Hunter Leise

CSCI 3202

Problem Set 3

**Problem 3.1**

(For b, c, d, and e, black nodes mean the variable is true, and red nodes mean the variable is false. Black leaf nodes mean that path matches the goal criteria, and red leaf nodes mean that path does not match the goal criteria. The variables are located on the right side of the graph.)

1. P(N, C, L) =0.7 \* 0.4 \* 0.2 = 0.056

P(~N, C, L) = 0.3 \* 0.4 \* 0.6 = 0.072

P(N, ~C, L) = 0.7 \* 0.6 \* 0.5 = 0.21

P(~N, ~C, L) = 0.3 \* 0.6 \* 0.8 = 0.144

P(L) = 0.056 + 0.072 + 0.21 + 0.144

= **0.482**

P(B) = P(B|L)P(L) + P(B|~L)P(~L)

= 0.9(0.482) + 0.2(0.518)

= **0.5374**

P(M) = P(M|L)P(L) + P(M|~L)P(~L)

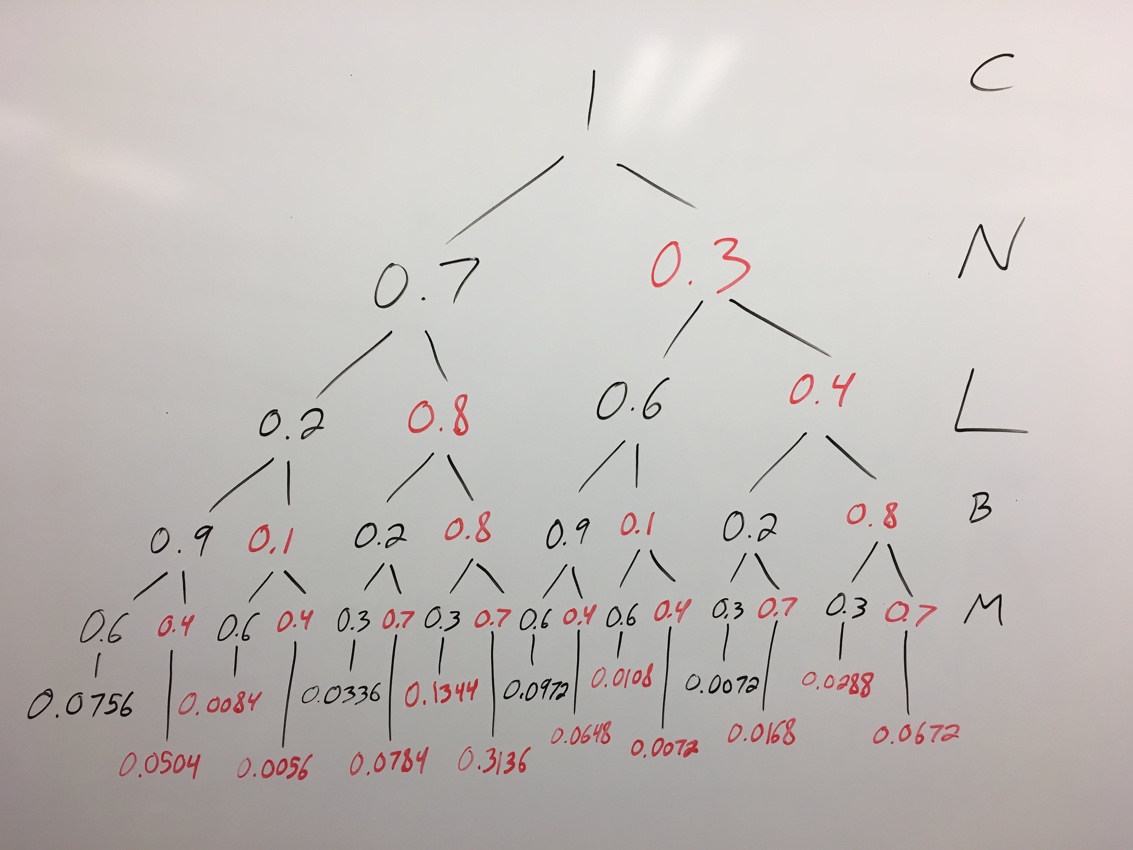
= 0.6(0.482) + 0.3(0.518)

