Exercise: 1

Meeting on December 17th

Task 1: Probabilites in "The Settlers of Catan"

In the board game "The Settlers of Catan" you roll two dice at the same time. The outcome, i.e. their sum, either randomly decides the distribution of commodities (2-6 and 8-12) or changing the position of the robber (7).

(a) What is the probability of rolling a 6 (fraction number)? Explain the meaning of denominator and numerator!

(b) What is the probability of rolling a 2 or 3?

(c) What is the probability of rolling a 6 if the last roll was a pair of threes?

(d) Why should you prefer the numbers 5, 6, 8 and 9 in this board game?

Lösung:

(a) There are 5 different results out of a possible 36 which result in a 6 (1-5, 5-1, 2-4, 4-2, 3-3):

$$P(6) = \frac{5}{36}$$

(b)

$$P(2 \cup 3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$$

(c) Still $\frac{5}{36}$, since the two events are completely independent.

(d) Higher probability of gaining commodities.

Task 2: Bayes' theorem und other formulae

(a) What is Bayes' theorem for the conditional probability $P(Y|X)$? Derive the theorem from the definition of conditional probability.

(b) Show that $P(A,B|C) = P(C|A)\frac{P(A)P(B)}{P(C)}$, if A and B are independent and both B and C are independent under condition A.

(c) Let D be dependent on C and B, and B dependent on A. All other combinations are independent of one another. Show:

$$P(A,B,C,D) = P(A)P(B|A)P(C)P(D|B,C)$$

(d) Let all S_i be independent of each other by condition D. What is the value of the normalization constant α for

$$P(D|S_1, S_2,S_3) = \alpha P(D)P(S_1|D)P(S_2|D)P(S_3|D) \text{ ?}$$

Lösung:

(a)

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad \Rightarrow \quad P(X \cap Y) = P(X|Y) \cdot P(Y)$$
$$\text{mit} \quad P(X \cap Y) = P(Y \cap X)$$
$$\text{folgt} \quad P(X|Y) \cdot P(Y) = P(Y|X) \cdot P(X)$$
$$\textbf{Bayes:} \quad \Rightarrow P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

(b)

$$P(A,B|C) \overset{\text{Bayes}}{=} \frac{P(C|B,A)P(B,A)}{P(C)} = \frac{P(C|A)P(B,A)}{P(C)} = \frac{P(C|A)P(A)P(B)}{P(C)}$$

(c)

$$\begin{aligned} P(A,B,C,D) &= \frac{P(A)}{P(A)}\frac{P(A,B)}{P(A,B)}\frac{P(A,B,C)}{P(A,B,C)}P(A,B,C,D) \\ &= P(A)\frac{P(A,B)}{P(A)}\frac{P(A,B,C)}{P(A,B)}\frac{P(A,B,C,D)}{P(A,B,C)} \\ &= P(A)P(B|A)P(C|A,B)P(D|A,B,C) \\ &= P(A)P(B|A)P(C)P(D|B,C) \end{aligned}$$

(d)

$$P(D|S_1, S_2,S_3) = \frac{P(D,S_1,S_2,S_3)}{P(S_1, S_2, S_3)}$$

analogous to (c):

$$= \frac{P(D)P(S_1|D)P(S_2|D,S_1)P(S_3|D,S_1,S_2)}{P(S_1, S_2, S_3)} = \frac{P(D)P(S_1|D)P(S_2|D)P(S_3|D)}{P(S_1, S_2, S_3)}$$
$$= \alpha P(D)P(S_1|D)P(S_2|D)P(S_3|D) \quad \text{with} \quad \alpha = \frac{1}{P(S_1, S_2, S_3)}$$

Task 3: Normalization of probabilities

Let $A = \{a_1, a_2, a_3, a_4\}$ and $B = \{b_1, b_2, b_3\}$, with all a_i and all b_i being elementary events. Given are $P(a_1|b_1) = 0.1$, $P(a_1|b_2) = 0.2$, $P(a_2|b_1) = 0.2$, and $P(a_4|b_1) = 0.4$.

(a) What is the value of $P(a_3|b_1)$?

(b) What is the value of $P(b_1|b_2)$?

(c) What is the value of $P(a_1|b_1 \cap b_2)$?

(d) What is the value of $P(a_1 \cup a_2 | b_1)$?

Lösung:

(a) $P(a_3|b_1) = 1 - (P(a_1|b_1) + P(a_2|b_1) + P(a_4|b_1)) = 0.3$

(b) $P(b_1|b_2) = 0$
since B cannot be in multiple states at the same time.

(c) You cannot make a statement about $P(a_1|b_1 \cap b_2)$ since:

$$P(a_1|b_1 \cap b_2) = \frac{P(a_1 \cap b_1 \cap b_2)}{P(b_1 \cap b_2)} = \frac{0}{0}$$

is not defined. This is similar to "$\Rightarrow$", since you can infer anything from a false statement.

