

## Solutions to Problem 1 of Homework 9 (5 Points)

Name: Anav Prasad (ap7152)

Due: 5PM on Monday, April 11

Collaborators:

Suppose that you are trying to carry out classification learning where C is the classification attribute and the rest are predictive attributes. You are given the following data set shown below.

X, Y, Z are predictive attributes, with values 1 and 2 and C is the classification attribute, with values 1, 2, and 3. The null value indicates nothing was recorded for that attribute. “#” is not an attribute, but rather the count of the occurrences in the training set. (imagine there are 10 lines in the training set with (1,1,1,1))

X	Y	Z	C	#
1	1	1	1	10
1	1	1	2	8
1	1	null	3	6
1	1	2	1	5
1	1	2	3	15
2	1	2	2	12
2	1	2	3	5
2	2	1	2	8
2	2	1	3	10
2	2	1	1	5
2	2	1	2	6
null	2	1	3	10

A. (5 points) How does Naive Bayes classify the instance  $X=1, Y=2, Z=1$ ? Do not use the Laplacian correction.

**Solution:**

To figure out the classification using Naive Bayes, we need to compute the following probabilities:

$$\left. \begin{aligned} P(C = 1 \mid X = 1, Y = 2, Z = 1) \\ P(C = 2 \mid X = 1, Y = 2, Z = 1) \\ P(C = 3 \mid X = 1, Y = 2, Z = 1) \end{aligned} \right\} (1)$$

So, before computing the above, let's compute the frequencies and auxiliary probabilities first:

$$\begin{aligned} \text{Freq}(C = 1) &= 10 + 5 + 5 = 20 \\ \text{Freq}(C = 2) &= 8 + 12 + 8 + 6 = 34 \\ \text{Freq}(C = 3) &= 6 + 15 + 5 + 10 + 10 = 46 \\ \text{Total} &= 20 + 34 + 46 = 100 \end{aligned}$$

$$\begin{aligned}
P(C = 1) &= \frac{Freq(C = 1)}{Total} \\
&= \frac{10 + 5 + 5}{100} \\
\therefore P(C = 1) &= \frac{1}{5} = 0.2 \\
P(C = 2) &= \frac{Freq(C = 2)}{Total} \\
&= \frac{8 + 12 + 8 + 6}{100} \\
\therefore P(C = 2) &= \frac{17}{50} = 0.34 \\
P(C = 3) &= \frac{Freq(C = 3)}{Total} \\
&= \frac{6 + 15 + 5 + 10 + 10}{100} \\
\therefore P(C = 3) &= \frac{23}{50} = 0.46 \\
P(X = 1 \mid C = 1) &= \frac{10 + 5}{20} = \frac{3}{4} = 0.75 \\
P(X = 1 \mid C = 2) &= \frac{8}{34} = \frac{4}{17} \approx 0.235 \\
P(X = 1 \mid C = 3) &= \frac{6 + 15}{46 - 10} = \frac{7}{12} = 0.58\bar{3} \\
P(Y = 2 \mid C = 1) &= \frac{5}{20} = \frac{1}{4} = 0.25 \\
P(Y = 2 \mid C = 2) &= \frac{8 + 6}{34} = \frac{7}{17} \approx 0.412 \\
P(Y = 2 \mid C = 3) &= \frac{10 + 10}{46} = \frac{10}{23} \approx 0.435 \\
P(Z = 1 \mid C = 1) &= \frac{10 + 5}{20} = \frac{3}{4} = 0.75 \\
P(Z = 1 \mid C = 2) &= \frac{8 + 8 + 6}{34} = \frac{11}{17} \approx 0.647 \\
P(Z = 1 \mid C = 3) &= \frac{10 + 10}{46 - 6} = \frac{1}{2} = 0.5
\end{aligned}$$

Now, let's compute the probabilities in (1):

$$\begin{aligned}
P(C = 1 \mid X = 1, Y = 2, Z = 1) &= P(C = 1) \cdot P(X = 1 \mid C = 1) \cdot P(Y = 2 \mid C = 1) \cdot P(Z = 1 \mid C = 1) \\
&= \frac{1}{5} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \\
\therefore P(C = 1 \mid X = 1, Y = 2, Z = 1) &= \frac{9}{320} = 0.028125
\end{aligned}$$

$$\begin{aligned}
P(C = 2 \mid X = 1, Y = 2, Z = 1) &= P(C = 2) \cdot P(X = 1 \mid C = 2) \cdot P(Y = 2 \mid C = 2) \cdot P(Z = 1 \mid C = 2) \\
&= \frac{17}{50} \cdot \frac{4}{17} \cdot \frac{7}{17} \cdot \frac{11}{17} \\
\therefore P(C = 2 \mid X = 1, Y = 2, Z = 1) &= \frac{154}{7225} \approx 0.0213 \\
P(C = 3 \mid X = 1, Y = 2, Z = 1) &= P(C = 3) \cdot P(X = 1 \mid C = 3) \cdot P(Y = 2 \mid C = 3) \cdot P(Z = 1 \mid C = 3) \\
&= \frac{23}{50} \cdot \frac{7}{12} \cdot \frac{10}{23} \cdot \frac{1}{2} \\
\therefore P(C = 3 \mid X = 1, Y = 2, Z = 1) &= \frac{7}{120} = 0.058\bar{3}
\end{aligned}$$

Therefore,

$$\operatorname{argmax}_C \{P(C \mid X = 1, Y = 2, Z = 1)\} = 3$$

Thus, the Naive Bayes Classifier would classify the given instance of  $X = 1, Y = 2, Z = 1$  as  $C = 3$ .  $\square$

## Solutions to Problem 2 of Homework 9 (5 Points)

Name: Anav Prasad (ap7152)

Due: 5PM on Monday, April 11

Collaborators:

B. (5 points) Using the Laplacian correction with  $\delta = 1$ , how does Naive Bayes classify the instance  $X=1, Y=2, Z=2$ ?

