

Introdução a Modelagem e Aprendizado

ELTDI - DAELN UTFPR-CT

Prof. Heitor S. Lopes (2026-2) - Exercício #3

Objetivo: utilizar métodos de agrupamentos para a análise de dados



Parte I: Identificação de motoristas

- Uma empresa fabricante de veículos elétricos disponibilizou um veículo de teste para observar o seu desempenho quando operado por diferentes motoristas. Cada motorista tinha liberdade de dirigir de acordo com sua conveniência no que se refere ao seu modo de dirigir e percurso. Os motoristas se revezavam em trechos distintos de estradas e/ou regiões urbanas, de modo que o veículo foi testado em uma grande diversidade situações e rodovias. Diversos dados foram coletados durante cada sessão de direção, incluindo, por exemplo, acelerações/giroscópio (em 3 eixos), tensão e voltagem das baterias, velocidade, posicionamento dos pedais (acelerador e freio), uso de ar condicionado, altitude do local, etc. O dataset **Xmotoristas.csv** contém 5000 amostras, cada uma correspondendo a um trecho de um motorista.
- Uma vez que os dados foram anonimizados, não se sabe quantos motoristas diferentes dirigiram o veículo e nem quantas vezes cada motorista fez isto. Importante ressaltar que neste estudo vale o “estilo de direção” próprio de um motorista. Isto quer dizer que um motorista terá preponderantemente o mesmo estilo, independentemente da estrada e do veículo. Também é possível que mais de um motorista tenha aproximadamente o mesmo estilo de direção.
- Portanto, utilizando métodos de agrupamento (em especial, K-means e t-SNEE) deseja-se saber: (1) quantos estilos de direção há neste grupo de motoristas; (2) qual a distribuição aproximada do número de trechos rodados para cada estilo de direção.



Parte 2: Caracterização de clientes de shopping

- a. O objetivo deste estudo é caracterizar em grupos os clientes que frequentam um determinado shopping de Curitiba, de modo a identificar *clusters* de clientes. Desta maneira, entendendo as características dos grupos pode-se potencializar as vendas com campanhas de *marketing* dirigidas a cada grupo. Neste *dataset*, cada cliente tem uma identificação, informação de gênero, idade, renda anual aproximada e gastos realizados no shopping em um determinado período (estes dois últimos dados – renda e gastos - estão em uma escala arbitrária).
- b. Utilize o dataset **ClientesShopping.csv** e aplique os métodos de agrupamento *K-means* e hierárquico. Tenha em mente que é necessário separar os clientes em grupos **que tenham significado**. O objetivo da análise dos dados, obviamente, é aumentar as vendas através de estratégias de marketing focadas em grupos específicos, **claramente categorizados** (p.ex. pessoas com alta renda, mas com gastos medianos/baixos). Deve-se caracterizar todos os grupos encontrados. Mostrar os parâmetros utilizados nos algoritmos e os resultados com gráficos comentados.