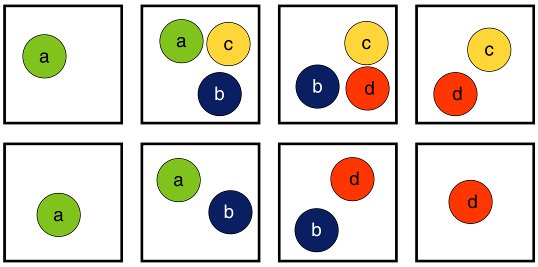
Bayesian Classification

Exercise 1 : Probability Basics (Conditional Independence)

There are eight boxes containing different colored balls as shown in the illustration below:



The balls can be green, blue, yellow, or red (also marked a, b, c, d in the figure). When picking one of the eight boxes at random, let A refer to the event "box contains a green ball," B to the event "box contains a blue ball," C to the event "box contains a yellow ball," and D to the event "box contains a red ball." Hence, $A \cap B$ is the event "box contains both a green and a blue ball," etc.

(a) Calculate $P(A)$, $P(B)$, $P(C)$, and $P(D)$.

Answer

$P(A) = P(B) = P(D) = \frac{4}{8} = \frac{1}{2}$; $P(C) = \frac{3}{8}$.

(b) Calculate $P(A \cap B)$, $P(A \cap C)$, $P(B \cap C)$, and $P(B \cap D)$.

Answer

$P(A \cap B) = P(B \cap C) = P(B \cap D) = \frac{2}{8} = \frac{1}{4}$, $P(A \cap C) = \frac{1}{8}$.

(c) Check all that apply:

[X] The events A and B are statistically independent.
[ ] The events A and C are statistically independent.
[ ] The events B and C are statistically independent.
[X] The events B and D are statistically independent.

Answer

A and B are independent because $P(A)P(B) = P(A \cap B)$ (or, equivalently, $P(A) = P(A \mid B)$). The same holds for B and D.

(d) Calculate $P(A \mid C)$, $P(B \mid C)$, and $P(A \cap B \mid C)$.

Answer

$P(A \mid C) = \frac{1}{3}$, $P(B \mid C) = \frac{2}{3}$, $P(A \cap B \mid C) = \frac{1}{3}$.

(e) Calculate $P(B \mid D)$, $P(C \mid D)$, and $P(B \cap C \mid D)$

Answer

$P(B \mid D) = \frac{2}{4} = \frac{1}{2}$, $P(C \mid D) = \frac{2}{4} = \frac{1}{2}$, $P(B \cap C \mid D) = \frac{1}{4}$.

(f) Check all that apply:

☐ The events A and B are conditionally independent given C.
☒ The events B and C are conditionally independent given D.

Answer

Since $P(B \mid D)P(C \mid D) = P(B \cap C \mid D)$, the events B and C are conditionally independent given D.

Exercise 2 : Bayes' Rule

A hospital database contains diagnoses ($C_1 \ldots C_5$) for 8 patients along with binary observations of symptoms $S_1 \ldots S_9$:

Patient	**Diagnosis**	**Symptoms**								
		S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
1	C_1	1	0	1	0	1	0	0	0	0
2	C_2	0	1	0	1	1	0	1	0	0
3	C_3	1	0	1	0	0	1	0	1	0
4	C_4	0	1	0	1	1	0	1	0	0
5	C_3	1	0	1	0	0	0	0	1	0
6	C_5	0	0	0	0	1	0	0	0	1
7	C_3	1	0	1	0	0	1	0	0	0
8	C_2	0	1	0	0	0	0	1	0	0

(a) Compute based on the database the prior probabilities $P(C_i)$ for each diagnosis.

Answer

$$
\begin{aligned}
P(C_1) &= \tfrac{1}{8} = 0.125 \\
P(C_2) &= \tfrac{2}{8} = \tfrac{1}{4} = 0.25 \\
P(C_3) &= \tfrac{3}{8} = 0.375 \\
P(C_4) &= \tfrac{1}{8} = 0.125 \\
P(C_5) &= \tfrac{1}{8} = 0.125
\end{aligned}
$$

(b) Compute based on the database the posterior probabilities $P(C_i \mid S_4)$ for each diagnosis.

