CSE 512: Homework 1   Due Wednesday September 11

1. **(2pts)** Alice, Bob, Carlos, Dima, and Estella are all elementary school kids, and the sit in the classroom in a circle. Alice is closest to the teacher, Mrs. Novid, and next to her is Bob, then Carlos, then Dima, then Estella.

   The students are somewhat well-spaced apart, but, you know kids. They like to cough on each other, touch each other's stuff, there's no privacy or social distancing. So disease transmission is indeed an issue.

   Transmission probabilities

   - If Mrs. Novid has Covid, then Alice has a 50% chance of getting it from Mrs. Novid.
   - If Alice has Covid, then Bob has 50% chance of getting Covid from Alice.
   - If Bob has Covid, then Carlos has 50% chance of getting Covid from Bob.
   - If Carlos has Covid, then Dima has 50% chance of getting Covid from Carlos.
   - If Dima has Covid, then Estella has 50% chance of getting Covid from Dima.

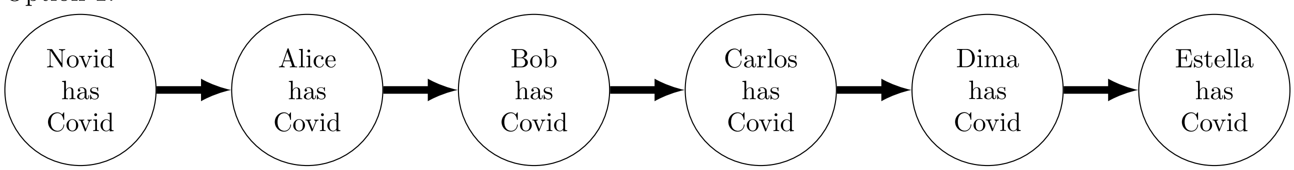
   However, even if no one else in the class has Covid, each child still has a 25% chance of getting Covid, from an outside source.

   (a) Draw a graphical model in which each node stands for a random variable, which represents the probability that a person has Covid. (e.g. node A may represent the event that Alice has Covid, node Bob that Bob has covid, etc...)

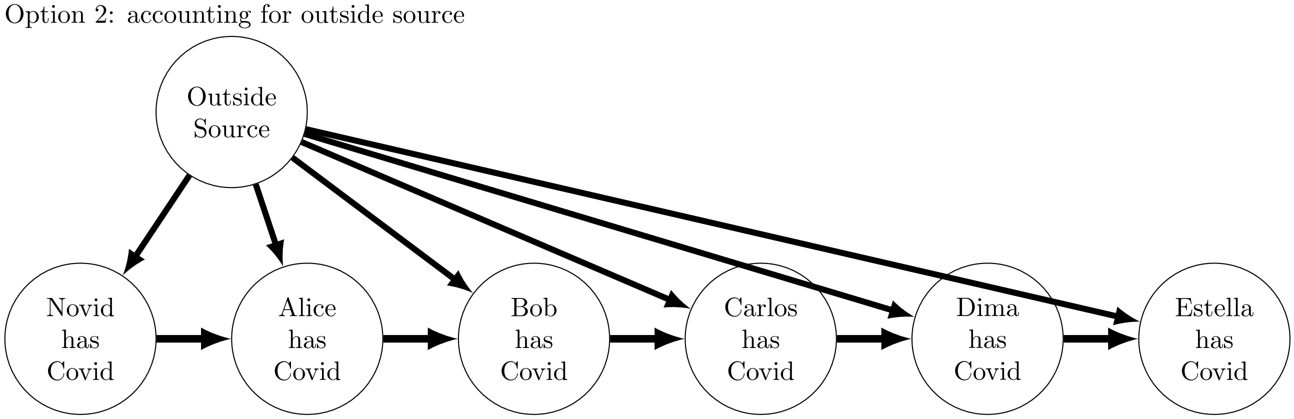
   Ans.

