

INDÍCE

- Introdução
- Resolução dos exercícios
- Conclusão

INTRODUÇÃO

Realizamos o presente relatório, onde o principal objetivo do trabalho foi adquirir sensibilidade para as questões fundamentais de teoria de informação, em particular entropia e informação mútua, assim como as suas aplicações práticas. Com este relatório pretendemos fazer um resumo do trabalho realizado, respondendo às questões que nos fazem refletir sobre os resultados obtidos.

RESOLUÇÃO DE EXERCÍCIOS

2- d)

Relação de Acceleration com MPG:

A relação positiva de 0.4136 sugere que veículos que aceleram mais rapidamente tendem a ter um melhor consumo de combustível.

Relação de Cylinders com MPG:

A forte relação negativa de -0.7761 indica que veículos com menos cilindros tendem a ter um melhor MPG, enquanto veículos com mais cilindros têm um MPG mais baixo.

Relação de Displacement com MPG:

A relação negativa de -0.8055 mostra que veículos com um menor deslocamento do motor tendem a ter um melhor consumo de combustível (MPG).

Relação de Horsepower com MPG:

A relação negativa de -0.7551 indica que veículos com menor potência do motor tendem a ter um melhor MPG.

Relação de ModelYear com MPG:

A relação positiva de 0.5873 sugere que, ao longo do tempo, os veículos têm melhorado seu consumo de combustível (MPG). À medida que o ano do modelo aumenta, o consumo de combustível tende a melhorar.

Relação de Weight com MPG:

A forte relação negativa de -0.8312 mostra que veículos mais leves tendem a ter um melhor MPG.

Relação de MPG com MPG:

A relação entre MPG e ela mesma é sempre 1.0, o que indica uma correlação perfeita.

7-c)

Acceleration:

O resultado que obtivemos através da entropia normal foi 3.496423557860517 o que se revelou um valor ligeiramente alto que por sua vez mostra alguma incerteza.

Cylinders:

O resultado que obtivemos através da entropia normal foi 1.590435690353271, o que relevou um valor de entropia normal.

Displacement:

O resultado que obtivemos através da entropia normal após binning foi 4.8740687856999925, o que revelou um valor muito alto, o que demonstra uma maior incerteza, comparado com a acceleration.

Horsepower:

O resultado que obtivemos através da entropia normal após binning foi 4.597999424808645, sendo um caso igual ao do displacement, demonstrando uma incerteza alta.

ModelYear:

O resultado que obtivemos através da entropia normal foi 3.6906425111722605, sendo um valor inferior aos anteriores, mas semelhante ao valor da entropia da acceleration.

Weight:

O resultado que obtivemos através da entropia normal após binning foi 6.040364750974289, sendo este o valor mais alto de entropia até agora, demonstrando assim a maior incerteza até então.

MPG:

O resultado que obtivemos através da entropia normal foi 4.835799622324453, o que é um valor elevado, tendo assim uma incerteza também elevada, tal como as entropias do displacement e do horsepower

Cálculo conjunto de dados completo:

A entropia completa (após o binning das 3 variáveis pedidas) é calculada através de todos os dados e o resultado foi 6.657726321303801. Este valor é bastante elevado, tal como esperado, sendo até o mais elevado entre todos calculados, o que significa que também tem a maior incerteza.

8-b)

A entropia normal é uma medida de incerteza em um conjunto de dados, enquanto a entropia de Huffman é uma versão comprimida da entropia normal obtida após a aplicação do algoritmo de codificação de Huffman.

Acceleration:

Entropia normal: 3.496423557860517

Entropia de Huffman: 0.8138413150698164

Observação: Entropia de Huffman é menor, indicando maior previsibilidade após a codificação de Huffman.

Cylinders:

Entropia normal: 1.5904356903532713

Entropia de Huffman: 0.7131947672488212

Observação: Entropia de Huffman é menor, sugerindo menor complexidade nos dados.

Displacement:

Entropia normal após binning: 4.8740687856999925

Entropia de Huffman após binning: 1.4565496924219286

Observação: Entropia de Huffman após "binning" é menor, indicando redução na complexidade dos dados.

Horsepower:

Entropia normal após binning: 4.597999424808645

Entropia de Huffman após binning: 1.5355722038768724

Observação: Entropia de Huffman após "binning" é menor, sugerindo menor incerteza.

ModelYear:

Entropia normal: 3.6906425111722605

Entropia de Huffman: 0.198347107438016

Observação: A Entropia de Huffman é muito baixa, indicando pouca desordem nos dados.

Weight:

Entropia normal após binning: 6.040364750974289

Entropia de Huffman após binning: 0.807466389775972

Observação: Entropia normal é alta, mas a de Huffman é baixa, sugerindo maior previsibilidade após "binning."

MPG:

Entropia normal: 4.835799622324453

Entropia de Huffman: 0.8847623589638332

Observação: Entropia de Huffman é significativamente menor do que a entropia normal, indicando uma redução na complexidade dos dados.

A aplicação da codificação de Huffman parece reduzir a entropia dos dados, tornando-os mais previsíveis e menos complexos, especialmente quando os valores não são igualmente prováveis.

8-c) Para melhorar a codificação de Huffman, duas abordagens eficazes são: "binning prévio", agrupando símbolos semelhantes antes da codificação, e o "ajuste da árvore de Huffman", que permite redistribuir símbolos na árvore para comprimentos de códigos mais uniformes. Ambas otimizam a eficiência na compressão de dados e transmissão de informações.

10-b) Relação da informação mútua com os coeficientes de correlação de Pearson.

A informação mútua entre MPG e as outras variáveis é dada por :

$$I(\text{Var}, \text{MPG}) = H(\text{Var}) - H(\text{Var}/\text{MPG}).$$

Quanto maior for o valor absoluto do coeficiente de correlação de Pearson menor será $H(\text{Var}/\text{MPG})$, e mais próxima estará a informação mútua do valor de $H(\text{Var})$.

11-b) Calculamos o MPG contudo verificamos que não correspondia com o ideal. Os valores apresentados são sempre mais pequenos como podemos ver nestas 2 imagens dos 9 primeiros elementos.(lado esquerdo mpg estimado; lado direito mpg tabelado).

<u>15.4061</u>	<u>18</u>
<u>14.2505</u>	<u>15</u>
<u>16.0505</u>	<u>18</u>
<u>15.8629</u>	<u>16</u>
<u>15.8474</u>	<u>17</u>
<u>10.8706</u>	<u>15</u>
<u>10.9826</u>	<u>14</u>
<u>11.1516</u>	<u>14</u>
<u>10.2444</u>	<u>14</u>

11-e) Calculando o valor normal obtivemos 18.5887, em seguida utilizamos o menor IM que resultou em 21.2167, que se nota um aumento mesmo sendo este pequeno. Em seguida calculamos o 3º IM e o resultado foi bastante diferente dos anteriores sendo quase o dobro do normal ,36.7902, ou seja aumenta muito.

CONCLUSÃO

Concluindo e tendo em conta o nosso principal objetivo, podemos concluir que adquirimos sensibilidade para as questões fundamentais da teoria de informação, como por exemplo através da realização da entropia e informação mútua, bem como as suas aplicações práticas. Após a realização dos exercícios concluimos também que fizemos um bom trabalho concluindo com qualidade o que era esperado.