Assignment 2: Naive Bayes and Text Classification

Benedikt Riegel

May 8, 2021

1 Task 1: Simple Bayes (20 Points)

1.1 Task (10 Points)

Define:

- $P(\text{Box x}) :=$ probability that Box x is chosen)
- $P(\text{Apple}) :=$ probability that an apple is chosen
- $P(\text{Orange}) :=$ probability that an orange is chosen

Given: $P(\text{Box 1}) = P(\text{Box 2})$

	Apples	Oranges
Box 1	4	10
Box 2	6	8

We assume $P(\text{Box 1}) + P(\text{Box 2}) = 1$.
$\implies P(\text{Apple} \mid \text{Box 1}) = 4/14 = 2/7$, $P(\text{Orange} \mid \text{Box 1}) = 10/14 = 5/7$,
$P(\text{Apple} \mid \text{Box 2}) = 6/14 = 3/7$ and $P(\text{Orange} \mid \text{Box 2}) = 8/14 = 4/7$.
$P(\text{Box 1}) = P(\text{Box 2}) = 1/2$.

What is the probability of choosing an apple?

$$\begin{aligned}
P(\text{Apple}) &= P(\text{Apple, Box 1}) + P(\text{Apple, Box 2}) \\
&= P(\text{Box 1}) \cdot P(\text{Apple} \mid \text{Box 1}) + P(\text{Box 2}) \cdot P(\text{Apple} \mid \text{Box 2}) \\
&= \frac{1}{2} \cdot \frac{2}{7} + \frac{1}{2} \cdot \frac{3}{7} \\
&= \frac{5}{14}
\end{aligned}$$

If an apple is chosen, what is the probability that it came from box 1?

$$\begin{aligned} P(\text{Box 1} \mid \text{Apple}) &= \frac{P(\text{Box 1, Apple})}{P(\text{Apple})} \\ &= \frac{P(\text{Apple, Box 1})}{P(\text{Apple})} \\ &= \frac{\frac{1}{2} \cdot \frac{2}{7}}{\frac{5}{14}} \\ &= \frac{1 \cdot 2 \cdot 14}{2 \cdot 7 \cdot 5} \\ &= \frac{28}{70} \\ &= \frac{2}{5} \end{aligned}$$

1.2 Task (10 Points)

Given: Given are two M&M bags from 1994 and 1996 and the probabilities of finding a specific colour in the two different bags. The probabilities are as follows:

	Yellow	Green	Other
1994	0.3	0.2	0.5
1996	0.16	0.24	0.6

Now scenario A :="one M&M of each bag is taken out, one is green and the other is yellow." occurs.
What is the probability of scenario B := "the yellow M&M came from the 1994 bag"? So it's asked for the probability

$$P(B \mid A) = \frac{P(B, A)}{P(A)}$$

First take a look at $P(B, A)$, notice that A and B implies that not only is the 1994 M&M yellow, but the 1996 M&M is green.

$$\begin{aligned} P(B, A) &= P(\text{1994 Bag was picked}) * P(\text{yellow M\&M}) \\ &\quad + P(\text{1996 Bag was picked}) * P(\text{green M\&M}) \\ &= 0.5 * 0.3 + 0.5 * 0.24 \\ &= 0.27 \end{aligned}$$

Notice that because of statement A we have to take one out of each bag, so we have to treat the probabilities for each bag to be 0.5. Now $P(A)$ can also be computed by summing up all the allowed combinations of M&M that fulfil scenario A.

A.1 1994 M&M is yellow $\implies$ 1996 M&M is green.

A.2 1996 M&M is green $\implies$ 1996 M&M is yellow.

This means there are only scenario $A.1$ and $A.2$ that fulfil A.

$$[A.1 \Leftrightarrow B \text{ and } A] \implies [P(A.1) = P(B, A) = 0.27]$$

Analogously $P(A.2)$ can be computed to be

$$\begin{aligned} P(A.2) &= 0.5 * 0.2 + 0.5 * 0.16 \\ &= 0.18 \end{aligned}$$

resulting in

$$\begin{aligned} P(A) &= P(A.1) + P(A.2) \\ &= 0.27 + 0.18 \\ &= 0.45 \end{aligned}$$

Therefore the answer to our previous question is

$$\begin{aligned} P(B \mid A) &= \frac{P(B, A)}{P(A)} \\ &= \frac{0.27}{0.45} \\ &= 0.6 \end{aligned}$$