   Either of the following answers will be accepted

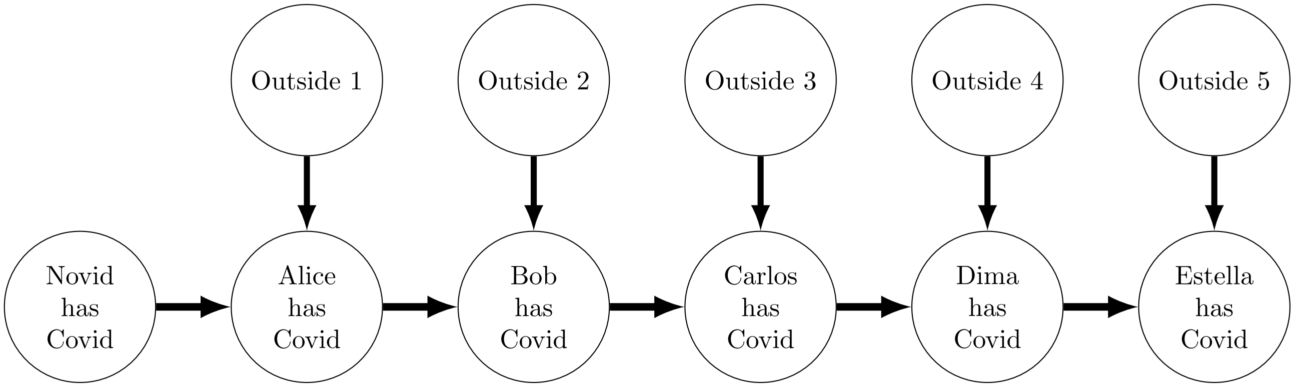
   Option 1:



   Option 2: accounting for outside source



   Option 3: accounting for outside source (this one is the "most correct" one)



Graders, be lenient with the inclusion of the outsider, but make sure the rest of the nodes are written correctly.

(b) We find out one day that Mrs. Novid is home sick, and has had Covid for the past few days. What is the probability that Estella has Covid? Give your answer to the nearest 0.001

Hint: The question is a little easier if you first try to find the probability that each person *doesn't* have Covid.

Ans.

From the chain, we see that the probabilities we need are:

$$\mathbf{Pr}(\text{Novid has Covid}), \mathbf{Pr}(\text{Alice has Covid}|\text{Novid has Covid}), \mathbf{Pr}(\text{Bob has Covid}|\text{Alice has Covid}), \ldots \text{ etc.}$$

Well, from the story, $\mathbf{Pr}(\text{Novid has Covid}) = 1$. The transition probabilities are a bit trickier. I think the easier thing to do here is to try to compute the probability that Alice *doesn't* have Covid, given that Novid has Covid. Let's define some things, or things will get messy.

$$\begin{aligned}
\sim A &= \text{Alice doesn't have Covid} \\
A_1 &= \text{Alice didn't catch Covid from Novid} \\
A_2 &= \text{Alice didn't catch Covid from someone else} \\
N &= \text{Novid has covid}
\end{aligned}$$

$$\mathbf{Pr}(\sim A|N) = \mathbf{Pr}(A_1, A_2|N) = \underbrace{\mathbf{Pr}(A_1|N)}_{0.5} \underbrace{\mathbf{Pr}(A_2|N)}_{=\mathbf{Pr}(A_2)=0.75} = 0.5 \cdot 0.75$$

Therefore, the probability that Alice has Covid given Novid has Covid $= \mathbf{Pr}(A) = 1 - 0.375 = 0.625$.

Bob's probability is a little trickier because it is not for certain that Alice has Covid. So, we have to consider several scenarios. We define the event that Alice has Covid as A. Now, we consider

$$\begin{aligned}
\sim B &= \text{Bob doesn't have Covid} \\
B_1 &= \text{Bob didn't catch Covid from Alice} \\
B_2 &= \text{Bob didn't catch Covid from someone else}
\end{aligned}$$

$$\begin{aligned}
\mathbf{Pr}(\sim B|A) &= \mathbf{Pr}(B_1, B_2|A) = \underbrace{\mathbf{Pr}(B_1|A)}_{0.5}\underbrace{\mathbf{Pr}(B_2)}_{=0.75} = 0.375. \\
\mathbf{Pr}(\sim B| \sim A) &= \mathbf{Pr}(B_1, B_2| \sim A) = \underbrace{\mathbf{Pr}(B_1| \sim A)}_{1}\underbrace{\mathbf{Pr}(B_2)}_{=0.75} = 0.75.
\end{aligned}$$

and now we can compute the probability of $B =$ Bob has Covid:

$$\mathbf{Pr}(\sim B) = \mathbf{Pr}(\sim B|A)\mathbf{Pr}(A) + \mathbf{Pr}(\sim B| \sim A)\mathbf{Pr}(\sim A) = 0.375 \cdot 0.625 + 0.75 \cdot 0.375 = 0.515625$$

and therefore $\mathbf{Pr}(B) = 0.484375$.

