

Parte I

[1] Quais as principais limitações de modelos de correlações?

Uma das principais limitações de modelos de correlação é a suposição de causalidade. O facto de os modelos estabelecerem correlações entre atributos que não apresentam razão lógica para estarem correlacionados, pode induzir o utilizador em erro. Neste caso, apesar de os modelos estabelecerem uma relação estatística entre os atributos, na verdade, esta relação não faz sentido.

Outra limitação é a direcionalidade, isto é, o valor da correlação estabelecida não nos informa qual dos atributos é a causa do outro. Assim, se existe uma correlação forte entre A e B, o valor do coeficiente de correlação não nos permite saber se A é a causa para B acontecer ou vice-versa.

Para além disso, os modelos de correlação apenas podem ser estabelecidos entre atributos numéricos.

[2] O que é um coeficiente de correlação e como é interpretado?

O coeficiente de correlação permite avaliar a força dos relacionamentos entre atributos numéricos, podendo ser interpretado como a distância do ponto à reta traçada pelo conjunto. A sua interpretação traduz-se através de valores negativos entre 0 e -1 e valores positivos entre 0 e 1, sendo que quanto mais próximo o valor estiver de 1 (na correlação positiva) e -1 (na correlação negativa mais forte é a relação entre os atributos).

[3] Qual a diferença entre uma correlação negativa e uma correlação positiva?

Uma correlação positiva representa uma relação positiva entre os atributos, isto é, se um aumenta, o mesmo acontece com o outro. O valor do coeficiente desta correlação representa a força desta afirmação.

Uma correlação negativa entre dois atributos representa, por outro lado, uma relação inversa entre ambos. Assim, se A e B apresentam uma correlação negativa, isto é, se A aumenta, B diminui. No entanto, é necessário manter, novamente, a atenção ao valor do coeficiente desta correlação. Para um coeficiente de correlação de -1 , esta afirmação tem bastante força, enquanto para -0.1 esta afirmação já se revela mais fraca.

a) Se dois atributos diminuem essencialmente à mesma taxa é uma correlação positiva ou negativa? Explique.

É uma correlação positiva pois à medida que um atributo diminui, o mesmo acontece com o outro atributo com que estamos a correlacionar.

[4] Como é medida a força de uma correlação? Quais os limites para essa força?

Quanto mais próximos, em módulo, de 1, mais forte será a relação. O limite é de 1 para correlações positivas e de -1 para correlações negativas.

[5] Consegue pensar em atributos que poderiam ser interessantes incluir no dataset estudado no exemplo da aula?

Serão atributos interessantes aqueles que nos ajudem a perceber mais sobre os atributos das correlações encontradas.

Relativamente ao dataset estudado no exemplo da aula, seria interessante incluir um atributo que representasse a quantidade de aquecedores a gás, o número de horas que os habitantes passam na casa.

Parte II

[1] Documente quais os atributos que podem influenciar ou explicar o consumo/eficiência de combustível num determinado veículo (mpg).

Com este estudo foi possível perceber que os atributos que mais podem influenciar a eficiência de consumo de combustível num determinado veículo são o *weight*, que apresentou um coeficiente de correlação de $-0,829$ e o *displacement*, que apresentou um coeficiente de $-0,803$.

Tanto o atributo *weight* correlacionado com o atributo *mpg*, como o atributo *displacement* correlacionado com o atributo *mpg*, apresentam uma correlação negativa. Assim, à medida que o peso do veículo aumenta, diminui a eficiência do mesmo. Relativamente à cilindrada, acontece o mesmo, por isso, à medida que a cilindrada do veículo aumenta, diminui também a eficiência de consumo de combustível do mesmo.