Answer

First, compute $P(S_4|C_i)$, $i = 1, \ldots, 5$:

$$
\begin{aligned}
P(S_4|C_1) &= \tfrac{P(S_4 \cap C_1)}{P(C_1)} = 0.0 \\
P(S_4|C_2) &= \tfrac{P(S_4 \cap C_2)}{P(C_2)} = 0.5 \\
P(S_4|C_3) &= \tfrac{P(S_4 \cap C_3)}{P(C_3)} = 0.0 \\
P(S_4|C_4) &= \tfrac{P(S_4 \cap C_4)}{P(C_4)} = 1.0 \\
P(S_4|C_5) &= \tfrac{P(S_4 \cap C_5)}{P(C_5)} = 0.0
\end{aligned}
$$

Then, compute the a-posteriori probabilities $P(C_i|S_4)$, $i = 1, \ldots, 5$:

$$\begin{aligned}
P(C_1|S_4) &= \frac{P(C_1)\cdot P(S_4|C_1)}{\sum_{j=1}^{5} P(C_j)\cdot P(S_4|C_j)} = \frac{\frac{1}{8}\cdot 0}{\frac{1}{8}\cdot 0+\frac{1}{4}\cdot\frac{1}{2}+\frac{3}{8}\cdot 0+\frac{1}{8}\cdot 1+\frac{1}{8}\cdot 0} = \frac{0}{\frac{2}{8}} = 0 \\
P(C_2|S_4) &= \frac{P(C_2)\cdot P(S_4|C_2)}{\sum_{j=1}^{5} P(C_j)\cdot P(S_4|C_j)} = \frac{\frac{1}{4}\cdot\frac{1}{2}}{\frac{1}{8}\cdot 0+\frac{1}{4}\cdot\frac{1}{2}+\frac{3}{8}\cdot 0+\frac{1}{8}\cdot 1+\frac{1}{8}\cdot 0} = \frac{\frac{1}{8}}{\frac{2}{8}} = \frac{1}{2} \\
P(C_3|S_4) &= \frac{P(C_3)\cdot P(S_4|C_3)}{\sum_{j=1}^{5} P(C_j)\cdot P(S_4|C_j)} = \frac{\frac{3}{8}\cdot 0}{\frac{1}{8}\cdot 0+\frac{1}{4}\cdot\frac{1}{2}+\frac{3}{8}\cdot 0+\frac{1}{8}\cdot 1+\frac{1}{8}\cdot 0} = \frac{0}{\frac{2}{8}} = 0 \\
P(C_4|S_4) &= \frac{P(C_4)\cdot P(S_4|C_4)}{\sum_{j=1}^{5} P(C_j)\cdot P(S_4|C_j)} = \frac{\frac{1}{8}\cdot 1}{\frac{1}{8}\cdot 0+\frac{1}{4}\cdot\frac{1}{2}+\frac{3}{8}\cdot 0+\frac{1}{8}\cdot 1+\frac{1}{8}\cdot 0} = \frac{\frac{1}{8}}{\frac{2}{8}} = \frac{1}{2} \\
P(C_5|S_4) &= \frac{P(C_5)\cdot P(S_4|C_5)}{\sum_{j=1}^{5} P(C_j)\cdot P(S_4|C_j)} = \frac{\frac{1}{8}\cdot 0}{\frac{1}{8}\cdot 0+\frac{1}{4}\cdot\frac{1}{2}+\frac{3}{8}\cdot 0+\frac{1}{8}\cdot 1+\frac{1}{8}\cdot 0} = \frac{0}{\frac{2}{8}} = 0
\end{aligned}$$

Exercise 3 : Naïve Bayes

Given is the following dataset to classify whether a dog is dangerous or well-behaved in character:

Color	**Fur**	**Size**	**Character** (C)
brown	ragged	small	well-behaved
black	ragged	big	dangerous
black	smooth	big	dangerous
black	curly	small	well-behaved
white	curly	small	well-behaved
white	smooth	small	dangerous
red	ragged	big	well-behaved

(a) Determine the parameters $P(A_i)$ and $P(B_{j=x_j} \mid A_i)$ for a Naïve Bayes classifier on this dataset.

