1 Question 4: Naive Bayes Classification (12 points)

1.1 Training Data

Age Group	Income Level	Gender	Previous Purchases	Purchase
Young	High	Female	Yes	**Yes**
Middle-aged	Medium	Male	No	**No**
Senior	Low	Female	Yes	**No**
Young	Medium	Male	Yes	**Yes**
Middle-aged	High	Female	Yes	**Yes**
Senior	Medium	Male	No	**No**
Young	High	Female	No	**Yes**
Middle-aged	Low	Female	No	**No**
Senior	High	Male	Yes	**Yes**
Young	Medium	Male	No	**Yes**

Table 1: Training dataset

New Customer Profile: Age Group = Young, Income Level = High, Gender = Female, Previous Purchases = Yes

1.2 Step 1: Calculate Priors

From the training data:

$$P(\text{Purchase} = \text{Yes}) = \frac{6}{10} = 0.6 \tag{1}$$

$$P(\text{Purchase} = \text{No}) = \frac{4}{10} = 0.4 \tag{2}$$

1.3 Step 2: Calculate Class-Conditional Likelihoods

1.3.1 For Purchase = Yes (6 samples)

$$P(\text{Age Group} = \text{Young} \mid \text{Purchase} = \text{Yes}) = \frac{4}{6} = 0.667 \tag{3}$$

$$P(\text{Income Level} = \text{High} \mid \text{Purchase} = \text{Yes}) = \frac{3}{6} = 0.5 \tag{4}$$

$$P(\text{Gender} = \text{Female} \mid \text{Purchase} = \text{Yes}) = \frac{2}{6} = 0.333 \tag{5}$$

$$P(\text{Previous Purchases} = \text{Yes} \mid \text{Purchase} = \text{Yes}) = \frac{4}{6} = 0.667 \tag{6}$$

Combined likelihood for Purchase = Yes:

$$L(\text{features} \mid \text{Purchase} = \text{Yes}) = 0.667 \times 0.5 \times 0.333 \times 0.667 \tag{7}$$

$$= 0.074 \tag{8}$$

1.3.2 For Purchase = No (4 samples)

$$P(\text{Age Group} = \text{Young} \mid \text{Purchase} = \text{No}) = \frac{0}{4} = 0 \tag{9}$$

$$P(\text{Income Level} = \text{High} \mid \text{Purchase} = \text{No}) = \frac{0}{4} = 0 \tag{10}$$

$$P(\text{Gender} = \text{Female} \mid \text{Purchase} = \text{No}) = \frac{2}{4} = 0.5 \tag{11}$$

$$P(\text{Previous Purchases} = \text{Yes} \mid \text{Purchase} = \text{No}) = \frac{1}{4} = 0.25 \tag{12}$$

Combined likelihood for Purchase = No:

$$L(\text{features} \mid \text{Purchase} = \text{No}) = 0 \times 0 \times 0.5 \times 0.25 \tag{13}$$

$$= 0 \tag{14}$$

1.4 Step 3: Calculate Unnormalized Posterior Scores

$$P(\text{Purchase} = \text{Yes} \mid \text{features}) \propto P(\text{Purchase} = \text{Yes}) \times L(\text{features} \mid \text{Purchase} = \text{Yes}) \tag{15}$$

$$\propto 0.6 \times 0.074 = 0.044 \tag{16}$$

$$P(\text{Purchase} = \text{No} \mid \text{features}) \propto P(\text{Purchase} = \text{No}) \times L(\text{features} \mid \text{Purchase} = \text{No}) \tag{17}$$

$$\propto 0.4 \times 0 = 0 \tag{18}$$

1.5 Step 4: Final Prediction

Since $P(\text{Purchase} = \text{Yes} \mid \text{features}) > P(\text{Purchase} = \text{No} \mid \text{features})$:

Final Prediction: Purchase = Yes

1.6 Laplace Smoothing Check

In this example, some feature combinations for the "No" class have zero counts, which results in zero probability. However, since we can still make a clear prediction, Laplace smoothing is not strictly necessary for this particular case, though it would typically be applied in practice to handle zero probabilities.