**Solution:**

Here, the auxiliary values and probabilities for the case of Naive Bayes Classification with Laplacian correction is as follows:

$$\begin{aligned}
 \delta &= 1 \\
 q &= q_X = q_Y = q_Z = 2 \\
 \text{Freq}(C = 1) &= 20 && \text{(Computed in Problem 1)} \\
 \text{Freq}(C = 2) &= 34 && \text{(Computed in Problem 1)} \\
 \text{Freq}(C = 3) &= 46 && \text{(Computed in Problem 1)} \\
 \text{Total} &= 100 && \text{(Computed in Problem 1)} \\
 P(C = 1) &= \frac{1}{5} = 0.2 && \text{(Computed in Problem 1)} \\
 P(C = 2) &= \frac{17}{50} = 0.34 && \text{(Computed in Problem 1)} \\
 P(C = 3) &= \frac{23}{50} = 0.46 && \text{(Computed in Problem 1)} \\
 P(X = 1 \mid C = 1) &= \frac{10 + 5 + \delta}{20 + q \cdot \delta} = \frac{16}{22} = \frac{8}{11} = 0.72\overline{7} \\
 P(X = 1 \mid C = 2) &= \frac{8 + \delta}{34 + q \cdot \delta} = \frac{9}{36} = \frac{1}{4} = 0.25 \\
 P(X = 1 \mid C = 3) &= \frac{6 + 15 + \delta}{46 - 10 + q \cdot \delta} = \frac{22}{38} = \frac{11}{19} \approx 0.579 \\
 P(Y = 2 \mid C = 1) &= \frac{5 + \delta}{20 + q \cdot \delta} = \frac{6}{22} = \frac{3}{11} = 0.27\overline{2} \\
 P(Y = 2 \mid C = 2) &= \frac{8 + 6 + \delta}{34 + q \cdot \delta} = \frac{15}{36} = \frac{5}{12} = 0.41\overline{6} \\
 P(Y = 2 \mid C = 3) &= \frac{10 + 10 + \delta}{46 + q \cdot \delta} = \frac{21}{48} = \frac{7}{16} = 0.4375 \\
 P(Z = 2 \mid C = 1) &= \frac{5 + \delta}{20 + q \cdot \delta} = \frac{6}{22} = \frac{3}{11} = 0.27\overline{2} \\
 P(Z = 2 \mid C = 2) &= \frac{12 + \delta}{34 + q \cdot \delta} = \frac{13}{36} = 0.36\overline{1} \\
 P(Z = 2 \mid C = 3) &= \frac{15 + 5 + \delta}{46 - 6 + q \cdot \delta} = \frac{21}{42} = \frac{1}{2} = 0.5
 \end{aligned}$$

Now, let's compute the probabilities for the classification:

$$\begin{aligned}
P(C = 1 \mid X = 1, Y = 2, Z = 2) &= P(C = 1) \cdot P(X = 1 \mid C = 1) \cdot P(Y = 2 \mid C = 1) \cdot P(Z = 2 \mid C = 1) \\
&= \frac{1}{5} \cdot \frac{8}{11} \cdot \frac{3}{11} \cdot \frac{3}{11} \\
\therefore P(C = 1 \mid X = 1, Y = 2, Z = 2) &= \frac{72}{6655} \approx 0.0108 \\
P(C = 2 \mid X = 1, Y = 2, Z = 2) &= P(C = 2) \cdot P(X = 1 \mid C = 2) \cdot P(Y = 2 \mid C = 2) \cdot P(Z = 2 \mid C = 2) \\
&= \frac{17}{50} \cdot \frac{1}{4} \cdot \frac{5}{12} \cdot \frac{13}{36} \\
\therefore P(C = 2 \mid X = 1, Y = 2, Z = 2) &= \frac{221}{17280} \approx 0.0128 \\
P(C = 3 \mid X = 1, Y = 2, Z = 2) &= P(C = 3) \cdot P(X = 1 \mid C = 3) \cdot P(Y = 2 \mid C = 3) \cdot P(Z = 2 \mid C = 3) \\
&= \frac{23}{50} \cdot \frac{11}{19} \cdot \frac{7}{16} \cdot \frac{1}{2} \\
\therefore P(C = 3 \mid X = 1, Y = 2, Z = 2) &= \frac{1771}{30400} \approx 0.0583
\end{aligned}$$

Therefore,

$$\operatorname{argmax}_C \{P(C \mid X = 1, Y = 2, Z = 2)\} = 3$$

Thus, the Naive Bayes Classifier with Laplacian Correction would classify the given instance of  $X = 1, Y = 2, Z = 2$  as  $C = 3$ .  $\square$