Now, we are on a role. Defining C, D, E as the probabilities that Carlos, Dima, and Estella have Covid, respectively, we can work our way down the chain.

$$\begin{aligned}
\mathbf{Pr}(\sim C|B) &= 0.5 \cdot 0.75 = 0.375. \\
\mathbf{Pr}(\sim C| \sim B) &= 1 \cdot 0.75 = 0.75. \\
\mathbf{Pr}(\sim C) &= \mathbf{Pr}(\sim C|B)\mathbf{Pr}(B) + \mathbf{Pr}(\sim C| \sim B)\mathbf{Pr}(\sim B) = 0.5684 \\
\mathbf{Pr}(C) &= 0.4316
\end{aligned}$$

$$\begin{aligned}
\mathbf{Pr}(\sim D|C) &= 0.5 \cdot 0.75 = 0.375. \\
\mathbf{Pr}(\sim D| \sim C) &= 1 \cdot 0.75 = 0.75. \\
\mathbf{Pr}(\sim D) &= \mathbf{Pr}(\sim D|C)\mathbf{Pr}(C) + \mathbf{Pr}(\sim D| \sim C)\mathbf{Pr}(\sim C) = 0.5881 \\
\mathbf{Pr}(D) &= 0.4119
\end{aligned}$$

$$\begin{aligned}
\mathbf{Pr}(\sim E|D) &= 0.5 \cdot 0.75 = 0.375. \\
\mathbf{Pr}(\sim E| \sim D) &= 1 \cdot 0.75 = 0.75. \\
\mathbf{Pr}(\sim E) &= \mathbf{Pr}(\sim E|D)\mathbf{Pr}(D) + \mathbf{Pr}(\sim E| \sim D)\mathbf{Pr}(\sim D) = 0.5956 \\
\mathbf{Pr}(E) &= 0.4044
\end{aligned}$$

So Estella has a 40.4 % chance of having Covid.

2. **(2 pts)** *Probability and statistics.* I have 4 children, Alexa, Siri, Googs, and Zuckie. Every morning I tell them to put on their socks.

   - Alexa only listens to me on Mondays and Thursdays and puts on her socks. The rest of the days, she puts on her socks only half of the time. She either puts on both her socks or none of her socks.
   - Siri always runs and gets her socks, but only puts one sock on.
   - Googs tells me all this random trivia about socks, but never puts on his socks.
   - Zuckie wears both his socks 4/7 of the time and sells the rest of them to CambridgeAnalytica.

   Assume the children all act independently. Round all answers to at least 3 significant digits.

   (a) **(0.5 pts)** What are the chances that either Alexa or Zuckie is wearing at least a sock?
   Ans. This is a classic example of the **intersection rule**

$$\mathbf{Pr}(A \text{ and } B) = \mathbf{Pr}(A) + \mathbf{Pr}(B) - \mathbf{Pr}(A \cap B).$$

Specifically, since

$$\mathbf{Pr}(\text{Alexa is wearing a sock}) = \frac{2}{7} + \frac{1}{2} \cdot \frac{5}{7} = \frac{9}{14}$$

$$\begin{aligned}
\mathbf{Pr}(\text{Alexa or Zuckie is wearing a sock}) &= \mathbf{Pr}(\text{Alexa is wearing a sock}) + \mathbf{Pr}(\text{Zuckie is wearing a sock}) \\
&\quad -\mathbf{Pr}(\text{both are wearing a sock}) \\
&= \frac{9}{14} + \frac{4}{7} - \frac{9}{14} \cdot \frac{4}{7} = \approx 84.7\%
\end{aligned}$$

(b) **(0.5 pt)** On a random day, a girl is wearing at least one sock. What are the chances that Alexa is wearing at least one sock?

