Introduction to Machine Learning - HW2 - Q2

Professors: Abolghasemi & Arabi
Student: Mohamad Mahdi Samadi
Student ID: 810101465

Q2. Given a training dataset, we're going to use Naive Bayes to predict some new unseen data.

Training dataset is:

f_1	f_2	f_3	y
0	0	1	0
0	1	0	0
1	1	0	0
0	0	1	0
1	1	1	1
1	0	0	1
1	1	0	1

And the test dataset is:

f_1	f_2	f_3
0	0	0
1	1	1
X	1	0
X	0	1

First, let's calculate the prior probabilities.

$P(y) = \frac{3}{7}$ and $P(\bar{y}) = \frac{4}{7}$

Next step is to calculate the conditional probability of each feature given a target class.

y	f_1	f_2	f_3
$y = 0$	$f_1 \rightarrow \frac{3}{4}$ and $\bar{f}_1 \rightarrow \frac{1}{4}$	$f_2 \rightarrow \frac{2}{4}$ and $\bar{f}_2 \rightarrow \frac{2}{4}$	$f_3 \rightarrow \frac{2}{4}$ and $\bar{f}_3 \rightarrow \frac{2}{4}$
$y = 1$	$f_1 \rightarrow \frac{4}{4}$ and $\bar{f}_1 \rightarrow \frac{0}{4}$	$f_2 \rightarrow \frac{3}{4}$ and $\bar{f}_2 \rightarrow \frac{1}{4}$	$f_3 \rightarrow \frac{2}{4}$ and $\bar{f}_3 \rightarrow \frac{2}{4}$

We want to know $P(y|F)$ where F is the vector of features. Since it is hard to calculate and we only want to know which class is more probable, we aim to calculate $P(y, F)$ for each of the two classes and compare them. The Bayes Law shows why the two of these are equivalent:

$$P(y|F) = \frac{P(F|y)P(y)}{P(F)} \rightarrow A = \frac{P(y|F)}{P(\bar{y}|F)} = \frac{\frac{P(F|y)P(y)}{P(F)}}{\frac{P(F|\bar{y})P(\bar{y})}{P(F)}} = \frac{P(F|y)P(y)}{P(F|\bar{y})P(\bar{y})}$$

$P(F)$ is the same for two probabilities of $P(y|F)$ and $P(\bar{y}|F)$. So we can compare the $P(F|y)P(y)$ term.

- $A > 1$: class $y = 1$ is more probable than the other class.
- $A < 1$: class $y = 0$ is more probable than the other class.
- $A = 1$: Two classes have the same probability.

The simplifying assumption is that all features are independent of each other. So using chain rule of probabilities we can rewrite the $P(F|y)$ term.

$$P(F|y) = P(f_1|y)P(f_2\,|\,f_1, y)P(f_3\,|\,f_1, f_2, y) = P(f_1\,|\,y)\,P(f_2\,|\,y)\,P(f_3\,|\,y)$$

$$\frac{P(y)\,P(f_1\,|\,y)\,P(f_2\,|\,y)\,P(f_3\,|\,y)}{P(\bar{y})\,P(f_1\,|\,\bar{y})\,P(f_2\,|\,\bar{y})\,P(f_3\,|\,\bar{y})} = \frac{P(y)}{P(\bar{y})} \times \frac{P(f_1|y)}{P(f_1|\bar{y}}) \times \frac{P(f_2|y)}{P(f_2|\bar{y}}) \times \frac{P(f_3|y)}{P(f_3|\bar{y})}$$

$i.\ F = \{0,\ 0,\ 0\}$

$$\frac{P(y)}{P(\bar{y})} \times \frac{P(f_1|y)}{P(f_1|\bar{y}}) \times \frac{P(f_2|y)}{P(f_2|\bar{y}}) \times \frac{P(f_3|y)}{P(f_3|\bar{y})} = \frac{3}{4} \times \frac{0}{1} \times \frac{1}{2} \times \frac{2}{2} = 0 < 1 \rightarrow \hat{y} = 0$$

$ii.\ F = \{1,\ 1,\ 1\}$

$$\frac{P(y)}{P(\bar{y})} \times \frac{P(f_1|y)}{P(f_1|\bar{y}}) \times \frac{P(f_2|y)}{P(f_2|\bar{y}}) \times \frac{P(f_3|y)}{P(f_3|\bar{y})} = \frac{3}{4} \times \frac{4}{3} \times \frac{3}{2} \times \frac{2}{2} = 1.5 > 1 \rightarrow \hat{y} = 1$$

$iii.\ F = \{X,\ 1,\ 0\}$

The value of f_1 is missing. So we'll assume that $P(f_1|y) = P(f_1|\bar{y})$.

$$\frac{P(y)}{P(\bar{y})} \times \frac{P(f_1|y)}{P(f_1|\bar{y}}) \times \frac{P(f_2|y)}{P(f_2|\bar{y}}) \times \frac{P(f_3|y)}{P(f_3|\bar{y})} = \frac{3}{4} \times 1 \times \frac{3}{2} \times \frac{2}{2} = \frac{18}{16} > 1 \rightarrow \hat{y} = 1$$

$iv.\ F = \{X,\ 0,\ 1\}$

Again, the value of f_1 is missing. So we'll make the same assumption.

$$\frac{P(y)}{P(\bar{y})} \times \frac{P(f_1|y)}{P(f_1|\bar{y})} \times \frac{P(f_2|y)}{P(f_2|\bar{y})} \times \frac{P(f_3|y)}{P(f_3|\bar{y})} = \frac{3}{4} \times 1 \times \frac{1}{2} \times \frac{2}{2} = \frac{6}{16} < 1 \rightarrow \hat{y} = 0$$

	$F = \{0, 0, 0\}$	$F = \{1, 1, 1\}$	$F = \{X, 1, 0\}$	$F = \{X, 0, 1\}$
$\hat{y}$	0	1	1	0