



Teoria de Informação – LEI 2024

TP1 - Entropia, Informação Mútua e Codificação de Huffman

Licenciatura de Engenharia Informática

2023/2024

Tomás Amaro Ricardo Simões Guilherme Carvalho

2023212581 2023212489 2023212131

Este relatório serve para responder às perguntas colocadas no enunciado do trabalho prático 1

**2.d)** Relação de MPG com a Aceleração do Carro -> O rendimento do combustível tende a aumentar com o aumento da aceleração do carro

Relação de MPG com o Número de Cilindros do Motor -> O rendimento do combustível começa a aumentar até aos 4 cilindros, mas depois tende a diminuir à medida que o número de cilindros aumenta

Relação de MPG com o Deslocamento do Ar por Pistão -> O rendimento do combustível tende a diminuir com o aumento do Deslocamento do Ar por Pistão

Relação de MPG com Cavalos de Potência -> O rendimento do combustível tende a diminuir com o aumento dos cavalos de potência do carro

Relação de MPG com Ano de Fabrico -> O rendimento do combustível tende a aumentar com o quão mais recente é o carro

Relação de MPG com o Peso do Carro -> O rendimento do combustível tende a diminuir com o aumento do peso do carro

**7.c)** A partir da observação dos resultados, podemos concluir que:

* A variável Cilindros tem um valor bastante baixo de entropia ou seja são valores bastante únicos como (4,6,8).

* A variável Aceleração e Ano de Fabrico têm uma entropia relativamente alta ou seja ainda apresentam alguns valores diversificados na matriz das informações.

* As variáveis Deslocamento de Ar por Pistão, Cavalos de Potência e Rendimento do Combustível têm uma entropia alta o que significa que ainda têm vários valores diversificados.

* A variável do Peso é a que apresenta maior entropia, o que significa que os seus valores ainda são bastante diversificados.

**8.b)** Comparando os valores de entropia utilizando a codificação de Huffman com os valores de entropia já calculados no ponto 7, podemos observar que estes são muito parecidos, sendo os da codificação de Huffman um pouco mais elevados que os do ponto 7

**8.c)** A variância dos comprimentos pode ser reduzida se fizermos um binning com intervalos maiores, de forma a reduzir a diversificação de valores diferentes, o que vai aumentar as probabilidades de cada valor e diminuir a quantidade de bits necessários para representar cada um deles. A importância é usar ainda menos bits para codificar a fonte de informação.

**10.b)** A variável MPG tem uma informação mútua superior a 0 com cada uma das outras variáveis, o que significa que sabendo algo sobre uma delas, obtemos informação da outra. Relacionando com os coeficientes de correlação de Pearson:

- Nas variáveis Cilindros do Motor, Deslocamento de Ar por Pistão, Cavalos de Potência e Peso, por causa de terem coeficientes de Correlação de Pearson negativos, sabemos que quando o MPG aumenta estas diminuem e vice-versa;

- Nas variáveis Aceleração e Ano de Fabrico, por causa de terem coeficientes de Correlação de Pearson positivos, sabemos que quando o MPG aumenta estas também aumentam e vice-versa

**11.b)** Os valores previstos na variável MPG diferem muito dos seus valores reais.

**11.f)**  O mean absolute error e o root mean square error dizem-nos que há uma grande diferença entre os valores previstos e os verdadeiros, tanto quando calculamos com as informações completas como quando substituímos a variável com menor informação mútua pelo seu valor médio. Quando substituímos a variável com maior informação mútua pelo seu valor médio, obtemos um valor ainda maior para o mean absolute error e para o root mean square error.