Ans. Here, we use Bayes' rule:

$$
\begin{aligned}
& \mathbf{Pr}(\text{Alexa is wearing a sock} \mid \text{a girl is wearing a sock}) \\
= \; & \frac{\mathbf{Pr}(\text{a girl is wearing a sock}|\text{Alexa is wearing a sock}) \cdot \mathbf{Pr}(\text{Alexa is wearing a sock})}{\mathbf{Pr}(\text{a girl is wearing a sock})} \\
= \; & \frac{\mathbf{Pr}(\text{Alexa is wearing a sock})}{\mathbf{Pr}(\text{Alexa is wearing a sock}) + \mathbf{Pr}(\text{Siri is wearing a sock}) - \mathbf{Pr}(\text{Alexa and Siri are both wearing socks})} \\
= \; & \frac{\mathbf{Pr}(\text{Alexa is wearing a sock})}{\mathbf{Pr}(\text{Alexa is wearing a sock}) + \mathbf{Pr}(\text{Siri is wearing a sock}) - \mathbf{Pr}(\text{Alexa is wearing a sock})\mathbf{Pr}(\text{Siri is wearing a sock})} \\
= \; & \frac{\frac{9}{14}}{\frac{9}{14} + 1 - \frac{9}{14} \cdot 1} = \frac{9}{14} \approx 64.3\%
\end{aligned}
$$

(c) **(0.5 pts)** What is the expected number of socks being worn by each child?

Ans. We directly use the expectation definition:

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot \mathbf{Pr}(\text{no socks}) + 1 \cdot \mathbf{Pr}(1 sock) + 2 \cdot \mathbf{Pr}(2 socks).$$

For Alexa,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot \frac{1}{2} \cdot \frac{5}{7} + 1 \cdot 0 + 2 \cdot \frac{9}{14} = \frac{9}{7} \approx 1.29.$$

For Siri,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1$$

For Googs,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot 1 + 1 \cdot 0 + 2 \cdot 0 = 0$$

And for Zuckie,

$$\mathbb{E}[\#\text{ socks}] = 0 \cdot 3/7 + 1 \cdot 0 + 2 \cdot 4/7 = 8/7 \approx 1.14.$$

(d) **(0.5 pts)** What is the variance in the number of socks being worn by each child?

Ans. We directly use the variance definition:

$$\mathbf{var}(\#\text{ socks}) = \mathbb{E}[(\#\text{ socks} - \mathbb{E}[\#\text{ socks}])^2).$$

For Siri and Googs, the mean and values are always the same, and they have 0 variance.

For Alexa,

$$\mathbf{var}(\#\text{ socks}) = (0 - \frac{9}{7})^2 \cdot \frac{5}{14} + (1 - \frac{9}{7})^2 \cdot 0 + (2 - \frac{9}{7})^2 \cdot \frac{9}{14} \approx 0.918.$$

And for Zuckie,

$$\mathbf{var}(\#\text{ socks}) = (0 - \frac{8}{7})^2 \cdot \frac{3}{7} + (1 - \frac{8}{7})^2 \cdot 0 + (2 - \frac{8}{7})^2 \cdot \frac{4}{7} \approx 0.98.$$

3. **(2 pts) Exponential distribution.** Wait time is often modeled as an exponential distribution, e.g.

$$\mathbf{Pr}(\text{I wait less than } x \text{ hours at the DMV}) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

and this cumulative distribution function is parametrized by some constant $\lambda > 0$. A random variable X distributed according to this CDF is denoted as $X \sim \exp(\lambda)$.

   (a) **(0.5 pts)** In terms of λ, give the probability distribution function for the exponential distribution.
   Ans. The PDF can be computed as just the derivative of the CDF, which comes to

$$p_\lambda(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

   (b) **(0.5 pt)** Show that if $X \sim \exp(\lambda)$, then the mean of X is $1/\lambda$ and the variance is $1/\lambda^2$.
   (You may use a symbolic integration tool such as Wolfram Alpha. If you do wish to do the integral by hand, my hint is to review integration by parts.)
   Ans. To compute the mean,

$$\mathbb{E}[X] = \int_0^\infty x p_\lambda(x) dx = \int_0^\infty \lambda x e^{-\lambda x} dx$$

   Now the rest is an exercise in integration. Using Wolfram Alpha,

$$\int_0^\infty \lambda x e^{-\lambda x} dx = \frac{e^{-\lambda x}}{\lambda}(\lambda x - 1)|_0^\infty = 0 - \left(-\frac{1}{\lambda}\right) = 1/\lambda$$