= **0.4446**

P(S) = P(S|M)P(M) + P(S|~M)P(~M)

= 0.8(0.4446) + 0.1(0.5554)

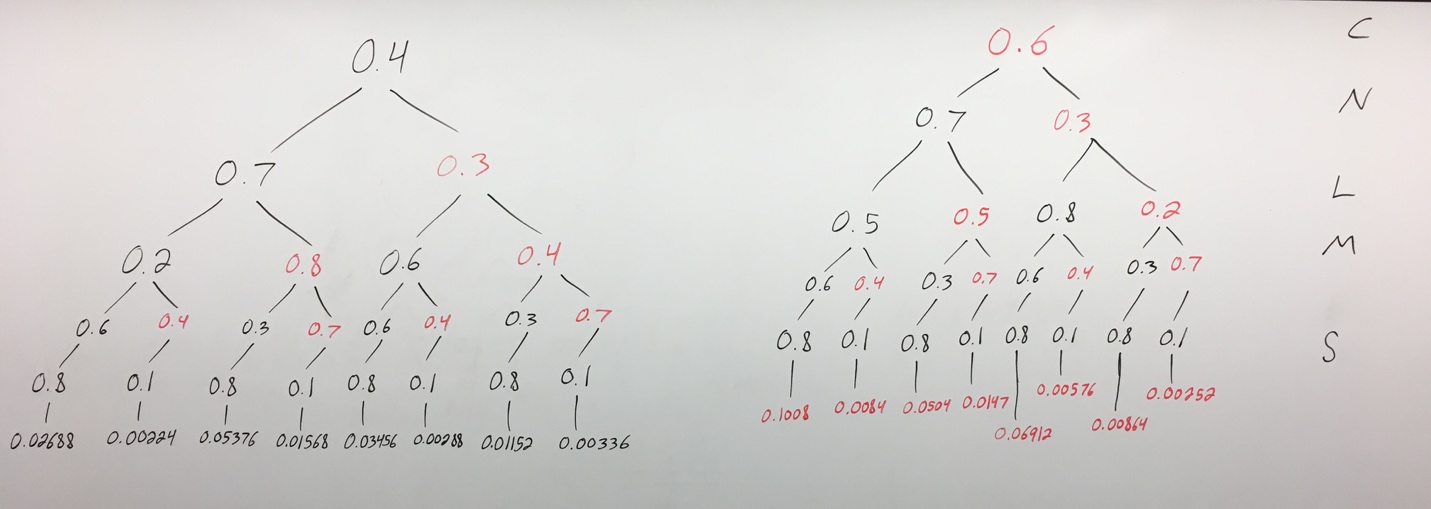
= **0.41122**



P(B, M|C) = summation of black leaf nodes / summation of all leaf nodes

= 0.2136 / 1

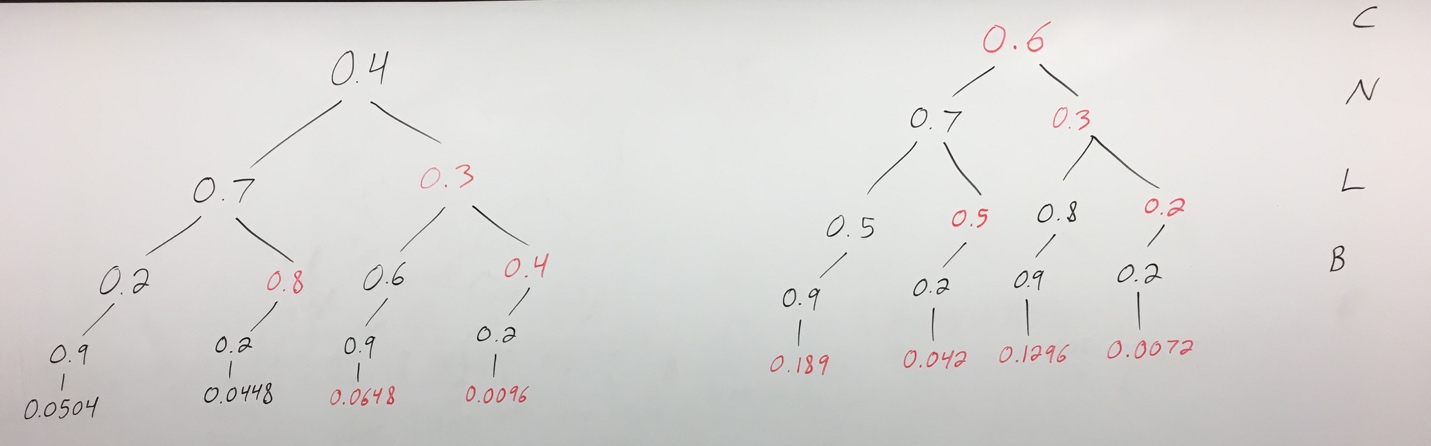
= **0.2136**



P(C|S) = summation of black leaf nodes / summation of all leaf nodes

= 0.15088 / 0.41122

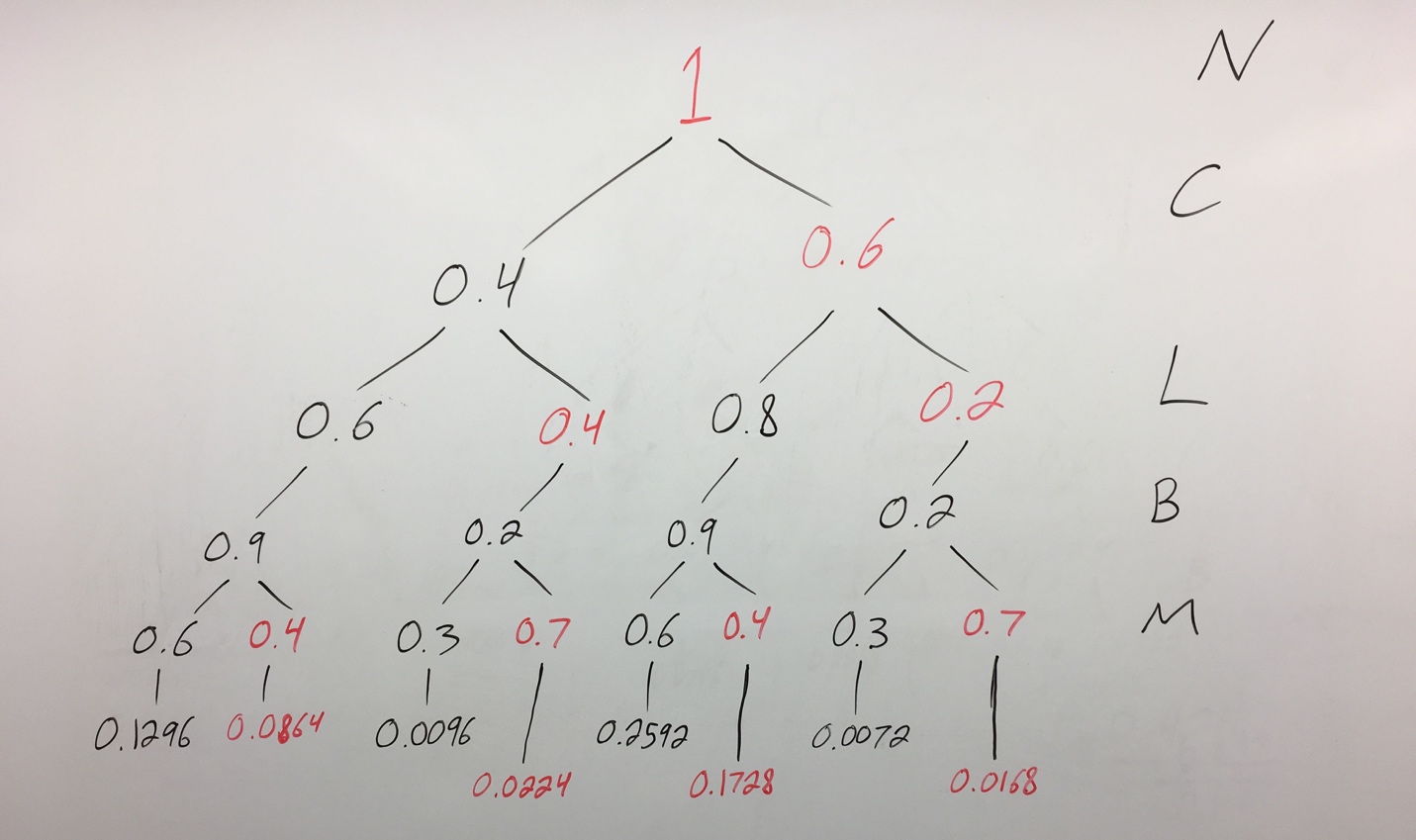
≈ **0.3669**



