

# LDATS2470 - Project 2023

Classification of people based on their health status by the  
use of SVM

Rousseau Mathieu, 67001800

,



UCLouvain  
Belgium  
23/12/2022

## 1 Introduction

The aim of this project is to analyse a range of biomedical voice measurements from 31 people where 23 of parkinson disease. There are around 6 voice measurements per patient so that in total we have a collection of 195 observations. Each one contains severall voice measures that are detailed below. The 'status' column indicate is the patient has the parkinson disease or not.

## 2 Research question

Using support vector machine algorithms we want to discriminate the patients based on their health status.

Firstly, we will begin with a basic descriptive analysis of the different variables composing this dataset. Then we will first try to perform an Hard Margin SVM. If needed, in case of the presence of outliers or a non linear dataset, we could investigate respectively the Soft Margin SVM or the use of the kernel trick.

## 3 Exploratory data analysis

## Appendix

### Description of the different variables

The **response variable** is *status* : 1 if the subject has the Parkinson disease and 0 if not.

The **explanatory variables** are the following :

- *name* : the subject name along the recording number.
- *mdvp.fo* : the **average** local fundamental frequency.
- *mdvp.fhi* : the **maximum** local fundamental frequency.
- *mdvp.flo* : the **minimum** local fundamental frequency.
- *mdvp.jitter\_perc* (%), *mdvp.jitter\_abs* (*Abs*), *mdvp.rap*, *mdvp.ppq*, *mdvp.ddp* : these are several measures of variation in fundamental frequency.
- *mdvp.apq*, *mdvp.shimmer*, *mdvp.shimmer\_db*, *shimmer.apq3*, *shimmer.apq5*, *shimmer.dda* : these are several measures of variation in amplitude.
- *nhr*, *hnr* : 2 measures of noise to tonal components in the voice.
- *rdpe*, *d2* : 2 nonlinear dynamical complexity measures.
- *dfa* : signal fractal scaling exponent.
- *spread1*, *spread2*, *ppe* : 3 nonlinear measures of fundamental frequency variation.