   The same result can be arrived at by using integration by parts.
   To compute the variance, recall that $\mathbf{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Then

$$\mathbb{E}[X^2] = \int_0^\infty x^2 p_\lambda(x) dx = \int_0^\infty \lambda x^2 e^{-\lambda x} dx = \exp(-\lambda x)\left(-\frac{2}{\lambda^2} - \frac{2x}{\lambda} - x^2\right)|_0^\infty = \frac{2}{\lambda^2}$$

$$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

   (c) **(0.5 pt)** Now suppose I run a huge server farm, and I am monitoring the server's ability to respond to web requests. I have m observations of delay times, $x_1, ..., x_m$, which I assume are i.i.d., distributed according to $\exp[\lambda]$ for some λ. Given these m observations, what is the maximum likelihood estimate $\hat{\lambda}$ of λ?
   Ans. First, we compute the likelihood of observations $x_1, ..., x_m$ given that they are i.i.d., distributed as $\exp[\lambda]$:

$$\mathbf{Pr}(x_1, ..., x_m) = \prod_{i=1}^m \mathbf{Pr}(x_i) = \prod_{i=1}^m (\lambda e^{-\lambda x_i}) = \lambda^m \exp\left(-\lambda \sum_{i=1}^m x_i\right).$$

   I would like to find λ which maximizes this quantity. However, this expression looks pretty complicated–not convex or concave.
   Let's use a trick that we are now pretty familiar with: take the log.

$$\log\left(\mathbf{Pr}(x_1, ..., x_m)\right) = m \log(\lambda) - \lambda \sum_{i=1}^m x_i$$

   This is a concave function of λ, so now we can find the maximum of the log probability by taking the derivative and setting it to 0:

$$\frac{\partial}{\partial \lambda} \log\left(\mathbf{Pr}(x_1, ..., x_m)\right) = \frac{m}{\lambda} - \sum_{i=1}^m x_i = 0 \Rightarrow \frac{1}{\hat{\lambda}_{\mathbf{MLE}}} = \frac{1}{m}\sum_{i=1}^m x_i \Rightarrow \hat{\lambda}_{\mathbf{MLE}} = \frac{m}{\sum_{i=1}^m x_i}.$$

(d) **(0.5 pt)** Given the estimate of $\hat{\lambda}$ in your previous question, is $1/\hat{\lambda}$ an unbiased estimate of the mean wait time? Is $1/\hat{\lambda}^2$ an unbiased estimate of the variance in wait time?

Ans. The term $1/\hat{\lambda}$ is indeed an unbiased estimator of the mean:

$$\mathbb{E}\left[\frac{1}{\hat{\lambda}}\right] = \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} x_i\right] = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[x_i] = \frac{1}{\lambda}$$

However, the term $1/\hat{\lambda}^2$ is a biased estimator of the variance:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\hat{\lambda}^2}\right] &= \mathbb{E}\left[\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right)^2\right] \\
&= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\underbrace{\mathbb{E}[x_i x_j]}_{\text{i.i.d.}} \\
&= \frac{1}{m^2}\left(\sum_{i\neq j}\mathbb{E}[x_i]\mathbb{E}[x_j] + \sum_{i=j}\mathbb{E}[x_i^2]\right).
\end{aligned}$$

Noting that $\mathbb{E}[x_i] = \mathbb{E}[x_j] = \frac{1}{\lambda}$ and

$$\mathbb{E}[x_i^2] = \mathbf{var}(x_i) + (\mathbb{E}[x_i])^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}$$

then

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\hat{\lambda}^2}\right] &= \frac{1}{m^2}\left(\sum_{i\neq j}\frac{1}{\lambda^2} + \sum_{i=j}\frac{2}{\lambda^2}\right) \\
&= \frac{1}{m^2}\left((m^2 - m)\frac{1}{\lambda^2} + m\frac{2}{\lambda^2}\right) \\
&= \frac{m+1}{m}\frac{1}{\lambda^2} \neq \frac{1}{\lambda^2}.
\end{aligned}$$

The bias, however, disappears in the limit of $m \to +\infty$.

4. **(2 pts)** *Independent or not independent.* Variables A and B are random variables for two distributions. Decide if A and B are independent. Justify your answer.