P(C, N|B) = summation of black leaf nodes / summation of all leaf nodes

= 0.0952 / 0.5374

≈ **0.17715**



P(M|B, ~N) = summation of black leaf nodes / summation of all leaf nodes

= 0.4056 / 0.704

≈ **0.5761**

**Problem 3.2**

1. Starting information =

=

= 1

Attribute A Set 1 =

= 0.6 \* (0.38998 + 0.52832)

= 0.6 \* (0.9183)

≈ 0.55098

Attribute A Set 2 =

= 0.4 \* (0.5 + 0.31128)

= 0.4 \* (0.811280

≈ 0.32451

Attribute A = 0.55098 + 0.324512

≈ 0.87549

Attribute B Set 1 =

= 0.875 \* (0.4613 + 0.5239)

= 0.875 \* (0.9852)

≈ 0.8621

Attribute B Set 2 =

= 0.125 \* (0+ 0)

= 0.125 \* (0)

= 0

Attribute B = 0.8621 + 0

≈ 0.8621

**Attribute B resulted in the most decrease in entropy, so B is the more informative attribute.**

1. Starting information =

=

= 1

Attribute A Set 1 =

= 0.5 \* (0.2575 + 0.4644)

= 0.5 \* (0.7219)

≈ 0.36095

Attribute A Set 2 =

= 0.5 \* (0.4644 + 0.2575)

= 0.5 \* (0.7219)

