

---

# Reconhecimento de Padrões Inteligência Geoespacial

2019/2020

*Project Assignment*

## Smartphone and Smartwatch Activity Recognition

---

### 1 Background

Monitoring human activity is important for many goals, being one of the most important ones people's health and well-being. Physical inactivity and unhealthy eating are two risk factors for disease that should be taken in consideration. Nowadays, personal and wearable devices, like smartwatches and smartphones possess computational power and sensors that may be used for human activity recognition.

Your job in this assignment is to classify human activities using accelerometer and gyroscope data collect from smartphones and smart-watches!

### 2 Dataset Description

Consider the dataset available at <https://archive.ics.uci.edu/ml/datasets/WISDM+Smartphone+and+Smartwatch+Activity+and+Biometrics+Dataset++>. This dataset contains the raw accelerometer and gyroscope sensor data collected from the smart-phone and smart-watch at a sampling rate of 20Hz. Data from 51 test subjects as they perform 18 activities for 3 minutes were recorded. The sensor data for each device (phone, watch) and type of sensor (accelerometer, gyroscope) is stored in a different directory (so there are 4 data directories). In each directory there are 51 files corresponding to the 51 test subjects.

In addition to the raw time-series sensor data, 91 features from the sensor data are also available. See the detailed dataset description document (*WISDM-dataset-description.pdf*) available with the data folder, for details.

**Note:** Feature data are available for all participants, with the exception of participant ID 1614.

### 3 Objective

Your task is to develop classifiers for human activity recognition. Consider three scenarios:

- **Scenario A (Binary Classifier):** where one wants to distinguish the **Jogging** activity from the others activities;
- **Scenario B (Three-class Classifier):** where one wants to distinguish **Non-hand-oriented activities** (walking, jogging, stairs, standing, kicking) from **General Hand-oriented activities** (dribbling, playing catch, typing, writing, clapping, brushing teeth, folding clothes), and from **Eating Hand-oriented activities** (eating pasta, eating soup, eating sandwich, eating chips, drinking).
- **Scenario C (18-class Classifier):** where one wants to distinguish all the 18 activities.

## 4 Practical Assignment

### 4.1 Data import

Develop scripts for feature data import. Organize data into four sub-sets, i.e.: features from smart-phone accelerometer; features from smart-phone gyroscope; features from smart-watch accelerometer; and features from smart-watch gyroscope.

### 4.2 Feature Selection and Reduction

Some of the supplied features may be useless, redundant or highly correlated with others. In this phase, you should consider to use feature selection and dimensionality reduction techniques, and see how they affect the performance of the pattern recognition algorithms. Analyse the distribution of the values of your features and compute the correlation between them. Make sure you know your features! Do not forget to present your findings in the final report.

### 4.3 Experimental Analysis

You should be able to design experiences in order to run the pattern recognition algorithms in the given data and evaluate their results. Define the appropriate performance metrics and justify your choices!

Run the experiments multiple times! To be able to present average results and standard deviations (of the metrics used) you should split the training set in parts and use cross-validation. At the end you should be able to choose the best classifier and evaluate them in a testing set.

Do not forget that manually inspecting the predictions of your algorithms can give you precious insights of where they might be failing (and why), and what you can do to improve them (e.g. what makes the algorithm fail in this particular case? what special characteristic does it have that makes it so hard? how can I make the algorithm better deal with those cases?). Go back and forward to the Pre-processing, Feature reduction and Feature Selection phases until you are satisfied with the results. It is a good idea to keep track of evolution of the performance of your algorithm during this process. Try to show these trends in your final report, to be able to justify all the issues involved (choosing parameters, model fit, etc.)

### 4.4 Pattern Recognition Methods

You can write your own code in your language of choice or use the functions and methods available in Matlab and in the Statistical Pattern Recognition STPRTool used in the classes (since you are already familiarized with it). The methods used in your work should be described as well as discussion of the parameters used. Try out different pattern recognition algorithms. You should try to understand how they perform differently in your data.

### 4.5 Results and Discussion

Present and discuss final results obtained in your Project assignment. This problem was already studied by other authors. Compare your results with the results from other sources. In this problem one important aspect is to evaluate among the four data-subsets the ones more appropriate for the different scenarios and activities.

### 4.6 Code & Graphical User Interface (GUI)

You should deliver your software code in MATLAB, or in any other programming language you used during the project.

For your project you should write code for a graphical user interface (GUI). The GUI should improve the interaction of the user with the code by providing options for data-loading, feature selection/dimensionality reduction, classification, post-processing, validation and visualization. Remember to comment your code. Write also a help section to your code that tells the purpose of the function, usage, and explanation of parameters.

## 5 Documentation

Write documentation (in Portuguese or in English) about your project. The documentation should include a cover page where course name, project title, date, names and student numbers of the authors are mentioned.

Describe the methods used for classification in such detail that the reader would be able to implement the same kind of functions for feature extraction and classification just based on your documentation and some basic background in pattern recognition. Always justify your choices, even when they are based on intuition. Do not forget to verify your assumptions! Include classification results with the given data to your documentation. At the end of your documentation you should have a list of all references used.

### 5.1 Requirements

Practical assignment is meant to be done in groups **of three** persons. If someone wants to work alone, this is also possible. Larger groups are not allowed.

### 5.2 Project Submission & Deadlines

#### 1. Project First Milestone (**Deadline: 17th April 2020!**)

Deliverables:

- Data Preprocessing (Scaling, Feature Reduction (PCA & LDA), Feature Selection, etc.);
- Minimum Distance classifier, Fisher LDA for Scenario A.
- Code + short report.

#### 2. Project Final Goal (**Deadline: 22th May 2020!**)

Deliverables:

- Data Preprocessing (Scaling, Feature Reduction (PCA & LDA), Feature Selection, etc.);
- Several classifiers;
- Final Report
- Matlab code + GUI.

#### 3. Presentation and Discussion (**27th May 2020!**)

### Acknowledgments

Credits to WISDM (Wireless Sensor Data Mining) Lab in the Department of Computer and Information Science of Fordham University and to UCI Machine Learning Repository for data supply and description. All publications based on the supplied dataset should cite “Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. IEEE Access, 7:133190-133202, Sept. 2019.”.