

Proposta - Trabalho Prático 2

Agrupamentos das características importantes de um carro

Base de dados e contextualização

A base de dados escolhida foi encontrada no Kaggle, denominada como Car Features and MSRP (<https://www.kaggle.com/CooperUnion/cardataset>). A base de dados está no formato CSV e é composta por 16 atributos e 11915 registros diferentes, fornecidos pelos websites Edmunds e Twitter, que descrevem um carro. Dentre os atributos, temos a montadora, o modelo do carro, o ano de fabricação, o tipo do motor de gasolina, o tipo de transmissão, o número de portas, o tamanho do carro e entre outros. O dataset foi escolhido por possuir uma documentação clara e com um bom número de registros e atributos de fácil entendimento, sendo possível realizar as operações de agrupamentos entre as montadoras, por exemplo.

Objetivo

Agrupar características importantes de um carro a partir do dataset definido, utilizando os algoritmos K-Means e Mistura Gaussiana. As características desejadas para o agrupamento são:

- Montadora do carro
- Categoria de mercado
- Tamanho do carro
- etc.