

CS221 Problem Workout **Solutions**

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 ($Y = H$) with probability λ and tails ($Y = T$) with probability $1 - \lambda$.
- If Y comes up heads ($Y = H$), 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 ($Y = T$), 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.

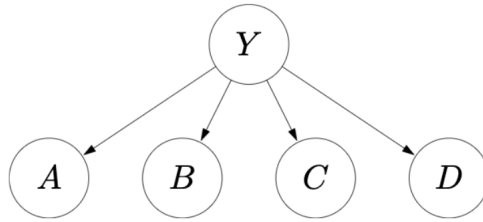


Figure 1: Bayesian network for a single candy bag.

a. (1 point)

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

Solution Solution for part (i) is shown in Figure 1.

(ii) What is the probability of generating a **Healthy** bag containing **Apple**, **Banana**, **Caramel**, and **not Dark-Chocolate**? For compactness, we will use the following notation to denote this possible outcome:

(**Healthy**, {**Apple**, **Banana**, **Caramel**}).

Solution By definition, we create a **Healthy** bag with probability λ , and include the candies with probability $p_1 p_1 (1 - p_1) p_1$, so the result is

$$\lambda p_1 p_1 (1 - p_1) p_1$$

(iii) What is the probability of generating a bag containing **Apple**, **Banana**, **Caramel**, and *not* **Dark-Chocolate**?

Solution The bag could be **Healthy** or **Tasty**. We have computed the probability for the **Healthy** case above. For a **Tasty** one, a similar computation gives

$$(1 - \lambda) p_2 p_2 (1 - p_2) p_2$$

so the result is:

$$\lambda p_1 p_1 (1 - p_1) p_1 + (1 - \lambda) p_2 p_2 (1 - p_2) p_2$$

(iv) What is the probability that a bag was a **Tasty** one, given that it contains **Apple**, **Banana**, **Caramel**, and *not* **Dark-Chocolate**?

Solution Using the definition of conditional probability, we get:

$$\frac{(1 - \lambda) p_2 p_2 (1 - p_2) p_2}{\lambda p_1 p_1 (1 - p_1) p_1 + (1 - \lambda) p_2 p_2 (1 - p_2) p_2}$$

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:

bag 1 :	(H ealthy, { A pple, B anana})
bag 2 :	(T asty, { C aramel, D ark-Chocolate})
bag 3 :	(H ealthy, { A pple, B anana})
bag 4 :	(T asty, { C aramel, D ark-Chocolate})
bag 5 :	(H ealthy, { A pple, B anana})

Estimate λ, p_1, p_2 by maximum likelihood.

Solution Out of 5 bags, 3 are **H**ealthy, so $\lambda = 3/5$. To estimate p_1 , we only consider the 3 healthy bags. For a **H**ealthy bag, the probability of adding **A**pple, **B**anana, not **C**aramel, and not **D**ark-Chocolate is $(p_1)^4$. For the three bags, the probability becomes $(p_1)^{12}$, which is maximized for $p_1 = 1$. Equivalently, to generate 3 **H**ealthy bags, we flip a (biased) coin of parameter p_1 12 times. Since we observe 12 "heads", the maximum likelihood estimate is $p_1 = 1$. To generate 2 **T**asty bags, we flip a (biased) coin of parameter p_2 8 times. Since we observe 0 "heads", the maximum likelihood estimate is $p_2 = 0$.

- $$\lambda = 3/5$$

- $$p_1 = 12/12 = 1$$

- $$p_2 = 0/8 = 0$$

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

Solution We just need to increment the counts in the previous solution by 1.

