



INF-0615 – APRENDIZADO DE MÁQUINA SUPERVISIONADO I

EXERCÍCIO 6

PROTOCOLO UM-CONTRA-TODOS (*ONE-VS-ALL*) ECONOMIC REGION PREDICTION

1 Descrição do Problema

O globo é composto por diversos países que apresentam suas próprias características, culturas, histórias e economias. No entanto, países vizinhos tendem a passar por processos históricos similares, inclusive muitos sendo originados de um mesma nação que um dia foi unificada (países do Leste Europeu, por exemplo, após o desmantelamento da União Soviética). Por consequência, os países e seus vizinhos tendem a ter economias similares entre si, o que muitas vezes os levam a formar blocos econômicos, mas que podem ser significativamente diferentes de outras regiões do mundo. Assim, os países da Europa Ocidental tem características econômicas diferentes daqueles países na América do Sul por exemplo.

Nesse contexto, o objetivo desse exercício é classificar, baseado em atributos políticos e econômicos de um país, a qual região do mundo ele pertence. Há um total de 34 atributos que serão utilizados para a predição da região, a qual está presente na coluna *continent*. Além disso, há 5 regiões possíveis (classes): **Central America**, **Eastern Africa**, **South-Eastern Asia**, **South America** e **Westen Europe**, ou seja, é um problema **multi-classe**.

2 Tarefas

Neste exercício, pedimos que você:

1. Inspeção os dados, verifique a distribuição das classes e os tipos dos atributos.
2. Normalize os dados para que fiquem mais bem preparados para o treinamento.
3. Treine 5 regressões logísticas, uma para cada região, seguindo o protocolo One-vs-All.
4. Execute cada modelo treinado no item anterior sobre o conjunto de validação. Tome a resposta de cada classificador como um voto, realize a contagem dos votos para cada exemplo e determine a classe predita.
5. Calcule a matriz de confusão e a acurácia balanceada. Ouve alguma classe favorecida?

3 Arquivos

Os arquivos disponíveis no Moodle são:

- *OvA_training_set.csv*: dados de treinamento;
- *OvA_validation_set.csv*: dados de validação;
- *Ex06.R*: código que implementa a solução do exercício;

4 Referências

1. *Economic Freedom of the World*. Kaggle. <https://www.kaggle.com/gsutters/economic-freedom>.