Assignment #1: Probability Models

General Instructions

- There are two problems. Carefully read all notes and questions.

- Answers should be clearly typed as in a technical report.

- Type out the whole solutions and answers, submit a PDF report on Canvas (or other associated material).

- Scores for rubrics are shown in [ ] within each question.

Problem 1: Naive Bayes for Customer Churn Using Observed Data [50 pts]

Background

You are a data scientist working for a subscription-based company. Your goal is to predict whether a customer will churn ($Y = 1$) or not churn ($Y = 0$) using behavioral data.

You will build two Naive Bayes models, each using a different feature set. For each model, you will:

1. Estimate probabilities from data

2. Apply Laplace smoothing

3. Compute posterior churn probabilities

4. Compare model conclusions

Given (common to both models)

Let Y represent the event churn. From historical data, the overall churn rate is:

- $P(Y = 1) = 0.3$

- $P(Y = 0) = 0.7$

Model A — Support-related features

Feature definitions

- ($X_1 = 1$): Customer opened many support tickets last month
- ($X_2 = 1$): Customer spent a lot of time on help / FAQ pages

Training dataset (historical sample)

Churn (Y)	$X_1 = 1$	$X_1 = 0$	Total
$Y = 1$	24	6	30
$Y = 0$	14	56	70

Churn (Y)	$X_2 = 1$	$X_2 = 0$	Total
$Y = 1$	25	5	30
$Y = 0$	13	57	70

(Note: each feature table is estimated independently, as in Naive Bayes.)

For an observed customer with $X_1 = 1, X_2 = 1$, answer the following questions:

Question A1 — Estimate probabilities

Using **Laplace smoothing**, estimate:

5 pts $P(X_1 = 1 \mid Y = 1)$, $P(X_1 = 1 \mid Y = 0)$

5 pts $P(X_2 = 1 \mid Y = 1)$, $P(X_2 = 1 \mid Y = 0)$

Answer for A1

$P(X_1 = 1 \mid Y = 1) = \frac{24+1}{30+2} = \frac{25}{32} \approx 0.781$

$P(X_1 = 1 \mid Y = 0) = \frac{14+1}{70+2} = \frac{15}{72} \approx 0.208$

$P(X_2 = 1 \mid Y = 1) = \frac{25+1}{30+2} = \frac{26}{32} \approx 0.812$

$P(X_2 = 1 \mid Y = 0) = \frac{13+1}{70+2} = \frac{14}{72} \approx 0.194$

Question A2 — Posterior probabilities

Using your smoothed estimates and Naive Bayes, compute the **normalized posterior probabilities**:

4 pts $P(Y = 1 \mid X_1 = 1, X_2 = 1)$

4 pts $P(Y = 0 \mid X_1 = 1, X_2 = 1)$

Answer for A2

$$P(Y = 1 \mid X_1 = 1, X_2 = 1) = P(Y = 1)P(X_1 = 1 \mid Y = 1)P(X_2 = 1 \mid Y = 1)$$
$$= \frac{3}{10} \cdot \frac{25}{32} \cdot \frac{26}{32} \approx 0.1904$$

$$P(Y = 0 \mid X_1 = 1, X_2 = 1) = P(Y = 0)P(X_1 = 1 \mid Y = 0)P(X_2 = 1 \mid Y = 0)$$
$$= \frac{7}{10} \cdot \frac{15}{72} \cdot \frac{14}{72} \approx 0.0283$$

Then, compute normalized posterior probabilities based on the result above

$P(Y = 1 \mid X_1 = 1, X_2 = 1) = \frac{0.1904}{0.1904 + 0.0283} \approx 0.8703$

$P(Y = 0 \mid X_1 = 1, X_2 = 1) = \frac{0.0283}{0.1904 + 0.0283} \approx 0.1296$

Question A3 — Prediction

Based on Model A:

1 pts Does the model predict churn or no churn?

1 pts What is the predicted probability?

Answer for A3

- The model predicts churn ($Y = 1$), since $P(Y = 1 \mid X_1 = 1, X_2 = 1) > P(Y = 0 \mid X_1 = 1, X_2 = 1)$.

- The predicted probability of churn is $P(Y = 1 \mid X_1 = 1, X_2 = 1) \approx 0.87$.