- $$\lambda = 4/7$$

- $$p_1 = 13/(13 + 1)$$

- $$p_2 = 1/(1 + 9)$$

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.

bag 1 : (? , {**A**pple, **B**anana, **C**aramel})
 bag 2 : (? , {**C**aramel, **D**ark-Chocolate})
 bag 3 : (? , {**A**pple, **B**anana, **C**aramel})
 bag 4 : (? , {**C**aramel, **D**ark-Chocolate})
 bag 5 : (? , {**A**pple, **B**anana, **C**aramel})

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:

Solution To evaluate $P(Y = T | \{A, B, C\})$ we plug in the parameter values in the formula in (a),(iv), obtaining $P(Y = T | \{A, B, C\}) = 0$. To evaluate $P(Y = T | \{C, D\})$ we use a similar formula obtaining

$$P(Y = T | \{C, D\}) = \frac{(1 - \lambda)(1 - p_2)^4}{\lambda(1 - p_1)^4 + (1 - \lambda)(1 - p_2)^4} = \frac{16}{17}$$

The resulting weighted dataset is:

- (**H**ealthy, $\{A, B, C\}$), 1×3
- (**T**asty, $\{A, B, C\}$), 0
- (**H**ealthy, $\{C, D\}$), $1/17 \times 2$
- (**T**asty, $\{C, D\}$), $16/17 \times 2$

(ii) M-step:

Solution Now we just do counts like in part (b). There are $3 + 2/17$ **H**ealthy bags out of 5. For p_1 , each (**H**ealthy, $\{A, B, C\}$) corresponds to 3 "heads" and 1 "tail" (probability $p_1 p_1 (1 - p_1) p_1$). Each (**H**ealthy, $\{C, D\}$) corresponds to 4 "tails" $((1 - p_1)^4)$. For p_2 , each (**T**asty, $\{C, D\}$) corresponds to 4 "tails" $((1 - p_2)^4)$. The new parameters are:

$$\begin{aligned} \lambda &= (3 + 2/17)/5 \\ p_1 &= 9/(9 + 3 + 4 * 2/17) \\ p_2 &= 0 \end{aligned}$$

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 $i = 0, \dots, M$. Let $X_i = (A_i, B_i, C_i, D_i)$.

Now you want to compute:

$$\mathbb{P}(Y_i = \mathbf{Healthy} \mid X_0 = (1, 1, 1, 0), \dots, X_i = (1, 1, 1, 0))$$

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

Suppose you have already computed the marginals:

$$f_i = \mathbb{P}(Y_i = \mathbf{Healthy} \mid X_0 = (1, 1, 1, 0), \dots, X_i = (1, 1, 1, 0))$$

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

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

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

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

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

Solution Emission: When $Y_{i+1} = \mathbf{Healthy}$, the probability of observing $X_{i+1} = (1, 1, 1, 0)$ is $p_1 p_1 (1 - p_1) p_1$ as in part (a),(ii).

Transition: There are two cases: either $Y_i = \mathbf{Healthy}$, in which case we transit to $Y_{i+1} = \mathbf{Healthy}$ with probability μ , or $Y_i = \mathbf{Tasty}$, in which case we transit to $Y_{i+1} = \mathbf{Healthy}$ with probability $1 - \mu$.

$$\propto ((1 - f_i)(1 - \mu) + f_i \mu) p_1 p_1 (1 - p_1) p_1$$

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

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

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).

Solution (Similar to the previous question)

Emission: When $Y_{i+1} = \mathbf{Tasty}$, the probability of observing $X_{i+1} = (1, 1, 1, 0)$ is $p_2 p_2 (1 - p_2) p_2$.

Transition: There are two cases: either $Y_i = \mathbf{Healthy}$, in which case we transit to $Y_{i+1} = \mathbf{Tasty}$ with probability $1 - \mu$, or $Y_i = \mathbf{Tasty}$, in which case we transit to $Y_{i+1} = \mathbf{Tasty}$ with probability μ .

$$\propto ((f_i)(1 - \mu) + (1 - f_i)\mu)p_2 p_2 (1 - p_2) p_2$$

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

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

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

Solution Since h and t have same proportionality constant, we get the true value of the probability by normalization:

$$h/(h + t)$$