(a) **(0.5 pts)** A and B are discrete random variables and have the following p.m.f.s

$$p_A(a) = \begin{cases} 0.25, & a = \text{red} \\ 0.25, & a = \text{blue} \\ 0.5, & a = \text{green} \end{cases}, \qquad p_B(b) = \begin{cases} 0.3, & b = \text{hat} \\ 0.3, & b = \text{T-shirt} \\ 0.2, & b = \text{skirt} \\ 0.2, & b = \text{shoes} \end{cases}$$

and $p_{A,B}(a,b)$ are defined by the table below

	a = red	a = blue	a = green
b = hat	0.075	0.075	0.15
b = T-shirt	0.075	0.075	0.15
b = skirt	0.05	0.05	0.1
b = shoes	0.05	0.05	0.1

Ans. A and B are **independent**. This can be shown by systematically verifying that

$$p_{A,B}(a,b) = p_A(a)p_B(b)$$

for every combination of a and b.

(b) **(1 pt)** A and B are uniform distributions, where

$$f_A(a) = \begin{cases} 1 & -1 \le a \le 0 \\ 0 & \text{else,} \end{cases} \qquad f_B(b) = \begin{cases} 1 & 0 \le b \le 1 \\ 0 & \text{else,} \end{cases},$$

$$f_{A,B}(a,b) = \begin{cases} 4/3 & |a+b| \le 1/2, \quad -1 \le a \le 0, \quad 0 \le b \le 1 \\ 0 & \text{else,} \end{cases}$$

Ans. A and B are **not independent**. For example, taking instantiations $a = -0.1$ and $b = 0.9$,

$$f_A(a) = 1, \quad f_B(b) = 1, \quad f_{A,B}(a,b) = 0 \neq f_A(a) \cdot f_B(b).$$

(c) **(0.5 pts)** A and B are Gaussian distributions, with the following properties:

$$\mathbb{E}[A] = 0, \quad \mathbb{E}[B] = 1, \quad \mathbb{E}[A^2] = 1, \quad \mathbb{E}[(B-1)^2] = 1/2, \quad \mathbb{E}[A(B-1)] = -1.$$

Writing in terms of the usual Gaussian distribution form, if we form a random vector as $X = \begin{bmatrix} A \\ B \end{bmatrix}$, then

$$\mu = \begin{bmatrix} \mathbb{E}[A] \\ \mathbb{E}[B] \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \mathbb{E}[(A-\mathbb{E}[A])^2] & \mathbb{E}[(A-\mathbb{E}[A])(B-\mathbb{E}[B])] \\ \mathbb{E}[(A-\mathbb{E}[A])(B-\mathbb{E}[B])] & \mathbb{E}[(B-\mathbb{E}[B])^2] \end{bmatrix}$$

Ans. A and B are **not independent**. This can be seen by looking at the off-diagonal elements of Σ. In particular,

$$\mathbb{E}[(A-\mathbb{E}[A])(B-\mathbb{E}[B])] = \mathbb{E}[A(B-1)] = -1 \neq 0$$

and therefore Σ is not diagonal.

5. **Naive Bayes. (2pts)** Consider the following dataset 1

Ans.

Feature A	Feature B	Decision (D)	Inference, Naive Bayes	Inference, full Bayes
1	1	1	-1	1
1	-1	-1	1	-1
-1	1	-1	1	-1
-1	-1	1	1	1
1	1	1	-1	1
-1	-1	1	1	1
1	-1	-1	1	-1
-1	-1	1	1	1
-1	1	-1	1	-1
-1	-1	1	1	1

(a) **(0.5 pts)** Using Naive Bayes, fill in the table below

Ans. Using the formula we computed above, and splitting

$$\mathbf{Pr}(A,B|D) = \mathbf{Pr}(A|D)\mathbf{Pr}(B|D) \tag{Naive Bayes}$$

we first compute

$$\begin{aligned}
\mathbf{Pr}(A=1|D=1) &= \frac{1}{3}\\
\mathbf{Pr}(A=1|D=-1) &= \frac{1}{2}\\
\mathbf{Pr}(B=1|D=1) &= \frac{1}{3}\\
\mathbf{Pr}(B=1|D=-1) &= \frac{1}{2}
\end{aligned}$$

and we note that $\mathbf{Pr}(A=-1|D=d) = 1-\mathbf{Pr}(A=1|D=d)$ and $\mathbf{Pr}(B=-1|D=d) = 1-\mathbf{Pr}(B=1|D=d)$ Additionally, we have the prior probabilities

