Assignment 3

Question 1: Sign up for project teams in the shared document: Working alone.

Question 2: Build a Naive Bayes classifier for the given training data with **add 1 smoothing** technique covered in the lecture slides:

Instance	Education Level	Career	Years of Experience	Salary
1	High School	Management	Less than 3	Low
2	High School	Management	3 to 10	Low
3	College	Management	Less than 3	High
4	College	Service	More than 10	Low
5	High School	Service	3 to 10	Low
6	College	Service	3 to 10	High
7	College	Management	More than 10	High
8	College	Service	Less than 3	Low
9	High School	Management	More than 10	High
10	High School	Service	More than 10	Low

First, note

$$\mathbb{P}(\text{low}) = \frac{6}{10} = \frac{3}{5} \quad \mathbb{P}(\text{high}) = \frac{4}{10} = \frac{2}{5}$$

For the education levels:

$$\begin{aligned}
\mathbb{P}(Y = \text{low} \mid X_1 = \text{high school}) &= \frac{\mathbb{P}(X_1 = \text{high school} \mid Y = \text{low})\mathbb{P}(Y = \text{low})}{\mathbb{P}(X_1 = \text{high school})} \\
&= \frac{\frac{4+1}{6+2} \cdot \frac{3}{5}}{\frac{1}{2}} = \boldsymbol{\frac{3}{4}} \\
\mathbb{P}(Y = \text{high} \mid X_1 = \text{high school}) &= \frac{\mathbb{P}(X_1 = \text{high school} \mid Y = \text{high})\mathbb{P}(Y = \text{high})}{\mathbb{P}(X_1 = \text{high school})} \\
&= \frac{\frac{1+1}{4+2} \cdot \frac{2}{5}}{\frac{1}{2}} = \boldsymbol{\frac{4}{15}} \\
\mathbb{P}(Y = \text{low} \mid X_1 = \text{college}) &= \frac{\mathbb{P}(X_1 = \text{college} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low})}{\mathbb{P}(X_1 = \text{college})} \\
&= \frac{\frac{2+1}{6+2} \cdot \frac{3}{5}}{\frac{1}{2}} = \boldsymbol{\frac{9}{20}} \\
\mathbb{P}(Y = \text{high} \mid X_1 = \text{college}) &= \frac{\mathbb{P}(X_1 = \text{college} \mid Y = \text{high}) \cdot \mathbb{P}(\text{high})}{\mathbb{P}(X_1 = \text{college})} \\
&= \frac{\frac{3+1}{4+2} \cdot \frac{2}{5}}{\frac{1}{2}} = \boldsymbol{\frac{8}{15}}
\end{aligned}$$

For the careers:

$$\begin{aligned}
\mathbb{P}(Y = \text{low} \mid X_2 = \text{management}) &= \frac{\mathbb{P}(X_2 = \text{management} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low})}{\mathbb{P}(X_2 = \text{management})} \\
&= \frac{\frac{2+1}{6+2} \cdot \frac{3}{5}}{\frac{1}{2}} = \boldsymbol{\frac{9}{20}} \\
\mathbb{P}(Y = \text{high} \mid X_2 = \text{management}) &= \frac{\mathbb{P}(X_2 = \text{management} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high})}{\mathbb{P}(X_2 = \text{management})} \\
&= \frac{\frac{3+1}{4+2} \cdot \frac{2}{5}}{\frac{1}{2}} = \boldsymbol{\frac{8}{15}} \\
\mathbb{P}(Y = \text{low} \mid X_2 = \text{service}) &= \frac{\mathbb{P}(X_2 = \text{service} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low})}{\mathbb{P}(X_2 = \text{service})} \\
&= \frac{\frac{4+1}{6+2} \cdot \frac{3}{5}}{\frac{1}{2}} = \boldsymbol{\frac{3}{4}} \\
\mathbb{P}(Y = \text{high} \mid X_2 = \text{service}) &= \frac{\mathbb{P}(X_2 = \text{service} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high})}{\mathbb{P}(X_2 = \text{service})} \\
&= \frac{\frac{1+1}{4+2} \cdot \frac{2}{5}}{\frac{1}{2}} = \boldsymbol{\frac{4}{15}}
\end{aligned}$$

For the levels of experience:

$$\mathbb{P}(Y = \text{low} \mid X_3 = \text{less than 3}) = \frac{\mathbb{P}(X_3 = \text{less than 3} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low})}{\mathbb{P}(X_3 = \text{less than 3})}$$
$$= \frac{\frac{2+1}{6+3} \cdot \frac{3}{5}}{\frac{3}{10}} = \mathbf{\frac{2}{3}}$$
$$\mathbb{P}(Y = \text{high} \mid X_3 = \text{less than 3}) = \frac{\mathbb{P}(X_3 = \text{less than 3} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high})}{\mathbb{P}(X_3 = \text{less than 3})}$$
$$= \frac{\frac{1+1}{4+3} \cdot \frac{2}{5}}{\frac{3}{10}} = \mathbf{\frac{8}{21}}$$
$$\mathbb{P}(Y = \text{low} \mid X_3 = \text{3 to 10}) = \frac{\mathbb{P}(X_3 = \text{3 to 10} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low})}{\mathbb{P}(X_3 = \text{3 to 10})}$$
$$= \frac{\frac{2+1}{6+3} \cdot \frac{3}{5}}{\frac{3}{10}} = \mathbf{\frac{2}{3}}$$
$$\mathbb{P}(Y = \text{high} \mid X_3 = \text{3 to 10}) = \frac{\mathbb{P}(X_3 = \text{3 to 10} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high})}{\mathbb{P}(X_3 = \text{3 to 10})}$$
$$= \frac{\frac{1+1}{4+3} \cdot \frac{2}{5}}{\frac{3}{10}} = \mathbf{\frac{8}{21}}$$
$$\mathbb{P}(Y = \text{low} \mid X_3 = \text{more than 10}) = \frac{\mathbb{P}(X_3 = \text{more than 10} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low})}{\mathbb{P}(X_3 = \text{more than 10})}$$
$$= \frac{\frac{2+1}{6+3} \cdot \frac{3}{5}}{\frac{4}{10}} = \mathbf{\frac{1}{2}}$$
$$\mathbb{P}(Y = \text{high} \mid X_3 = \text{more than 10}) = \frac{\mathbb{P}(X_3 = \text{more than 10} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high})}{\mathbb{P}(X_3 = \text{more than 10})}$$
$$= \frac{\frac{2+1}{4+3} \cdot \frac{2}{5}}{\frac{4}{10}} = \mathbf{\frac{3}{7}}$$