Answer

Class priors:

$$\begin{aligned}
P(\mathit{C}{=}\text{well-behaved}) &= \tfrac{4}{7} \\
P(\mathit{C}{=}\text{dangerous}) &= \tfrac{3}{7}
\end{aligned}$$

Attribute-value probabilities given class:

Attribute	Value (x_j)	Class (c_i)	$P(B_{j=x_j} \mid A_i)$
Color	brown	well-behaved	$\frac{1}{4} = 0.25$
Color	brown	dangerous	0.00
Color	black	well-behaved	$\frac{1}{4} = 0.25$
Color	black	dangerous	$\frac{2}{3} = 0.6\overline{6}$
Color	white	well-behaved	$\frac{1}{4} = 0.25$
Color	white	dangerous	$\frac{1}{3} = 0.3\overline{3}$
Color	red	well-behaved	$\frac{1}{4} = 0.25$
Color	red	dangerous	0.00
Fur	ragged	well-behaved	$\frac{1}{2} = 0.50$
Fur	ragged	dangerous	$\frac{1}{3} = 0.3\overline{3}$
Fur	smooth	well-behaved	0.00
Fur	smooth	dangerous	$\frac{2}{3} = 0.6\overline{6}$
Fur	curly	well-behaved	$\frac{1}{2} = 0.50$
Fur	curly	dangerous	0.00
Size	small	well-behaved	$\frac{3}{4} = 0.75$
Size	small	dangerous	$\frac{1}{3} = 0.3\overline{3}$
Size	big	well-behaved	$\frac{1}{4} = 0.25$
Size	big	dangerous	$\frac{2}{3} = 0.6\overline{6}$

(b) Classify the new example $\mathbf{x} = (\mathsf{black}, \mathsf{ragged}, \mathsf{small})$ using your Naïve Bayes classifier.

Answer

For reduced verbosity, let the following events be defined: $A_1 = (\mathsf{C}{=}\mathsf{well\text{-}behaved})$, $A_2 = (\mathsf{C}{=}\mathsf{dangerous})$, $B_1 = (\mathsf{C}{=}\mathsf{black})$, $B_2 = (\mathsf{C}{=}\mathsf{ragged})$, $B_3 = (\mathsf{C}{=}\mathsf{small})$.

With the Naïve Bayes assumption we have:

$$
\begin{aligned}
P(A_1 \mid B_1, B_2, B_3) &= \frac{P(A_1) \cdot P(B_1, B_2, B_3 \mid A_1)}{P(B_1, B_2, B_3)} \\
&\overset{\mathsf{NB}}{\approx} \frac{P(A_1) \cdot \prod_{j=1}^{3} P(B_j \mid A_1)}{\sum_{i=1}^{2} P(A_i) \prod_{j=1}^{3} P(B_j \mid A_i)}
\end{aligned}
$$

and equivalently for A_2; The denominator in both cases is:

$(P(A_1) \cdot P(B_1 \mid A_1) \cdot P(B_2 \mid A_1) \cdot P(B_3 \mid A_1)) +$
$(P(A_2) \cdot P(B_1 \mid A_2) \cdot P(B_2 \mid A_2) \cdot P(B_3 \mid A_2)) = \left(\frac{4}{7} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3}{4}\right) + \left(\frac{3}{7} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}\right) = \frac{3}{56} + \frac{2}{63} =$
$\frac{43}{504} \approx 0.085$

Hence, we get:

$P(A_1 \mid B_1, B_2, B_3) \approx \frac{\frac{4}{7} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3}{4}}{\frac{43}{504}} = \frac{27}{43} \approx 0.628$

and

$P(A_2 \mid B_1, B_2, B_3) \approx \frac{\frac{3}{7} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}}{\frac{43}{504}} = \frac{16}{43} \approx 0.372$

Thus, A_1 is more likely under the Naïve Bayes assumption.