

# Exercício 4

**Data de entrega: 31/10, as 7:00 (da manha).**

Use os dados [do arquivo cluster-data.csv](#) (os dados sao uma media de 30 medidas por vez da pessoa 1 do dataset [Activity Recognition from Single Chest-Mounted Accelerometer Data Set](#)

Rode o kmeans nos dados, com numero de restarts = 5

Use alguma metrica interna (algum Dunn, Silhouette, Calinski-Harabaz index) - apenas uma -para escolher o k entre 2 e 10.

O arquivo [cluster-data-class.csv](#) contem a classe correta de cada ponto. Use alguma medida externa (Normalized/adjusted Rand, Mutual information, variation of information) para decidir no k.

Plote os graficos correspondentes das 2 metricas (interna e externa) para os varios valores de k (extra).

## Detalhes R

A função [cluster.stats](#) do pacote fpc computa varias metricas internas, e externas se alt.clustering for fornecido

## Detalhes Python

Sklearn tem varias [metricas de cluster](#) implementadas

---

Last modified: Sun Oct 16 21:35:40 BRST 2016