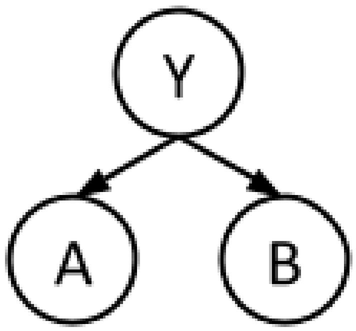
CS 188
Spring 2026

Regular Discussion 9 Solutions

1 Naive Bayes

In this question, we will train a Naive Bayes classifier to predict class labels Y as a function of input features A and B. Y, A, and B are all binary variables, with domains 0 and 1. We are given 10 training points from which we will estimate our distribution.

A	1	1	1	1	0	1	0	1	1	1
B	1	0	0	1	1	1	1	0	1	1
Y	1	1	0	0	0	1	1	0	0	0



1. What are the maximum likelihood estimates for the tables $P(Y)$, $P(A|Y)$, and $P(B|Y)$?

Y	$P(Y)$
0	3/5
1	2/5

A	Y	$P(A\|Y)$
0	0	1/6
1	0	5/6
0	1	1/4
1	1	3/4

B	Y	$P(B\|Y)$
0	0	1/3
1	0	2/3
0	1	1/4
1	1	3/4

2. Consider a new data point ($A = 1$, $B = 1$). What label would this classifier assign to this sample?

$$\begin{aligned}
P(Y=0, A=1, B=1) &= P(Y=0)P(A=1|Y=0)P(B=1|Y=0) && (1)\\
&= (3/5)(5/6)(2/3) && (2)\\
&= 1/3 && (3)\\
P(Y=1, A=1, B=1) &= P(Y=1)P(A=1|Y=1)P(B=1|Y=1) && (4)\\
&= (2/5)(3/4)(3/4) && (5)\\
&= 9/40 && (6)\\
& && (7)
\end{aligned}$$

Our classifier will predict label 0.

3. Let's use Laplace Smoothing to smooth out our distribution. Compute the new distribution for $P(A|Y)$ given Laplace Smoothing with $k = 2$.

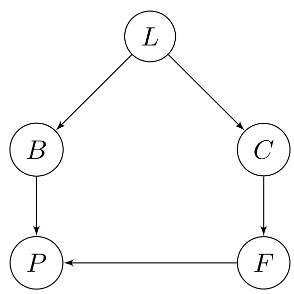
A	Y	$P(A\|Y)$
0	0	3/10
1	0	7/10
0	1	3/8
1	1	5/8

2 Bayes Nets: The CS 188 Project

Consider the following Bayes Net modeling a student's experience working on a CS 188 programming project. All variables are boolean.

- L: Works **L**ate into the night
- B: **B**ugs are introduced into the code
- C: Drinks **C**offee
- F: Maintains high **F**ocus
- P: **P**asses the Gradescope autograder

The structure of the network is as follows: Working late (L) directly influences both whether bugs are introduced (B) and whether the student drinks coffee (C). Drinking coffee (C) directly influences focus levels (F). Finally, both the presence of bugs (B) and the student's focus (F) directly influence whether the project passes the autograder (P).



(a) Write the expression for the joint probability distribution $P(L, B, C, F, P)$ factored according to the Bayes Net structure above. $P(L, B, C, F, P) = P(L)P(B|L)P(C|L)P(F|C)P(P|B, F)$

(b) You are given the following Conditional Probability Tables (CPTs) for the network:

L	$P(L)$
$+l$	0.5
$-l$	0.5

L	B	$P(B|L)$
$+l$	$+b$	0.8
$-l$	$+b$	0.2

L	C	$P(C|L)$
$+l$	$+c$	0.9
$-l$	$+c$	0.5

C	F	$P(F|C)$
$+c$	$+f$	0.7
$-c$	$+f$	0.3

B	F	P	$P(P|B, F)$
$+b$	$+f$	$+p$	0.4
$-b$	$+f$	$+p$	0.9
$+b$	$-f$	$+p$	0.1
$-b$	$-f$	$+p$	0.6

Calculate the probability that the student works late, does **not** introduce bugs, drinks coffee, maintains high focus, and passes the autograder. (i.e. Calculate $P(+l, -b, +c, +f, +p)$).

Using the factorization from the previous part:

$$\begin{aligned}
P(+l, -b, +c, +f, +p) &= P(+l) \cdot P(-b| + l) \cdot P(+c| + l) \cdot P(+f| + c) \cdot P(+p| - b, +f) \\
&= 0.5 \cdot (1 - 0.8) \cdot 0.9 \cdot 0.7 \cdot 0.9 \\
&= 0.5 \cdot 0.2 \cdot 0.9 \cdot 0.7 \cdot 0.9 \\
&= 0.1 \cdot 0.9 \cdot 0.7 \cdot 0.9 \\
&= 0.0567
\end{aligned}$$

(c) (OPTIONAL) We wish to calculate the probability distribution of passing the autograder given that the student stayed up late, $P(P \mid +l)$, using Variable Elimination. The initial factors are: $P(+l)$, $P(B \mid +l)$, $P(C \mid +l)$, $P(F \mid C)$, and $P(P \mid B, F)$.

Assume we eliminate the hidden variables in the order: C**, then** F**, then** B.

For each variable eliminated, write down the sum operation and the resulting new factor.

1. Eliminate C:
2. Eliminate F:
3. Eliminate B:

1. **Eliminate** C: We collect all factors containing C: $P(C \mid +l)$ and $P(F \mid C)$. We multiply them and sum out C.

$$f_1(F, +l) = \sum_c P(c \mid +l) P(F \mid c)$$

2. **Eliminate** F: We collect all factors containing F. This includes our new factor $f_1(F, +l)$ and $P(P \mid B, F)$. We multiply them and sum out F.

$$f_2(P, B, +l) = \sum_f f_1(f, +l) P(P \mid B, f)$$

3. **Eliminate** B: We collect all factors containing B. This includes our new factor $f_2(P, B, +l)$ and $P(B \mid +l)$. We multiply them and sum out B.

$$f_3(P, +l) = \sum_b P(b \mid +l) f_2(P, b, +l)$$

Note: After these steps, to get $P(P \mid +l)$, we would simply normalize $f_3(P, +l)$ (though since we started with the condition $+l$ pushed into the tables, $f_3(P, +l)$ is exactly equal to the joint $P(P, +l)$, which requires normalization by $P(+l)$ to become a conditional distribution).