Use your model to classify the following new instances:

Instance	Education Level	Career	Years of Experience
1	High School	Service	Less than 3
2	College	Retail	Less than 3
3	Graduate	Service	3 to 10

Instance 1:

$$\begin{aligned}\mathbb{P}(Y = \text{low} \mid X = \text{hs, s, } < 3) &= \mathbb{P}(X_1 = \text{high school} \mid Y = \text{low}) \cdot \mathbb{P}(X_2 = \text{service} \mid Y = \text{low}) \\ &\quad \cdot \mathbb{P}(X_3 = \text{less than 3} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low}) \\ &= \frac{4+1}{6+2} \cdot \frac{4+1}{6+2} \cdot \frac{2+1}{6+3} \cdot \frac{3}{5} = 0.078 \\ \mathbb{P}(Y = \text{high} \mid X = \text{hs, s, } < 3) &= \mathbb{P}(X_1 = \text{high school} \mid Y = \text{high}) \cdot \mathbb{P}(X_2 = \text{service} \mid Y = \text{high}) \\ &\quad \cdot \mathbb{P}(X_3 = \text{less than 3} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high}) \\ &= \frac{1+1}{4+2} \cdot \frac{1+1}{4+2} \cdot \frac{1+1}{4+3} \cdot \frac{2}{5} = 0.0126\end{aligned}$$

Now:

$$\mathbb{P}(Y = \text{low} \mid X = \text{hs, s, } < 3) > \mathbb{P}(Y = \text{high} \mid X = \text{hs, s, } < 3)$$

The predicted class label for **Instance 1** is: **low**.

Instance 2:

$$\begin{aligned}\mathbb{P}(Y = \text{low} \mid X = \text{college, r, } < 3) &= \mathbb{P}(X_1 = \text{college} \mid Y = \text{low}) \cdot \mathbb{P}(X_2 = \text{retail} \mid Y = \text{low}) \\ &\quad \cdot \mathbb{P}(X_3 = \text{less than 3} \mid Y = \text{low}) \\ &= \frac{2+1}{6+2} \cdot \frac{0+1}{6+3} \cdot \frac{2+1}{6+3} \cdot \frac{3}{5} = 0.008 \\ \mathbb{P}(Y = \text{high} \mid X = \text{college, r, } < 3) &= \mathbb{P}(X_1 = \text{college} \mid Y = \text{high}) \cdot \mathbb{P}(X_2 = \text{retail} \mid Y = \text{high}) \\ &\quad \cdot \mathbb{P}(X_3 = \text{less than 3} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high}) \\ &= \frac{3+1}{4+2} \cdot \frac{0+1}{4+3} \cdot \frac{1+1}{4+3} \cdot \frac{2}{5} = 0.010\end{aligned}$$

Now:

$$\mathbb{P}(Y = \text{low} \mid X = \text{college, r, } < 3) < \mathbb{P}(Y = \text{high} \mid X = \text{college, r, } < 3)$$

The predicted class label for **Instance 2** is: **high**.

Instance 3:

$$\begin{aligned}\mathbb{P}(Y = \text{low} \mid X = \text{grad, s, 3 to 10}) &= \mathbb{P}(X_1 = \text{grad} \mid Y = \text{low}) \cdot \mathbb{P}(X_2 = \text{service} \mid Y = \text{low}) \\ &\quad \cdot \mathbb{P}(X_3 = \text{3 to 10} \mid Y = \text{low}) \cdot \mathbb{P}(Y = \text{low}) \\ &= \frac{0+1}{6+3} \cdot \frac{4+1}{6+2} \cdot \frac{2+1}{6+3} \cdot \frac{3}{5} = 0.013 \\ \mathbb{P}(Y = \text{high} \mid X = \text{grad, s, 3 to 10}) &= \mathbb{P}(X_1 = \text{grad} \mid Y = \text{high}) \cdot \mathbb{P}(X_2 = \text{service} \mid Y = \text{high}) \\ &\quad \cdot \mathbb{P}(X_3 = \text{3 to 10} \mid Y = \text{high}) \cdot \mathbb{P}(Y = \text{high}) \\ &= \frac{0+1}{4+3} \cdot \frac{1+1}{4+2} \cdot \frac{1+1}{4+3} \cdot \frac{2}{5} = 0.005\end{aligned}$$

Now:

$$\mathbb{P}(Y = \text{low} \mid X = \text{grad, s, 3 to 10}) > \mathbb{P}(Y = \text{high} \mid X = \text{grad, s, 3 to 10})$$

The predicted class label for **Instance 3** is: **low**.
For **Questions 3** and **4**, see "Question 3 and 4.pdf."