Willis Allstead
4/4/18
CS 491

Lab 4

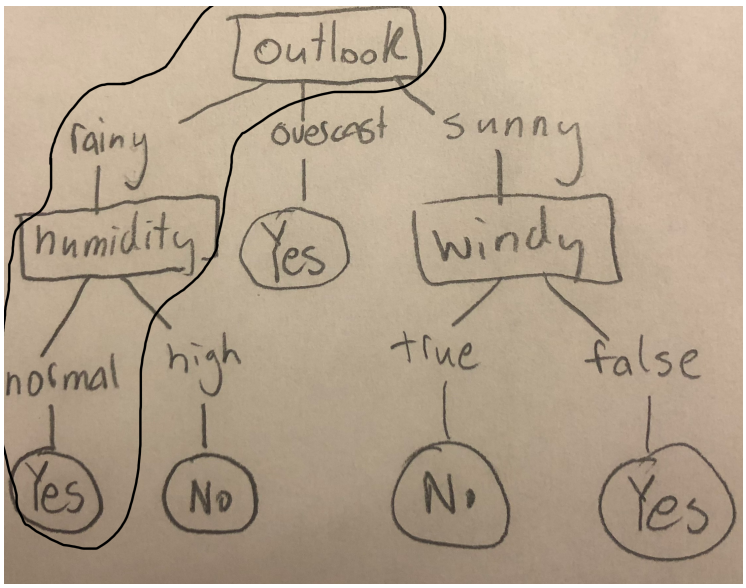
Data

1) Using the python code provided in the prompt and a cosine similarity function, I found the cosine similarity of the methods section of my paper and three papers I referenced.

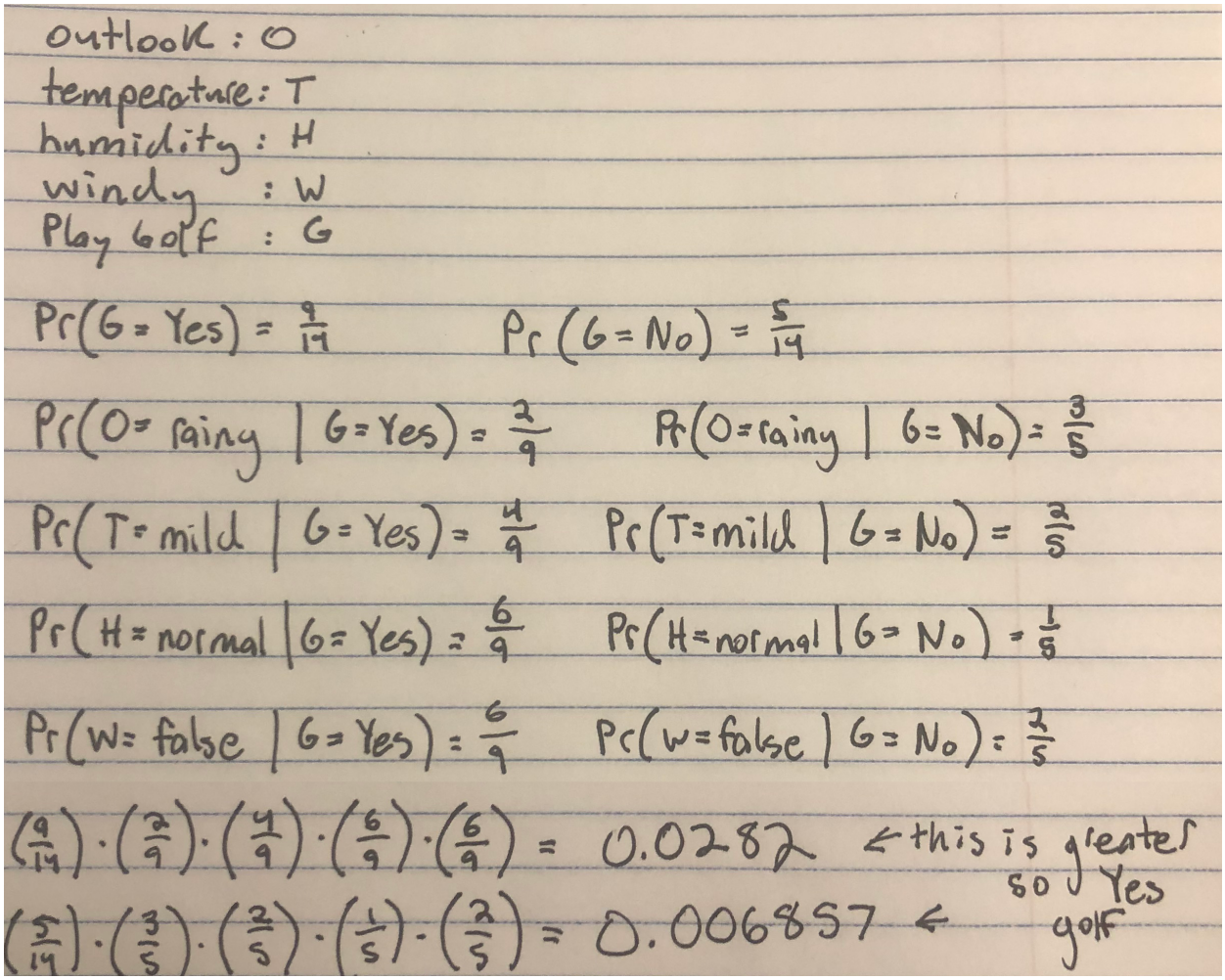
My results were:
Similarity to *Retrieving Similar Lyrics for Music Recommendation System*
= 0.06226378
Similarity to *Ranking Lyrics for Online Search*
= 0.02393933
Similarity to *Lyric-based Music Recommendation* (A paper written in this class a previous year I think)
= 0.03975788

While the similarity between my methods section and these papers' methods sections seems low given these numbers near 0, from looking at the papers and comparing similarity rankings from 1-3, I would have ranked them in the same way. Also, making use of stemming algorithms and similar ways of getting rid of noise in the sets of words would probably result in more meaningful data.

Supervised Learning



2) When outlook is rainy, temperature is mild, humidity is normal and it is not windy the path circled in black is the one that will be followed. As you can see, the temperature and wind level aren't even required to know the answer based on the dataset.



3) As is shown in the calculations above, "Yes" is the most likely decision for playing golf when outlook is rainy, temperature is mild, humidity is normal and it is not windy.

Data instance	Outlook	Temp	Humidity	Windy	Similarity	Label	K	Prediction
11	1	1	1	0	3	Yes	1	Yes
8	1	1	0	1	3	No	2	?
10	0	1	1	1	3	Yes	3	Yes
9	1	0	1	1	3	Yes	4	Yes
13	0	0	1	1	2	Yes	5	Yes
1	1	0	0	1	2	No	6	Yes
2	1	0	0	0	1	No	7	Yes
7	0	0	1	0	1	Yes	8	Yes
3	0	0	0	1	1	Yes	9	Yes
14	0	1	0	0	1	No	10	Yes

3) Using *k*-NN to predict whether we will golf given that the outlook is rainy, temperature is mild, humidity is normal and it is not windy, we have the values in the prediction column above for each *k* from 1-10

Data instances are numbered from 1-14 from the top of the given table to the bottom.