4033/5033 Assignment: Naive Bayes

Sasank Sribhashyam

In this assignment, we will practice the Naive Bayes classifier in theory. Table 1 contains 10 students, each described by two features 'Course' and 'Hour' and one binary label 'GPA'. For the two features, value '1' means 'yes' and value '0' means 'no'.

Table 1. Student Data Set

Student ID	Take more than 3 courses?	Study more than 2 hour per day?	GPA (A/B)
01	1	1	B
02	1	0	B
03	1	0	B
04	0	0	A
05	1	1	B
06	0	1	A
07	0	1	B
08	1	1	B
09	1	1	A
10	1	0	A

Task 1. Sam is a new student who takes less than 3 courses and studies less than 2 hours per day. Elaborate on how to train a Naive Bayes classifier based on Table 1 and apply it to infer Sam's GPA. In particular, you need to show the estimates of all key probabilities for the NB classifier and explain how to combine them to make inference for Sam.

To calculate the probabilities for the Naive Bayes classifier, let's follow the steps outlined in the previous response:

Step 1: Compute the prior probabilities:

- $P(GPA = A) = \frac{4}{10} = 0.4$
- $P(GPA = B) = \frac{6}{10} = 0.6$

Step 2: Compute the conditional probabilities:

For the feature 'Take more than 3 courses':

- $P(Takemorethan3courses = 0|GPA = A) = \frac{2}{4} = 0.5$
- $P(Takemorethan3courses = 0|GPA = B) = \frac{1}{6} = 0.16$

For the feature 'Study more than 2 hours per day':

- $P(Studymorethan2hoursperday = 0|GPA = A) = \frac{2}{4} = 0.5$

$P(\text{Study more than 2 hours per day} = 0|\text{GPA} = B) = \frac{2}{6} = 0.33$ (rounded to two decimal places).

Step 3: Use the Naive Bayes formula to calculate the posterior probabilities for Sam's features:

- $P(GPA = A|Sam's features) = 0.4 \times 0.5 \times 0.5 = 0.1$
- $P(GPA = B|Sam's features) = 0.6 \times 0.16 \times 0.33 \approx 0.0316$ (rounded to four decimal places)

Step 4: Normalize the probabilities:

Calculate the total probability for both classes:

Total Probability $= P(GPA = A|Sam's features) + P(GPA = B|Sam's features)$
$\approx 0.1 + 0.0316 \approx 0.1316$ (rounded to four decimal places)

Now, normalize the probabilities:

- $\frac{P(GPA=A|Sam's features)}{Total Probability} \approx \frac{0.1}{0.1316} \approx 0.0131$ (rounded to four decimal places)
- $\frac{P(GPA=B|Sam's features)}{Total Probability} \approx \frac{0.0316}{0.1316} \approx 0.00415$ (rounded to four decimal places)

Step 5: Choose the class with the highest probability:

Since $P(GPA = A|Sam's features) > P(GPA = B|Sam's features)$, Sam's GPA is predicted as 'A' with an approximate probability of 0.0131.

So, based on the Naive Bayes classifier, Sam's GPA is predicted to be 'A.'