Assignment 5

Task 1:
P(T1) = Temp above 80 (first email)
P(NOT(T1)) = Temp below 80 (first email)
P(Maine) = P(M) = 0.05
P(Sahara) = P(NOT(M)) = 0.95
P(T1 | M) = 0.2
P(T1 | NOT(M)) = 0.9

Part a:
Need to find P(M | NOT(T1)).
P(NOT(T1)) = P(NOT(T1) AND M) + P(NOT(T1) AND NOT(M))
= P(NOT(T1) | M)P(M) + P(NOT(T1) | NOT(M))P(NOT(M))
= (1 - P(T1 | M))P(M) + (1 - P(T1 | NOT(M)))(1-P(M))
= (1 - 0.2)(0.05) + (1 - 0.9)(1 - 0.05)
= (0.8)(0.05) + (0.1)(0.95)
= 0.04 + 0.095 = 0.135

P(M | NOT(T1)) * P(NOT(T1)) = P(NOT(T1) | M) * P(M)
P(M | NOT(T1)) = (P(NOT(T1) | M) * P(M))/P(NOT(T1))
= ((1 - P(T1 | M)) * P(M))/P(NOT(T1))
= ((1 - 0.2) * 0.05) / 0.135
= (0.04) / (0.135) = 0.2963

P(M | NOT(T1)) = 0.2963

Part b:
I am assuming that the sensor will stay in a fixed location and won't be moved between days. According to the problem, each daily high is conditionally independent from previous daily highs, given that they are read at the same location (which I have assumed). That means that the probability of reading a daily high below 80 degrees is the same from day-to-day. Therefore, given that the first e-mail reads under 80 degrees, the probability that the second e-mail will also read below 80 degrees P(NOT(T2)) = P(NOT(T1)) = 0.135.

Part c:
For Part C, I will operate under the same assumption as part B. That being the case, I can use the product rule to calculate the probability of the first three emails all indicating daily highs under 80. P(NOT(T1) & NOT(T2) & NOT(T3)) = NOT(T1) * NOT(T2) * NOT(T3) = 0.135 * 0.135 * 0.135 = 0.00246

Task 2:
Part a:
# of numbers in joint distribution table (assuming no conditional independence) = (5^1)*(7^10) - 1 = 1412376244

Part b:

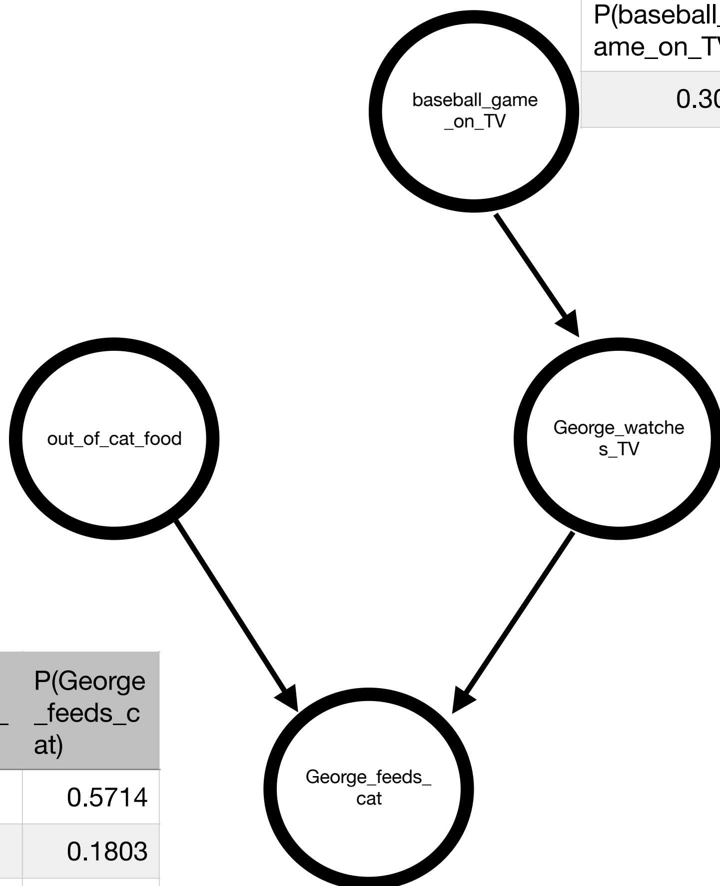
Given A, P(B1) does not depend on whether B2/B3/…/B10 are true. This removes 9 numbers for each B variable, so we are down 90 more variables. P(Bi | A, Bj) = P(Bi | A).

Given NOT(A), P(B1) does not depend on whether B2/B3/…/B10 are true. This removes 9 numbers for each B variable, so we are down 90 more variables. P(Bi | NOT(A), Bj) = P(Bi | NOT(A)).

So, number of independent numbers in joint distribution table = 1412376244 - 90 - 90 = 1412376064.

After accounting for the conditional independence relations, the space complexity of the representation of this joint distribution goes down from exponential in n to linear in n. The amount of numbers needed to store in this representation is the same as the number of independent numbers in the joint distribution table, so 1412376064.

Task 3:



P(baseball_game_on_TV)
0.3041

P(out_of_cat_food)
0.1671

baseball_game_on_TV	P(George_watches_TV)
T	0.5586
F	0.2795

out_of_cat_food	George_watches_TV	P(George_feeds_cat)
F	T	0.5714
T	F	0.1803
T	T	0.0417
F	F	0.9639

Task 4:
For this task, I used Numbers (OSX equivalent of Excel) to do my work. The txt file automatically populates the spreadsheet once it is opened. Basically, I marked off each row that fulfilled the conditions I was looking for (using a one, so I could count them), and then used the built-in SUM functionality to count the marked rows. From there, it was just math. The data (along with the counts I made) is contained in the file named training_data.xlsx. I've added tables to the above Bayesian Network to reflect the newly found conditional probabilities.

Task 5:
Part a:
The Markov Blanket of node L consists of the following nodes: G, Q, P, K, M

Part b:
No matter what, the following is always true:
$P(A \mid B) = P(A, B) / P(B)$

So, if $P(A) = 0.8$ and $P(F \mid A) = 0.8$
then $P(F \mid A) = P(F, A)/ P(A)$
$P(F, A) = P(F \mid A) * P(A)$
$= 0.8 * 0.8 = 0.64$

$P(F, A) = 0.64$

Part c:
$P(M, NOT(C) \mid H)$

Given H is true, $P(M) = 0.1$
Given H is true, $P(H) = 0.6$
$P(NOT(C)) = 0.4$

$P(M, NOT(C) \mid H) * P(H) = P(H \mid M, NOT(C)) * P(M, NOT(C))$
$P(M, NOT(C) \mid H) = (P(H \mid M, NOT(C)) * P(M, NOT(C))) / P(H)$
$= (P(H \mid NOT(C)) * (P(M) * P(NOT(C)))) / P(H)$
$= (0.1 * (0.1 * 0.4)) / 0.6$
$= 0.00667$

$P(M, NOT(C) \mid H) = 0.00667$