(d) $P(a_1 \cup a_2|b_1) = \frac{P((a_1 \cup a_2) \cap b_1)}{P(b_1)} = \frac{P((a_1 \cap b_1) \cup (a_2 \cap b_1))}{P(b_1)}$
all a_i are independent of each other since they are elementary events:
$= \frac{P(a_1 \cap b_1) + P(a_2 \cap b_1)}{P(b_1)} = \frac{P(a_1 \cap b_1)}{P(b_1)} + \frac{P(a_2 \cap b_1)}{P(b_1)} = P(a_1|b_1) + P(a_2|b_1) = 0.3$

Task 4: Probability table

You are analyzing part of speech in English texts. You know from a corpus analysis that 99% of verbs in gerund form end on “ing“. A total of 2.3% of all words end on “ing“. Around 1.5% of all words are verbs in gerund form.

(a) Building on that, what is the probability of a word ending on “ing“ if it is a verb in gerund form?

(b) What is the probability of a word both ending on “ing“ and being in gerund form? Are the two classes independent of each other?

(c) Write down the probability table of the given problem.

Lösung: Class “Word is verb in gerund form“: $W \in \{VG, \overline{VG}\}$
Class “Word ending in -ing“: $E \in \{I, \overline{I}\}$
Given are the probabilities $P(E = I|W = VG) = 99\%$, $P(E = I) = 2{,}3\%$ and $P(W = VG) = 1{,}5\%$.

a) We are looking for $P(W = VG|E = I)$. Using Bayes' theorem we get:

$$\begin{aligned} P(W = VG|E = I) &= \frac{P(E = I|W = VG) \cdot P(W = VG)}{P(E = I)} \\ &= \frac{99\% \cdot 1{,}5\%}{2{,}3\%} \\ &\approx 64{,}6\% \end{aligned}$$

b) We are looking for $P(W = VG \cap E = I)$. Using $P(A \cap B) = P(A|B) \cdot P(B)$ we get:

$$\begin{aligned} P(W = VG \cap E = I) &= P(W = VG|E = I) \cdot P(E = I) \\ &= 64{,}6\% \cdot 2{,}3\% \\ &\approx 1{,}49\% \end{aligned}$$

The classes $W = VG$ und $E = I$ are independet of each other since

$$\begin{aligned} P(W = VG) \cdot P(E = I) &= 1{,}5\% \cdot 2{,}3\% = 0{,}0345\% \\ &\neq 1{,}49\% = P(W = VG \cap E = I) \end{aligned}$$

c) Using the results from b) we get:

	VG	$\overline{VG}$	
I	1,49%	0,81%	2,3%
$\overline{I}$	0,01%	97,69%	97,7%
	1,5%	98,5%	100%

Task 5: Naive Bayes

Given are the five events S_1 bis S_5. You want to classify event S_5 from a given table of events using the probabilistic Naive Bayes model.

(a) What is a classification?

(b) What are the assumptions on S_1 bis S_4 made by a Naive Bayes model?

(c) Here are data about wheather events. Do the Naive Bayes assumptions apply here in general?

S_1: Outlook	S_2: Temperature	S_3: Humidity	S_4: Windy	S_5: Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

(d) Compute the probability of “Play=yes“ under „sunny, cool, high, true“. What is the classification result?

Lösung:

(a) Assigning a class with given data.

(b) Conditional independence of S_1 to S_4 with known S_5.

(c) Not in general. Even if I know that I can “play“ today, temperature and humidity are still certainly dependent on each other.

(d) Naive Bayes:

$$P(S_5|S_1,S_2,S_3,S_4) = \alpha P(S_5)P(S_1|S_5)P(S_2|S_5)P(S_3|S_5)P(S_4|S_5)$$
$$P(\text{no}|\text{sunny,cool,high,true}) = ?$$
$$P(\text{yes}|\text{sunny,cool,high,true}) = ?$$

From the table: 14 test cases, 5 times ”no” and 9 times ”yes”, so $P(\text{no}) = \frac{5}{14}$ and $P(\text{yes}) = \frac{9}{14}$.

„Outlook = sunny“ appears 3 times at “Play = no“ und 2 times at „Play = yes“, so $P(\text{sunny}|\text{no}) = \frac{3}{5}$ and $P(\text{sunny}|\text{yes}) = \frac{2}{9}$. The approach is analogous for the other variables:

$$P(\text{no}|\text{sunny,cool,high,true}) = \alpha \cdot \frac{5}{14} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} = \alpha \cdot 0.0206$$
$$P(\text{yes}|\text{sunny,cool,high,true}) = \alpha \cdot \frac{9}{14} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} = \alpha \cdot 0.0053$$

Finally: $P(\text{yes}) = 0.0053/(0.0053 + 0.0206) = 20.5\%$ and $P(\text{no}) = 79.5\%$,
Result of Classification: “no“.