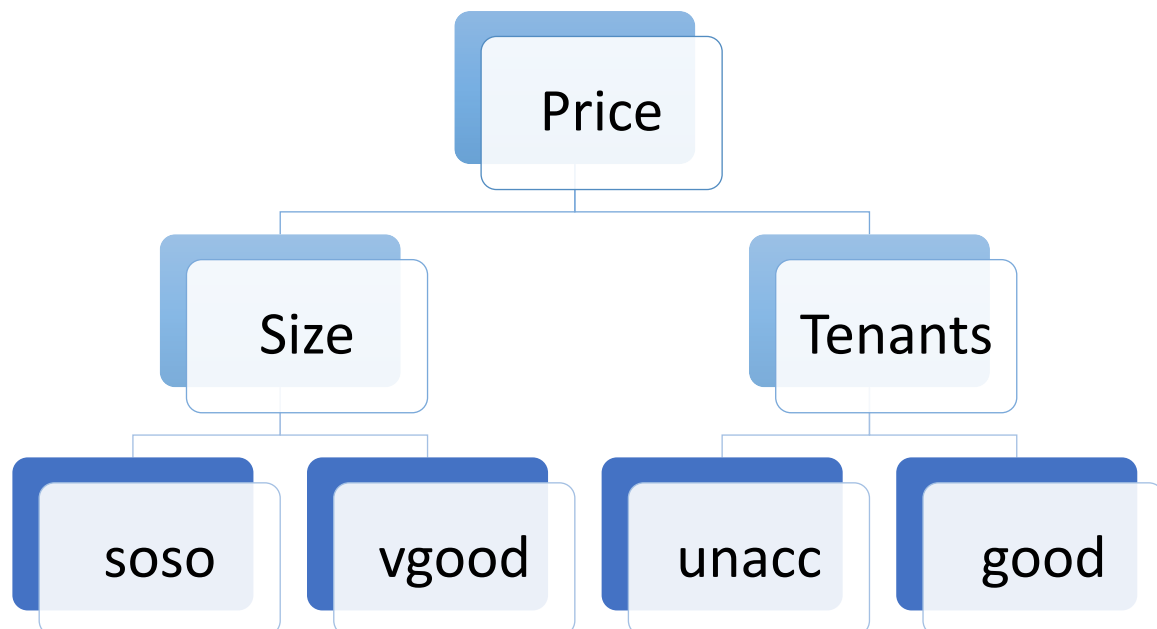


Figure 1: An example of a decision tree classifier with 4 leaf nodes

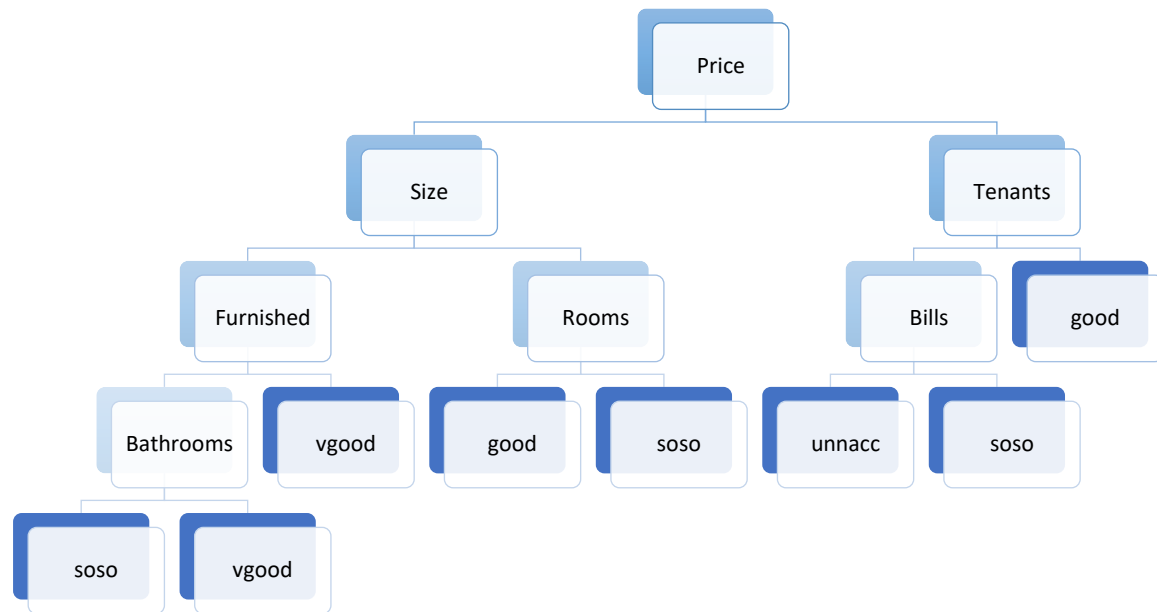


Figure 2: An example of a decision tree classifier with 8 leaf nodes