

# IA048 – Aprendizado de Máquina

## Atividade 2 – Classificação

Turma A – 1º semestre de 2024

Prof: Levy Boccato Email: lboccato@dca.fee.unicamp.br

Prof: Romis Attux Email: attux@unicamp.br

## Introdução

Nesta atividade, vamos abordar o problema de reconhecimento de atividades humanas (HAR, do inglês *human activity recognition*) a partir de informações capturadas por sensores de *smartphones*. Em particular, vamos trabalhar com a base de dados UCI HAR [Anguita et al., 2013], que contém registros de sensores inerciais presentes em um *smartphone* preso à cintura de 30 sujeitos realizando atividades cotidianas. Cada pessoa realizou seis atividades, as quais correspondem aos seguintes rótulos:

Atividade	Rótulo
Caminhar	0
Subir escadas	1
Descer escadas	2
Sentado	3
Em pé	4
Deitado	5

Foram capturadas as amostras dos três eixos ( $x$ ,  $y$  e  $z$ ) do acelerômetro (ACC, do inglês *accelerometer*) e do giroscópio (GYR, do inglês *gyroscope*) presentes no *smartphone*, empregando uma taxa de amostragem de 50 Hz. O conjunto completo de amostras foi particionado aleatoriamente em treinamento (70% dos voluntários) e teste (30% dos voluntários). Para facilitar a visualização da tarefa de HAR, apresentamos na Figura 1 exemplos de registros de ACC e GYR para quatro atividades humanas.

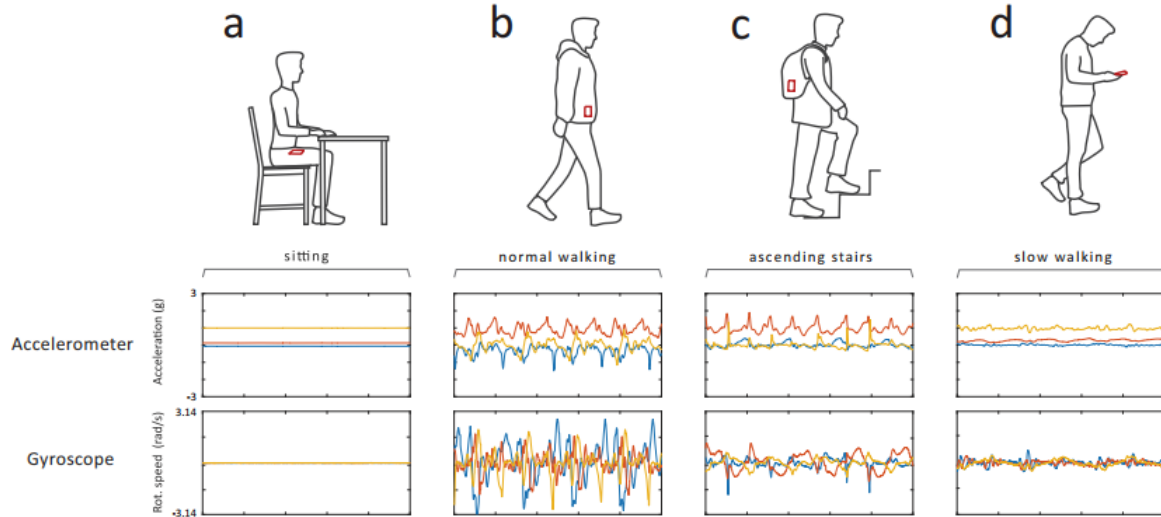


Figura 1: Exemplos de dados brutos de ACC e GYR coletados em um cenário realista. **a)** uma pessoa está sentada apoiada em uma mesa com o *smartphone* em um bolso frontal da calça; **b)** uma pessoa caminha normalmente (cerca de 1,9 passos por segundo) com o *smartphone* no bolso da jaqueta; **c)** uma pessoa sobe escadas com o *smartphone* em sua mochila; **d)** uma pessoa está caminhando lentamente (cerca de 1,4 passos por segundo) enquanto segura o aparelho em sua mão. Adaptada de [Straczekiwics et al., 2021]

## Primeira parte

Primeiramente, será explorada uma versão do conjunto de dados na qual já houve pré-processamento e extração de características. No caso, cada amostra contém 561 atributos derivados de uma mesma janela de 2,56 s dos 6 sinais disponíveis (ACC:  $x,y,z$ ; GYR:  $x,y,z$ ), considerando suas representações tanto no domínio do tempo quanto no domínio da frequência.

- a) Construa uma solução para este problema baseada no modelo de regressão logística. Descreva a abordagem escolhida para resolvê-lo (*softmax*, classificadores binários combinados em um esquema um-contra-um ou um-contra-todos). Obtenha, então, a matriz de confusão para o classificador considerando os dados do conjunto de teste. Além disso, adote uma métrica global para a avaliação do desempenho (médio) deste classificador. Discuta os resultados obtidos.
- b) Considere, agora, a técnica *k-nearest neighbors* (kNN). Adotando um esquema de validação cruzada, mostre como o desempenho do classificador, computado com a mesma métrica adotada no item a)) varia em função do parâmetro  $k$ . Escolhendo, então, o melhor valor para  $k$ , apresente a matriz de confusão para os dados de teste e o desempenho medido nesse conjunto. Comente os resultados obtidos, inclusive estabelecendo uma comparação com o desempenho da regressão logística.

## Segunda parte

Agora, vamos utilizar os dados “brutos” combinados de ACC e GYR como entradas dos classificadores. Para isso, devemos recorrer aos registros disponibilizados no diretório ‘Inertial Signals’, os quais estão separados por eixo e por sensor, sendo que cada amostra individual agora é formada por 128 valores (atributos), que correspondem às amplitudes instantâneas de aceleração (ACC) ou velocidade angular (GYR) dentro de uma janela de 2,56 s.

- c) Monte, então, a nova matriz de entrada concatenando os seis sinais temporais e, então, repita o procedimento experimental detalhado nos itens a) e b). Ao final, com base no desempenho obtido, faça uma análise comparativa entre a abordagem do item anterior e a abordagem baseada nos sinais “brutos” empregada nesta segunda parte.

## Referências

- [Anguita et al., 2013] Anguita, D., Alessandro Ghio, L. Oneto, Xavier Parra and Jorge Luis Reyes-Ortiz. “A *Public Domain Dataset for Human Activity Recognition using Smartphones*.” The European Symposium on Artificial Neural Networks (2013).
- [Strackiewicz et al., 2021] Strackiewicz, M., James, P. & Onnela, JP. “A *systematic review of smartphone-based human activity recognition methods for health research*”. npj Digit. Med. 4, 148 (2021).