Model B — Pricing and competition features

Feature definitions

- ($Z_1 = 1$): Customer changed pricing plan last month

- ($Z_2 = 1$): Customer received a competitor's marketing email

Training dataset (historical sample)

Churn (Y)	$Z_1 = 1$	$Z_1 = 0$	Total
$Y = 1$	18	12	30
$Y = 0$	21	49	70

Churn (Y)	$Z_2 = 1$	$Z_2 = 0$	Total
$Y = 1$	15	15	30
$Y = 0$	28	42	70

For an observed customer with $Z_1 = 1, Z_2 = 1$, answer the following questions:

Question B1 — Estimate probabilities

Using **Laplace smoothing**, estimate:

5 pts $P(Z_1 = 1 \mid Y = 1)$, $P(Z_1 = 1 \mid Y = 0)$

5 pts $P(Z_2 = 1 \mid Y = 1)$, $P(Z_2 = 1 \mid Y = 0)$

Answer for B1

$P(Z_1 = 1 \mid Y = 1) = \frac{18+1}{30+2} = \frac{19}{32} \approx 0.593$

$P(Z_1 = 1 \mid Y = 0) = \frac{21+1}{70+2} = \frac{22}{72} \approx 0.305$

$P(Z_2 = 1 \mid Y = 1) = \frac{15+1}{30+2} = \frac{16}{32} = 0.5$

$P(Z_2 = 1 \mid Y = 0) = \frac{28+1}{70+2} = \frac{29}{72} \approx 0.402$

Question B2 — Posterior probabilities

Compute the normalized posterior probabilities:

4 pts $P(Y = 1 \mid Z_1 = 1, Z_2 = 1)$

4 pts $P(Y = 0 \mid Z_1 = 1, Z_2 = 1)$

Answer for B2

$$P(Y = 1 \mid Z_1 = 1, Z_2 = 1) = P(Y = 1)P(Z_1 = 1 \mid Y = 1)P(Z_2 = 1 \mid Y = 1)$$
$$= \frac{3}{10} \cdot \frac{19}{32} \cdot \frac{16}{32} \approx 0.0890$$

$$P(Y = 0 \mid Z_1 = 1, Z_2 = 1) = P(Y = 0)P(Z_1 = 1 \mid Y = 0)P(Z_2 = 1 \mid Y = 0)$$
$$= \frac{7}{10} \cdot \frac{22}{72} \cdot \frac{29}{72} \approx 0.0861$$

Then, compute normalized posterior probabilities based on the result above

$P(Y = 1 \mid Z_1 = 1, Z_2 = 1) = \frac{0.0890}{0.0890+0.0861} \approx 0.5083$

$P(Y = 0 \mid Z_1 = 1, Z_2 = 1) = \frac{0.0861}{0.0890+0.0861} \approx 0.4916$

Question B3 — Prediction

Based on Model B:

1 pts Does the model predict churn or no churn?

1 pts What is the predicted probability?

Answer for B3

- The model predicts churn ($Y = 1$), since $P(Y = 1 \mid Z_1 = 1, Z_2 = 1) > P(Y = 0 \mid Z_1 = 1, Z_2 = 1)$.

- The predicted probability of churn is $P(Y = 1 \mid Z_1 = 1, Z_2 = 1) \approx 0.5083$.

Model comparison and interpretation

Compare the predictions from Model A and Model B:

5 pts Why are the predicted churn probabilities so different?

5 pts Which model's prediction is more trustworthy? Why?

Answer

- The reason why the predicted churn probabilities are so different is because the two models use the different feature with different strengths of association with churn.

- Model A is more trustworthy because the posterior probability is much higher and further from the prior. On the other hand, model B's prediction is close to 0.5, which is less useful to predict.

Problem 2 [50 pts]

Complete the Hidden Markov Model (HMM) Forward/Backtracking table using the following transition and emission probabilities.

Transition

From → To	Sunny	Cloudy	Rainy
Sunny	0.33	0.67	0.00
Cloudy	0.33	0.00	0.67
Rainy	0.33	0.33	0.33

Emission

Weather → Behavior	Walk	Umbrella
Sunny	$4/4 = 1.0$	$0/3 = 0.0$
Cloudy	$2/3 \approx 0.67$	$1/3 \approx 0.33$
Rainy	$1/3 \approx 0.33$	$2/3 \approx 0.67$

You can manually compute (show the steps), code (submit through GitHub), or use a spreadsheet (submit the spreadsheet) to work out the solutions.

Answer for Problem2

I summarized what I did in the file 'HMM.ipynb': `https://github.com/fumiya2001/Math485/blob/main/HW01/HMM.ipynb` by using a table.

Day	Observation	? −>	Sunny	Cloudy	Rainy
		$V0(?)$	0.33	0.33	0.33
1	Walk	$P(W \mid ?)$	1.00	0.67	0.33
		$V1(?) = V0(?)P(W \mid ?)$	0.33	0.22	0.11
2		$V1(S)P(? \mid S)$	0.11	0.22	0.00
		$V1(C)P(? \mid C)$	0.07	0.00	0.15
		$V1(R)P(? \mid R)$	0.04	0.04	0.04
	Umbrella	$P(U \mid ?)$	0	0.33	0.67
		$V2(?) = \max(?)P(U \mid ?)$	0	0.07	0.10
3		$V2(S)P(? \mid S)$	0.00	0.00	0.00
		$V2(C)P(? \mid C)$	0.02	0.00	0.05
		$V2(R)P(? \mid R)$	0.03	0.03	0.03
	Walk	$P(W \mid ?)$	1.00	0.67	0.33
		$V3(?) = \max(?)P(U \mid ?)$	0.03	0.02	0.02