

Principles of Data Mining: HW 04

Alvin Lin

August 2018 - December 2018

Question 3a

What stopping criteria did you use?

I used recursion depth as the stopping criteria. The algorithm stopped after branching a certain number of times.

Question 3b

Did you use any pruning or post-pruning?

No.

Question 3c

What splitting decision were you using?

I used weighted Gini index to determine the purity of the split.

Question 3d

What structure did your final decision tree classifier have?

```
for row in f:
    data = preprocess_training_data_row(row)
    if data[3] >= 25: # Sugar
        if data[1] >= 30: # FlourOrOats
            category = 'Muffin'
        else:
            category = 'Cupcake'
    else:
        if data[4] >= 18: # Butter
```

```
        category = 'Cupcake'
    else:
        category = 'Muffin'
```

Question 3e

Run the original training data back through your classifier. What was the accuracy of your resulting classifier on the training data?

On the training data, there were 42 correct classifications and 19 misclassifications, giving us an accuracy rate of 68.9%.

Question 3f

Did your program actually create the classifier program, or did it generate the attribute list and thresholds for you to hand-code in?

My program generated a decision tree stored as nodes and then translated those nodes into Pythonic if-else statements.

Question 3g

What else did you learn along the way?

I learned a lot more about recursion in Python since I had forgotten the mechanics of how it worked. It was a good refresher on how to use recursion in Python and the application was something I had never done before.

Question 3h

What can you conclude?

Overall, none of the attributes are by themselves a very good indicator of whether a recipe is cupcake or muffin. It seems that sugar is the best starting attribute since cupcakes and muffins are most divided by this attribute. Using this in combination with FlourOrOats and Butter lead me to the best distinction.

If you have any questions, comments, or concerns, please contact me at
alvin@omgimanerd.tech