CS221 Problem Workout

Week 8

1) **[CA session] Problem 1: The Bayesian Bag of Candies Model**

You have a lot of candy left over from Halloween, and you decide to give them away to your friends. You have four types of candy: **A**pple, **B**anana, **C**aramel, **D**ark-Chocolate. You decide to prepare candy bags using the following process.

- For each candy bag, you first flip a (biased) coin Y which comes up heads ($Y = \text{H}$) with probability λ and tails ($Y = \text{T}$) with probability $1 - \lambda$.
- If Y comes up heads ($Y = \text{H}$), you make a **H**ealthy bag, where you:
  (a) Add one **A**pple candy with probability p_1 or nothing with probability $1 - p_1$;
  (b) Add one **B**anana candy with probability p_1 or nothing with probability $1 - p_1$;
  (c) Add one **C**aramel candy with probability $1 - p_1$ or nothing with probability p_1;
  (d) Add one **D**ark-Chocolate candy with probability $1 - p_1$ or nothing with probability p_1.
- If Y comes up tails ($Y = \text{T}$), you make a **T**asty bag, where you:
  (a) Add one **A**pple candy with probability p_2 or nothing with probability $1 - p_2$;
  (b) Add one **B**anana candy with probability p_2 or nothing with probability $1 - p_2$;
  (c) Add one **C**aramel candy with probability $1 - p_2$ or nothing with probability p_2;
  (d) Add one **D**ark-Chocolate candy with probability $1 - p_2$ or nothing with probability p_2.

For example, if $p_1 = 1$ and $p_2 = 0$, you would deterministically generate: **H**ealthy bags with one **A**pple and one **B**anana; and **T**asty bags with one **C**aramel and one **D**ark-Chocolate. For general values of p_1 and p_2, bags can contain anywhere between 0 and 4 pieces of candy.

Denote A, B, C, D random variables indicating whether or not the bag contains candy of type **A**pple, **B**anana, **C**aramel, and **D**ark-Chocolate, respectively.

a. (*1 point*)

(i) Draw the Bayesian network corresponding to process of creating a single bag.

(ii) What is the probability of generating a **H**ealthy bag containing **A**pple, **B**anana, **C**aramel, and not **D**ark-Chocolate? For compactness, we will use the following notation to denote this possible outcome:

$$(\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}, \textbf{C}\text{aramel}\}).$$

(iii) What is the probability of generating a bag containing **A**pple, **B**anana, **C**aramel, and *not* **D**ark-Chocolate?

(iv) What is the probability that a bag was a **T**asty one, given that it contains **A**pple, **B**anana, **C**aramel, and *not* **D**ark-Chocolate?

b. (*1 point*)

You realize you need to make more candy bags, but you've forgotten the probabilities you used to generate them. So you try to estimate them looking at the 5 bags you've already made:

$$\begin{aligned}
&\textit{bag } 1: && (\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}\}) \\
&\textit{bag } 2: && (\textbf{T}\text{asty}, \{\textbf{C}\text{aramel}, \textbf{D}\text{ark-Chocolate}\}) \\
&\textit{bag } 3: && (\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}\}) \\
&\textit{bag } 4: && (\textbf{T}\text{asty}, \{\textbf{C}\text{aramel}, \textbf{D}\text{ark-Chocolate}\}) \\
&\textit{bag } 5: && (\textbf{H}\text{ealthy}, \{\textbf{A}\text{pple}, \textbf{B}\text{anana}\})
\end{aligned}$$

Estimate λ, p_1, p_2 by maximum likelihood.

Estimate λ, p_1, p_2 by maximum likelihood, using Laplace smoothing with parameter 1.

c. (*1 point*) You find out your little brother had been playing with your candy bags, and had mixed them up (in a uniformly random way). Now you don't even know which ones were **H**ealthy and which ones were **T**asty. So you need to re-estimate λ, p_1, p_2, but now without knowing whether the bags were **H**ealthy or **T**asty.

$$
\begin{array}{ll}
\textit{bag } 1: & (? \ , \ \{\mathbf{A}\text{pple}, \mathbf{B}\text{anana}, \mathbf{C}\text{aramel}\}) \\
\textit{bag } 2: & (? \ , \ \{\mathbf{C}\text{aramel}, \mathbf{D}\text{ark-Chocolate}\}) \\
\textit{bag } 3: & (? \ , \ \{\mathbf{A}\text{pple}, \mathbf{B}\text{anana}, \mathbf{C}\text{aramel}\}) \\
\textit{bag } 4: & (? \ , \ \{\mathbf{C}\text{aramel}, \mathbf{D}\text{ark-Chocolate}\}) \\
\textit{bag } 5: & (? \ , \ \{\mathbf{A}\text{pple}, \mathbf{B}\text{anana}, \mathbf{C}\text{aramel}\})
\end{array}
$$

You remember the EM algorithm is just what you need. Initialize with $\lambda = 0.5, p_1 = 0.5, p_2 = 0$, and run one step of the EM algorithm.

(i) E-step:

(ii) M-step:

d. (*1 point*)

You decide to make candy bags according to a new process. You create the first one as described above. Then with probability μ, you create a second bag of the same type as the first one (Healthy or Tasty), and of different type with probability $1-\mu$. Given this type, the bag is filled with candy as before. Then with probability μ, you create a third bag of the same type as the second one (Healthy or Tasty), and of different type with probability $1-\mu$. And so on, you repeat the process M times. Denote Y_i, A_i, B_i, C_i, D_i the variables at each time step, for $i = 0, \ldots, M$. Let $X_i = (A_i, B_i, C_i, D_i)$.

Now you want to compute:

$$\mathbb{P}(Y_i = \mathbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0))$$

exactly for all $i = 0, \ldots, M$, and you decide to use the forward-backward algorithm.

Suppose you have already computed the marginals:

$$f_i = \mathbb{P}(Y_i = \mathbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0))$$

for some $i \geq 0$. Recall the first step of the algorithm is to compute an intermediate result *proportional* to

$$\mathbb{P}(Y_{i+1} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

(i) Write an expression that is **proportional** to

$$\mathbb{P}(Y_{i+1} = \mathbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

in terms of f_i and the parameters p_1, p_2, λ, μ.

(ii) Write an expression that is **proportional** to

$$\mathbb{P}(Y_{i+1} = \textbf{T}\text{asty} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

in terms of f_i and the parameters of the model p_1, p_2, λ, μ. The proportionality constant should be the same as in (i).

(iii) Let h be the answer for part (i), and t for part (ii). Write an expression for

$$\mathbb{P}(Y_{i+1} = \textbf{H}\text{ealthy} \mid X_0 = (1,1,1,0), \ldots, X_i = (1,1,1,0), X_{i+1} = (1,1,1,0))$$

in terms of h, t and the parameters of the model p_1, p_2, λ, μ.