

Introdução à Modelagem e Aprendizado

ELTDI-DAELN UTFPR-CT

Prof. Heitor S. Lopes (2026-2) - Exercício #7

Objetivo: extração de *features* e agrupamento de séries temporais



- Quatro cruzamentos de capitais brasileiras foram monitorados para contabilizar o número de veículos que passaram por hora. Foram coletados 6 meses de dados (de 01-novembro-2015 até 29-abril-2016), num total de 4344 observações (182 dias x 24 horas) para cada um dos 4 cruzamentos. Estas informações estão na planilha *veiculos-cruzamentos.csv*.
- Separe o dataset em 6 partes, uma para cada mês. Depois utilize a biblioteca **Python TSFEL** para extrair *features* das séries temporais de cada cruzamento. Assim, teremos um vetor de *features* para cada mês e para cada cruzamento. Carregue estes vetores no Orange, **com a identificação do cruzamento e do mês**.
- Utilizando T-SNEE observe a formação de clusters e/ou *outliers* e explique o que foi observado.
- Para corroborar a observação qualitativa do item anterior, utilize o algoritmo de clusterização hierárquica para inferir sobre a similaridade/dissimilaridade do comportamento do tráfego entre cada um dos 4 cruzamentos, mês a mês. Quantos grupos distintos são observados ? Onde há melhor e pior agrupamento ? **O que se pode descobrir de interessante sobre isto ?**
- Escolha um das séries temporais (cruzamento x mês) e utilize a ferramenta “*Neighbors*” para encontrar qual a outra série temporal mais similar a esta. Confira na classificação hierárquica se este resultado está de acordo com a clusterização.