
TRABALHO DE MATEMÁTICA

Entropia de uma Base de dados 2.0pt
Alexandre Garcia de Oliveira

Author

Henrique Rodrigues de Freitas
FATEC - Rubens Lara
31/05/2023

Contents

1	Exercícios	3
1.1	Descrição da Base de dados	3
1.2	Resultados	3
1.3	Quantas Classes	3
1.3.1	Resposta	3
1.4	Quais Classes	3
1.4.1	Resposta	3
2	Execução	3
2.1	Entropia	3
2.1.1	Resposta	3
2.2	Max Entropia	5
2.2.1	Resposta	5
2.2.2	Programação	5
2.3	Código	5

1 Exercícios

1.1 Descrição da Base de dados

1.2 Resultados

A base de dados escolhida foi a +Milionária da Lotérica da Caixa, os dados escolhidos e lidos foram todos os números das bolas vencedoras do 1º ao 10º sorteio.

CAMPO: Campo números bolas vencedoras

1.3 Quantas Classes

1.3.1 Resposta

Temos 1 Classe para cada número das bolas vencedoras, então temos: 60 classes

1.4 Quais Classes

1.4.1 Resposta

Classe do número das bolas vencedoras: [1, 3, 7, 15, 23, 44, 13, 16, 35, 41, 42, 47, 1, 9, 17, 30, 31, 44, 6, 23, 25, 33, 34, 47, 6, 16, 21, 24, 26, 45, 1, 19, 22, 32, 39, 45, 9, 12, 35, 44, 47, 48, 1, 4, 5, 16, 38, 50, 6, 11, 12, 14, 15, 18, 4, 6, 10, 42, 47, 48]

Lembrando que cada Classe equivale a um número da bola: Então temos 60 classes para calcularmos.

2 Execução

2.1 Entropia

2.1.1 Resposta

A formula da entropia é:

$$H = -\sum_{x \in \text{Classes}} p(x) * \log_2(p(x))$$

Os dados fornecidos são:

[1, 3, 7, 15, 23, 44, 13, 16, 35, 41, 42, 47, 1, 9, 17, 30, 31, 44, 6, 23, 25, 33, 34, 47, 6, 16, 21, 24, 26, 45, 1, 19, 22, 32, 39, 45, 9, 12, 35, 44, 47, 48, 1, 4, 5, 16, 38, 50, 6, 11, 12, 14, 15, 18, 4, 6, 10, 42, 47, 48]

Ao contar o número de elementos nessa lista, temos um total de 60 números.

Para calcular a entropia dos dados, precisamos determinar a probabilidade de ocorrência de cada valor único na sequência e, em seguida, usar essas probabilidades para calcular a entropia.

Contar a frequência de cada valor na sequência: O valor 1 ocorre 4 vezes. O valor 3 ocorre 1 vez. O valor 4 ocorre 3 vezes. O valor 5 ocorre 1 vez. O valor 6 ocorre 4 vezes. O valor 7 ocorre 1 vez. O valor 9 ocorre 2 vezes. O valor 10 ocorre 1 vez. O valor 11 ocorre 1 vez. O valor 12 ocorre 2 vezes. O valor 13 ocorre 1 vez. O valor 14 ocorre 1 vez. O valor

15 ocorre 2 vezes. O valor 16 ocorre 4 vezes. O valor 17 ocorre 2 vezes. O valor 18 ocorre 1 vez. O valor 19 ocorre 1 vez. O valor 21 ocorre 1 vez. O valor 22 ocorre 1 vez. O valor 23 ocorre 2 vezes. O valor 24 ocorre 1 vez. O valor 25 ocorre 1 vez. O valor 26 ocorre 1 vez. O valor 30 ocorre 1 vez. O valor 31 ocorre 1 vez. O valor 32 ocorre 1 vez. O valor 33 ocorre 1 vez. O valor 34 ocorre 1 vez. O valor 35 ocorre 2 vezes. O valor 38 ocorre 1 vez. O valor 39 ocorre 1 vez. O valor 41 ocorre 1 vez. O valor 42 ocorre 2 vezes. O valor 44 ocorre 3 vezes. O valor 45 ocorre 2 vezes. O valor 47 ocorre 4 vezes. O valor 48 ocorre 2 vezes. O valor 50 ocorre 1 vez. Calcular a probabilidade de ocorrência de cada valor, dividindo a frequência pelo número total de dígitos (60 no caso): A probabilidade do valor 1 é $4/60 = 0.0667$. A probabilidade do valor 3 é $1/60 = 0.0167$. A probabilidade do valor 4 é $3/60 = 0.05$. A probabilidade do valor 5 é $1/60 = 0.0167$. A probabilidade do valor 6 é $4/60 = 0.0667$. A probabilidade do valor 7 é $1/60 = 0.0167$. A probabilidade do valor 9 é $2/60 = 0.0333$. A probabilidade do valor 10 é $1/60 = 0.0167$. A probabilidade do valor 11 é $1/60 = 0.0167$. A probabilidade do valor 12 é $2/60 = 0.0333$. A probabilidade do valor 13 é $1/60 = 0.0167$. A probabilidade do valor 14 é $1/60 = 0.0167$. A probabilidade do valor 15 é $2/60 = 0.0333$. A probabilidade do valor 16 é $4/60 = 0.0667$. A probabilidade do valor 17 é $2/60 = 0.0333$. A probabilidade do valor 18 é $1/60 = 0.0167$. A probabilidade do valor 19 é $1/60 = 0.0167$. A probabilidade do valor 21 é $1/60 = 0.0167$. A probabilidade do valor 22 é $1/60 = 0.0167$. A probabilidade do valor 23 é $2/60 = 0.0333$. A probabilidade do valor 24 é $1/60 = 0.0167$. A probabilidade do valor 25 é $1/60 = 0.0167$. A probabilidade do valor 26 é $1/60 = 0.0167$. A probabilidade do valor 30 é $1/60 = 0.0167$. A probabilidade do valor 31 é $1/60 = 0.0167$. A probabilidade do valor 32 é $1/60 = 0.0167$. A probabilidade do valor 33 é $1/60 = 0.0167$. A probabilidade do valor 34 é $1/60 = 0.0167$. A probabilidade do valor 35 é $2/60 = 0.0333$. A probabilidade do valor 38 é $1/60 = 0.0167$. A probabilidade do valor 39 é $1/60 = 0.0167$. A probabilidade do valor 41 é $1/60 = 0.0167$. A probabilidade do valor 42 é $2/60 = 0.0333$. A probabilidade do valor 44 é $3/60 = 0.05$. A probabilidade do valor 45 é $2/60 = 0.0333$. A probabilidade do valor 47 é $4/60 = 0.0667$. A probabilidade do valor 48 é $2/60 = 0.0333$. A probabilidade do valor 50 é $1/60 = 0.0167$. Calcular a entropia usando a fórmula da entropia:

$$\begin{aligned} \text{Entropia} = & -(0.0667 * \log_2(0.0667) + 0.0167 * \log_2(0.0167) + 0.05 * \log_2(0.05) + 0.0167 \\ & * \log_2(0.0167) + 0.0667 * \log_2(0.0667) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) \\ & + 0.0167 * \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0167 * \\ & \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0667 * \log_2(0.0667) \\ & + 0.0333 * \log_2(0.0333) + 0.0167 * \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0167 * \\ & \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0167 * \log_2(0.0167) \\ & + 0.0167 * \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0167 * \\ & \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0167 * \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0167 * \\ & \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.05 * \log_2(0.05) + 0.0333 * \log_2(0.0333) + 0.0667 \\ & * \log_2(0.0667) + 0.0333 * \log_2(0.0333) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) \\ & + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0333 * \log_2(0.0333) + 0.0167 * \\ & \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0333 * \log_2(0.0333) \\ & + 0.0167 * \log_2(0.0167) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0167 * \\ & \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + \\ & 0.0333 * \log_2(0.0333) + 0.0167 * \log_2(0.0167) + 0.0333 * \log_2(0.0333) + 0.05 * \log_2(0.05)) \end{aligned}$$

$$H = 5.048$$

2.2 Max Entropia

2.2.1 Resposta

No caso temos 60 dígitos, então temos 60 classes:

A entropia máxima ocorre quando todos os valores possíveis têm a mesma probabilidade de ocorrência. Como temos 60 dígitos, podemos calcular a entropia máxima usando:

$$HMaxima = -\sigma(p(x)) * \log_2(p(x))$$

onde $p(x)$ é a probabilidade igual para cada valor possível.

Como temos 60 dígitos, a probabilidade de cada valor ocorrer é $1/60$, pois todos os valores têm a mesma chance.

Portanto, a entropia máxima é:

$$H \text{ Máxima} = -(60 * (1/60) * \log_2(1/60))$$

$$H \text{ Máxima} = -\log_2(1/60)$$

Calculando esse valor, encontramos a entropia máxima dos dados:

$$HMaxima = 5.9069.$$

2.2.2 Programação

2.3 Código

Essa questão eu utilizei o Python na IDE **Spyder (Python 3.9)**:

```

# -*- coding: utf-8 -*-
"""
Created on Wed May 31 23:54:36 2023

@author: Henrique Freitas
"""
import numpy as np

def calcular_entropia(dados):
    valores_unicos, contagem = np.unique(dados, return_counts=True) # Obtém valores únicos e suas contagens
    probabilidades = contagem / len(dados) # Calcula as probabilidades de cada valor
    entropia = -np.sum(probabilidades * np.log2(probabilidades + 1e-10)) # Calcula a entropia, adicionando uma pequena quantidade para evitar o valor
    return entropia

def calcular_maxima_entropia(dados):
    num_valores = np.max(dados) + 1 # Obtém o número de valores distintos
    probabilidade_maxima = 1 / num_valores # Probabilidade uniforme para cada valor
    maxima_entropia = -np.log2(probabilidade_maxima) # Calcula a máxima entropia
    return maxima_entropia

# Dados fornecidos
bolas = np.array([1, 3, 7, 15, 23, 44, 13, 16, 35, 41, 42, 47, 1, 9, 17, 30, 31, 44, 6, 23, 25, 33, 34, 47, 6, 16, 21, 24, 26, 45, 1, 19, 22, 32, 39, 4

entropia = calcular_entropia(bolas)
maxima_entropia = calcular_maxima_entropia(bolas)

print("Entropia:", entropia)
print("Máxima Entropia:", maxima_entropia)

```

Figura 1: Código em Python utilizado para calcular todo o trabalho

```

Entropia: 5.048394340054162
Máxima Entropia: 5.672425341971495

In [5]:

```

Figura 2: Resultado da Entropia e Máxima entropia