**INTRODUCTION MACHINE LEARNING**

**EXERCISE 6**

A red sign with white text

Description automatically generated

|  |
| --- |
| **TEACHER:** |
| Johannes Kiesel |

|  |
| --- |
| **GROUP:** |
| Group 16 |
|  |
| **SUBMITTED BY:** |
| Aaron Perez Herrera |
| Cesar Fernando Gamba Tiusaba |
| Chun Ting Lin |
| Olubunmi Emmanuel Ogunleye |

Exercise 1 : Probability Basics (1 Points)

Which of the following statements are true?

* According to the Kolmogorov axioms the statement P(A) - P(A) = 0 holds.
* A function that fulfills the Kolmogorov axioms is a probability measure.
* Two events are statistically independent P(An B) = P(A) + P(B).
* Each subset A of a sample space 12 is an event.

The true statements are.

* A function that fulfills the Kolmogorov axioms is a probability measure. (true)
* Each subset A of a sample space 1 is an event. (true)

Exercise 3 : Bayes' Rule (2+3=5 Points)

A hospital database contains diagnoses (C1 ... C5) for 8 patients along with binary observations of symptoms S1 ... S9:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Patient** | **Diagnosis** | **Symptoms** | | | | | | | | |
| **S1** | **S2** | **S3** | **S4** | **S5** | **S6** | **S7** | **S8** | **S9** |
| **1** | **C1** | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| **2** | **C2** | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| **3** | **C3** | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| **4** | **C4** | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| **5** | **C3** | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| **6** | **C5** | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| **7** | **C3** | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| **8** | **C2** | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

1. Compute based on the database the prior probabilities P(Ci) for each diagnosis.

The total number of patients is 8.

* .The number of patients diagnosed with C1 is 1, so P(C1) = 1/8 = 0.125.
* The number of patients diagnosed with C2 is 2, so P(C2) = 2/8 = 0.25.
* The number of patients diagnosed with C3 is 3, so P(C3) = 3/8 = 0.375.
* The number of patients diagnosed with C4 is 1, so P(C4) = 1/8 =0.125.
* The number of patients diagnosed with C5 is 1, so P(C5) = 1/8 = 0.125.

1. Compute based on the database the posterior probabilities P(Ci | S4) for each diagnosis.

Posterior probabilities P(Ci|S4) of the diagnoses Ci given symptom S4 are calculated as follows:

The posterior probability of Ai given B is given by,

Hence, first P(B|Ai) is to be calculated.

Therefore, the posterior probability of S4 given C is calculated as:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

The next probabilities

Exercise 4

Part (a): Calculate P(A), P(B), P(C), and P(D):

Each event corresponds to the probability of a specific color of balls (green, blue, yellow, or red) being

in one of the eight boxes. Using the image provided:

Green balls (a) appear in 3 boxes.

Blue balls (b) appear in 4 boxes.

Yellow balls (c) appear in 2 boxes.

Red balls (d) appear in 4 boxes.

The probabilities are calculated as:

A math equations with numbers

AI-generated content may be incorrect.

Part (b): Calculate P(A and B), P(A and C), P(B and C), and P(B and D):

The joint probabilities are calculated by counting the boxes containing both colors:

A math equations with numbers

AI-generated content may be incorrect.

Part (c): Check for statistical independence:

Pair (A, B):

P(A) = 0.375, P(B) = 0.5, P(A and B) = 0.125 P(A) × P(B) = 0.375 × 0.5 = 0.1875

Since P(A and B) = 0.125 != 0.1875, A and B are not independent.

Pair (A, C):

P(A) = 0.375, P(C) = 0.25, P(A and C) = 0.125 P(A) × P(C) = 0.375 × 0.25 = 0.09375

Since P(A and C) = 0.125 != 0.09375, A and C are not independent.

Pair (B, C):

P(B) = 0.5, P(C) = 0.25, P(B and C) = 0.125 P(B) × P(C) = 0.5 × 0.25 = 0.125

Since P(B and C) = 0.125 = 0.125, B and C are independent.

Pair (B, D):

P(B) = 0.5, P(D) = 0.5, P(B and D) = 0.25 P(B) × P(D) = 0.5 × 0.5 = 0.25

Since P(B and D) = 0.25 = 0.25, B and D are independent.

Part (d): Calculate P(A|C), P(B|C), and P(A and B|C):

Conditional probabilities are calculated using the formula:

A black text on a white background

AI-generated content may be incorrect.

From the data:

A math equations with numbers

AI-generated content may be incorrect.

Part (e): Calculate P(B|D), P(C|D), and P(B and C|D):

Similarly, conditional probabilities given D are:

A black text on a white background

AI-generated content may be incorrect.

From the data:

A math equations with numbers

AI-generated content may be incorrect.

Part (f): Check for conditional independence:

Two events A and B are conditionally independent given C if:

A black text with a white background

AI-generated content may be incorrect.

From (d), A number and exclamation mark

AI-generated content may be incorrect.Thus, A and B are not conditionally independent given C