Question 1:

You observe that the presence of three specific words—*study*, *free*, and *money*—can help distinguish between spam and regular emails. This insight leads you to model the problem using a Naive Bayes classifier and construct the following dataset. (1 means the word is present in the email, 0 means the word is not present in the email.)

Calculate the probability that an email containing the words—*study=0*, *free=1*, and *money=1* is classified as spam.

Study	Free	Money	Category
1	0	0	Regular
0	0	1	Regular
1	0	0	Regular
1	1	0	Regular
0	1	0	Spam
0	1	0	Spam
0	1	0	Spam
0	1	0	Spam
0	1	1	Spam
0	1	1	Spam
0	1	1	Spam
0	1	1	Spam

C	P(C)	
Spam	8/12	= 2/3
Regular	4/12	= 1/3

C	Free	P(Free|C)
Spam	1	1
Spam	0	0
Regular	1	1/4
Regular	0	3/4

C	Study	P(Study|C)
Spam	1	0
Spam	0	1
Regular	1	3/4
Regular	0	1/4

C	Mny	P(Mny|C)
Spam	1	1/2
Spam	0	1/2
Regular	1	1/4
Regular	0	3/4

P(C | study=0, free=1, money=1) = ______________________________

$$P(C \mid S, F, M)$$

$$= P(C)\cdot P(S\mid C)\cdot P(F\mid C)\cdot P(M\mid C)$$

$$= P(c)\cdot P(-s\mid c)\cdot P(+f\mid c)\cdot P(+m\mid c)$$

$$= \alpha \begin{bmatrix} \overbrace{P(+c)}^{2/3} & \overbrace{P(-s\mid +c)}^{1} & \overbrace{P(+f\mid +c)}^{1} & \overbrace{P(+m\mid +c)}^{1/2} \\ \underbrace{P(-c)}_{1/3} & \underbrace{P(-s\mid -c)}_{1/4} & \underbrace{P(+f\mid -c)}_{1/4} & \underbrace{P(+m\mid -c)}_{1/4} \end{bmatrix} = \alpha \begin{bmatrix} 1/3 \\ 1/192 \end{bmatrix} = \begin{bmatrix} 64/65 \\ 1/65 \end{bmatrix}$$

Question 2: The joint distribution $P(A, B, C, D)$ is equal to the product of these probability distribution tables.

$P(A,B,C,D) = P(A) \cdot P(B|A) \cdot P(C|B) \cdot P(D|C)$

A	$P(A)$
+a	0.8
–a	0.2

A	B	$P(B\|A)$
+a	+b	0.9
+a	–b	0.1
–a	+b	0.6
–a	–b	0.4

B	C	$P(C\|B)$
+b	+c	0.8
+b	–c	0.2
–b	+c	0.8
–b	–c	0.2

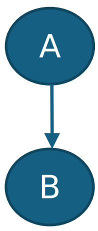
C	D	$P(D\|C)$
+c	+d	0.25
+c	–d	0.75
–c	+d	0.5
–c	–d	0.5

Draw Bayes Net that represent a distribution consistent with the table. Ensure that the network has the **minimal number of edges.**

$A \rightarrow B \rightarrow C \rightarrow D$

C is dependent on B. His solution seems to be wrong.

Question 3: Consider a simple Bayes Net shown below. A and B both can take on only the values true and false (A ∈ {true, false} and B ∈ {true, false}).



A	P(A)
true	1/2
false	1/2

A	B	P(B|A)
true	true	1
true	false	0
false	true	**p**
false	false	1-**p**

$P(A,B) = P(A) \cdot P(B|A)$

$P(B) = \sum_A P(A) \cdot P(B|A)$

$\hookrightarrow P(+a)P(B|+a) + P(-a)P(B|-a)$

$= \frac{1}{2}(1) + \frac{1}{2}(p)$

Find the value of the **p** in the conditional probability table above, if you are told that the P(B=true) = ¾

$= \frac{1}{2} + \frac{1}{2}p = 3/4 \Longrightarrow \frac{1}{2}p = 1/4$

$\boxed{\therefore p = 1/2}$

Question 4: Find values for the probabilities aand bin joint probability table below so that the binary variables X and Y are independent.

X	Y	$P(X,Y)$
t	t	$3/5$
t	f	$1/5$
f	t	a
f	f	b

Independent: $P(Y|X) = P(Y)$.

$3/5 + 1/5 + a + b = 1$ [Total prob must sum to 1]

$\boxed{a + b = 1/5}$ ①

If independent: $P(X,Y) = P(X)\,P(Y)$.