$$\mathbf{Pr}(D=1) = 6/10, \qquad \mathbf{Pr}(D=-1) = 4/10$$

so

$$\begin{aligned}
\mathbf{Pr}(D=1|A=1,B=1) &\propto \frac{1}{3}\cdot\frac{1}{3}\cdot\frac{6}{10} = \frac{1}{15}\\
\mathbf{Pr}(D=1|A=1,B=-1) &\propto \frac{1}{3}\cdot\frac{2}{3}\cdot\frac{6}{10} = \frac{2}{15}\\
\mathbf{Pr}(D=1|A=-1,B=1) &\propto \frac{2}{3}\cdot\frac{1}{3}\cdot\frac{6}{10} = \frac{2}{15}\\
\mathbf{Pr}(D=1|A=-1,B=-1) &\propto \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{6}{10} = \frac{4}{15}\\
\mathbf{Pr}(D=-1|A=1,B=1) &\propto \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{4}{10} = \frac{1}{10}\\
\mathbf{Pr}(D=-1|A=1,B=-1) &\propto \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{4}{10} = \frac{1}{10}\\
\mathbf{Pr}(D=-1|A=-1,B=1) &\propto \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{4}{10} = \frac{1}{10}\\
\mathbf{Pr}(D=-1|A=-1,B=-1) &\propto \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{4}{10} = \frac{1}{10}
\end{aligned}$$

(b) **(0.5 pts)** Using the probabilities you have computed above, compute the Naive Bayes inference of your training data, by filling in the 4th column. Give the training set error rate using this prediction.

Ans. We pick the decision D which maximizes probability given A and B. This leads to an error rate of 60%.

(c) **(0.5 pts)** Using full Bayes, fill in the table below, in terms of $p_{11} = \mathbf{Pr}(A=1, B=1), p_{1,-1} = \mathbf{Pr}(A=1, B=-1), p_{-1,1} = \mathbf{Pr}(A=-1, B=1), p_{-1,-1} = \mathbf{Pr}(A=-1, B=-1)$

$$\mathbf{Pr}(D=1|A=1,B=1) = \qquad \mathbf{Pr}(D=-1|A=1,B=1) =$$

$$\mathbf{Pr}(D=1|A=1,B=-1) = \qquad \mathbf{Pr}(D=-1|A=1,B=-1) =$$

$$\mathbf{Pr}(D=1|A=-1,B=1) = \qquad \mathbf{Pr}(D=-1|A=-1,B=1) =$$

$$\mathbf{Pr}(D=1|A=-1,B=-1) = \qquad \mathbf{Pr}(D=-1|A=-1,B=-1) =$$

Ans. Using the formula we computed above, and splitting

$$\mathbf{Pr}(A,B|D) = \mathbf{Pr}(A|D,B)\mathbf{Pr}(B|D) \tag{full Bayes}$$

so we compute

$$\begin{aligned}
\mathbf{Pr}(A=1|D=1,B=1) &= 1\\
\mathbf{Pr}(A=1|D=1,B=-1) &= 0\\
\mathbf{Pr}(A=1|D=-1,B=1) &= 0\\
\mathbf{Pr}(A=1|D=-1,B=-1) &= 1\\
\mathbf{Pr}(B=1|D=1) &= \frac{2}{3}\\
\mathbf{Pr}(B=1|D=-1) &= \frac{1}{2}
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{Pr}(D=1|A=1,B=1) &\propto 1\cdot\frac{2}{3} = \frac{2}{3}\\
\mathbf{Pr}(D=1|A=1,B=-1) &\propto 0\cdot\frac{1}{3} = 0\\
\mathbf{Pr}(D=1|A=-1,B=1) &\propto 0\cdot\frac{2}{3} = 0\\
\mathbf{Pr}(D=1|A=-1,B=-1) &\propto 1\cdot\frac{1}{3} = \frac{1}{3}\\
\mathbf{Pr}(D=-1|A=1,B=1) &\propto 0\cdot\frac{1}{2} = 0\\
\mathbf{Pr}(D=-1|A=1,B=-1) &\propto 1\cdot\frac{1}{2} = \frac{1}{2}\\
\mathbf{Pr}(D=-1|A=-1,B=1) &\propto 1\cdot\frac{1}{2} = \frac{1}{2}\\
\mathbf{Pr}(D=-1|A=-1,B=-1) &\propto 0\cdot\frac{1}{2} = 0
\end{aligned}$$

(d) **(0.5 pts)** Using the probabilities you have computed above, compute the full Bayes inference of your training data, by filling in the 5th column. Give the training set error rate using this prediction.

Ans. Again, pick the decision D which maximizes probability given A and B. This time, we have a 0% error rate!