≈ 0.36095

Attribute A = 0.36095 + 0.36095

≈ 0.7219

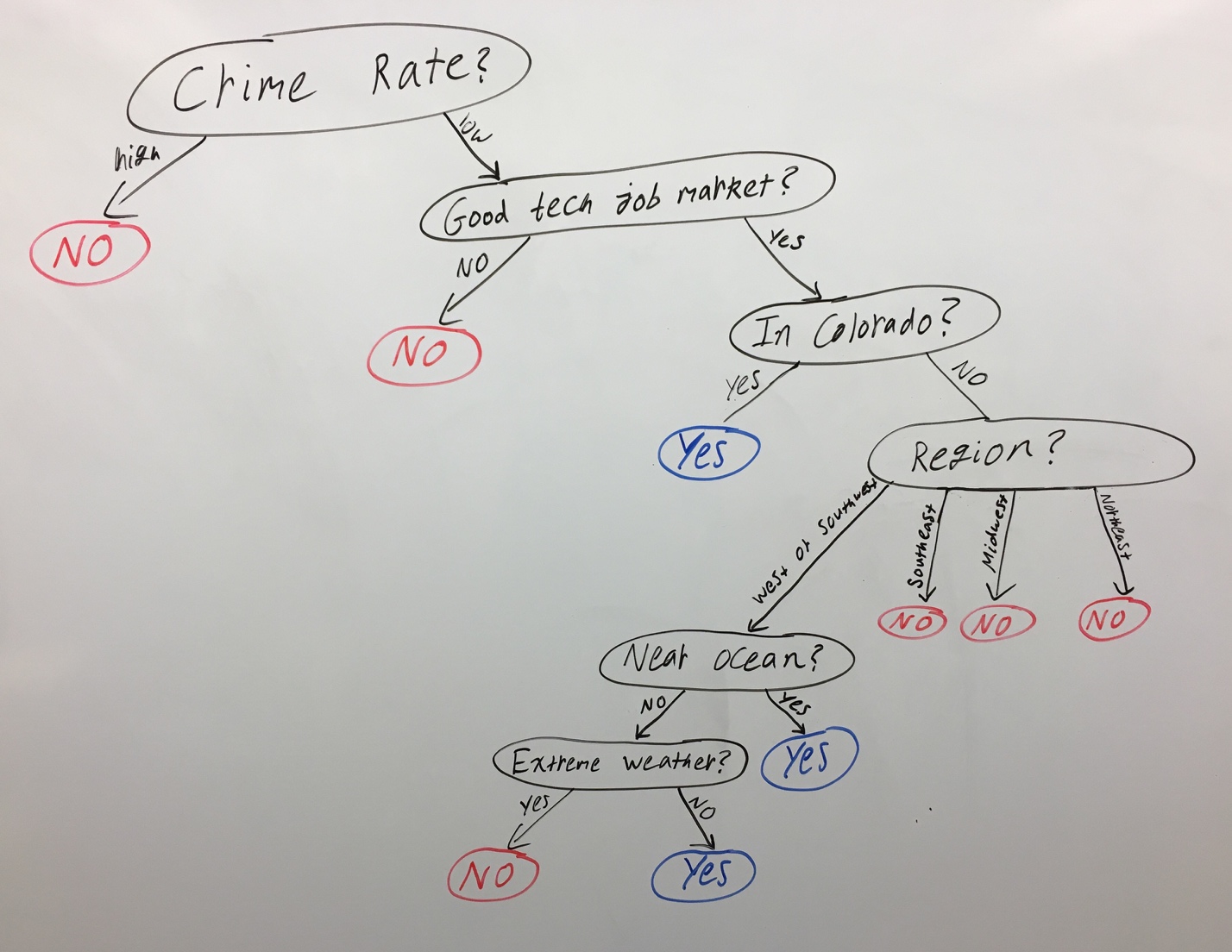
**The largest value of X such that Attribute B is more informative than Attribute A (0.7219) is 27 with an entropy value of ≈0.7197.**

I got this answer by trying different values of x in the following code:

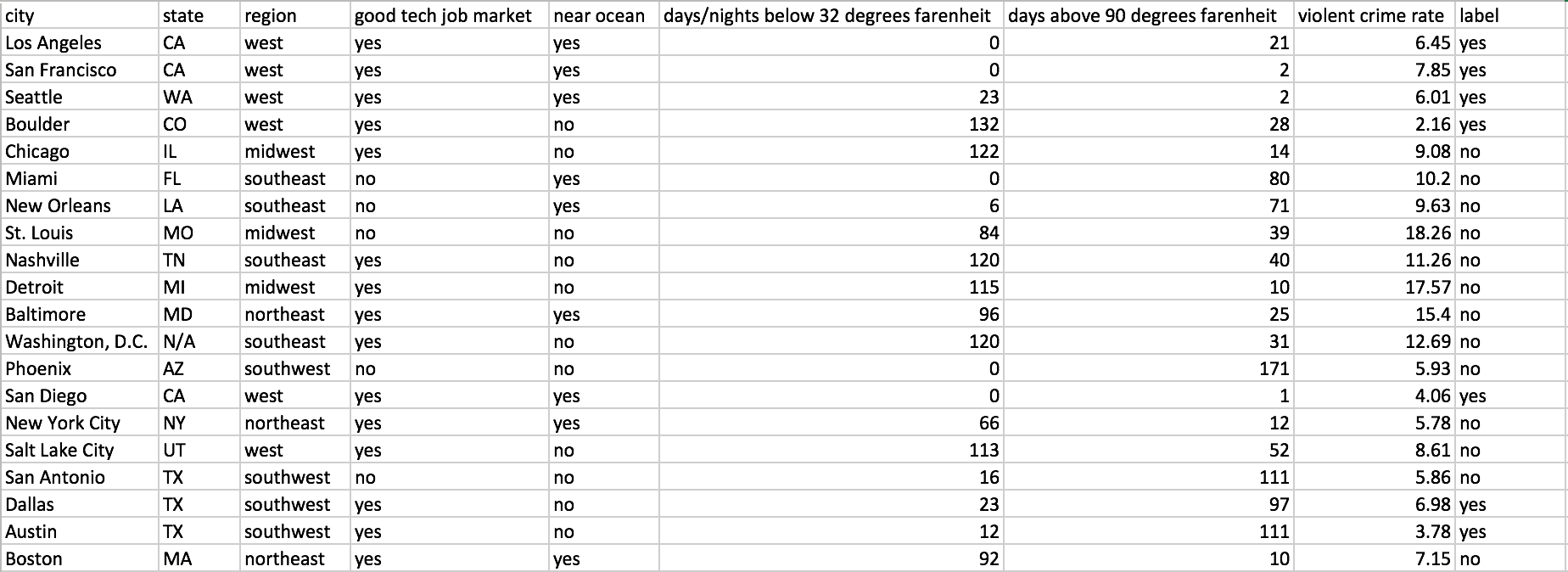
from math import log2  
  
 x = 27  
  
 set1\_yes = x  
 set1\_no = 50  
  
 set2\_yes = x  
 set2\_no = 50  
  
 set3\_yes = (100 - 2 \* x)  
 set3\_no = 0  
  
 set1\_total = set1\_yes + set1\_no  
 set2\_total = set2\_yes + set2\_no  
 set3\_total = set3\_yes + set3\_no  
  
 set1\_weight = set1\_total / (set1\_total + set2\_total + set3\_total)  
 set2\_weight = set2\_total / (set1\_total + set2\_total + set3\_total)  
 set3\_weight = set3\_total / (set1\_total + set2\_total + set3\_total)  
  
 pa = set1\_yes / set1\_total *# P(set 1 yes)* pb = set1\_no / set1\_total *# P(set 1 no)* pc = set2\_yes / set2\_total *# P(set 2 yes* pd = set2\_no / set2\_total *# P(set 2 no)* pe = set3\_yes / set3\_total *# P(set 3 yes)* pf = set3\_yes / set3\_total *# P(set 3 no)* set1\_entropy = set1\_weight \* ((-pa \* log2(pa)) + (-pb \* log2(pb)))  
 set2\_entropy = set2\_weight \* ((-pc \* log2(pc)) + (-pd \* log2(pd)))  
 set3\_entropy = set3\_weight \* ((-pe \* log2(pe)) + (-pf \* log2(pf)))  
  
 entropy = set1\_entropy + set2\_entropy + set3\_entropy  
 print(entropy)

**Problem 3.3**

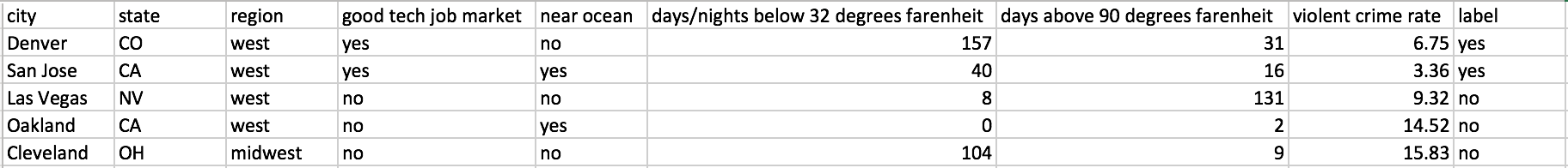
1. Would I like to live here?



1. Training Examples:



Test Examples:



Output Tree:

1. The tree produced by the program recognizes that the crime rate is the most useful attribute for splitting the data. It’s actually much more concise than the tree I made in part A by not even needing to use the temperature and ocean attributes. I find it really interesting that it used the crime rate of Salt Lake City as the border between labels. This results in the correct categorization, but the crime rate isn’t the reason I wouldn’t want to live in Salt Lake City.

I definitely see my own decision process in the computer-produced tree. Of course, there are so many factors to why I would choose to live somewhere, but given the few attributes that I listed, the computer did a really good job!

1. Both my cognitive tree and the computer-generated tree outputs the correct prediction for all five test cases.