$P(X=t) = 3/5 + 1/5 = 4/5$ $\qquad$ $P(Y=t) = 3/5 + a$

$P(X=f) = a + b = 1/5$ $\qquad$ $P(Y=f) = 1/5 + b$

~~**Question 5:** Aliens can be friendly or not; 75% are friendly. Friendly aliens arrive during the day 90% of the time, while unfriendly ones always arrive at night. If an alien arrives at night, how likely is it to be friendly?~~

$P(X=t, Y=t) = P(X=t)\,P(Y=t) = \left(4/5\right)\left(3/5 + a\right) = \underline{\underline{3/5}}$

$a = 0.15 \longrightarrow \boxed{3/10}$

$3/10 + b = 1/5 \Longrightarrow b = 1/20$

$$\boxed{\begin{aligned} a &= 3/10 \\ b &= 1/20 \end{aligned}}$$

Question 5: Aliens can be friendly or not; 75% are friendly. Friendly aliens arrive during the day 90% of the time, while unfriendly ones always arrive at night. If an alien arrives at night, how likely is it to be friendly?

Alien	P(F)
friendly	0.75
not friendly	0.25

D	F	P(D|F)
+d	+f	0.9
−d	+f	0.1
+d	−f	0
−d	−f	1

$$P(F \mid D) = P(F) \cdot P(D \mid F)$$

$$P(+f \mid -d) \ ?$$

$$\alpha \begin{bmatrix} \overbrace{P(+f)}^{0.75} \ \overbrace{P(-d \mid +f)}^{0.1} \\ \underbrace{P(-f)}_{0.25} \ \underbrace{P(-d \mid -f)}_{1} \end{bmatrix} = \alpha \begin{bmatrix} 0.075 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.2307 \\ 0.7692 \end{bmatrix}$$

$$\Longrightarrow \boxed{P(+f \mid -d) = 0.2307}$$

Laplace smoothing: $\dfrac{\text{count}(x) + \lambda}{N + \lambda|x|}$

Question 6: (Repeat with Laplace smoothing with $\lambda=1$)

You observe that the presence of three specific words—*study*, *free*, and *money*—can help distinguish between spam and regular emails. This insight leads you to model the problem using a Naive Bayes classifier and construct the following dataset. (1 means the word is present in the email, 0 means the word is not present in the email.)

Calculate the probability that an email containing the words—*study=0*, *free=1*, and *money=1* is classified as spam.

Study	Free	Money	Category
1	0	0	Regular
0	0	1	Regular
1	0	0	Regular
1	1	0	Regular
0	1	0	Spam
0	1	0	Spam
0	1	0	Spam
0	1	0	Spam
0	1	1	Spam
0	1	1	Spam
0	1	1	Spam
0	1	1	Spam

C	P(C)
Spam	8/12 → $\frac{8+1}{12+1(2)} = 9/14$
Regular	4/12 → $\frac{4+1}{12+1(2)} = 5/14$

C	Free	P(Free|C)
Spam	1 8/8 →	9/10
Spam	0	1/10
Regular	1 1/4 →	$\frac{1+1}{4+1(2)} = \frac{1}{3}$
Regular	0	2/3

C	Study	P(Study|C)
Spam	1 0/8 →	1/10
Spam	0 8/8 →	9/10
Regular	1 3/4 →	$\frac{4}{6} = 2/3$
Regular	0 1/4 →	$\frac{2}{6} = 1/3$

C	Mny	P(Mny|C)
Spam	1 4/8 →	5/10 = 1/2
Spam	0	1/2
Regular	1 1/4 →	1/3
Regular	0	2/3

P(C | study=0, free=1, money=1) = ____________________

$P(C \mid -s, +f, +m)$

$$= \alpha \begin{bmatrix} \overbrace{P(+c)}^{9/14}\overbrace{P(-s|+c)}^{9/10}\overbrace{P(+f|+c)}^{9/10}\overbrace{P(+m|+c)}^{1/2} \\ \underbrace{P(-c)}_{5/14}\underbrace{P(-s|-c)}_{1/3}\underbrace{P(+f|-c)}_{1/3}\underbrace{P(+m|-c)}_{1/3} \end{bmatrix} = \alpha \begin{bmatrix} 0.26036 \\ 0.013226 \end{bmatrix}$$

$$= \begin{bmatrix} 0.95 \\ 0.05 \end{bmatrix}$$ → answer.