COMPUTATIONAL INTELLIGENCE (CI-MAI)
Project proposal: Parkinson Telemonitoring

Lluís A. Belanche, Albert Fiol

Abstract

The Parkinson disease is a degenerative disorder of the central nervous system. Symptoms such as shaking are very common and well-known, but in the advanced stages of the disease thinking and behaviour problems may arise. Unfortunately, there is not yet an easy lab test to diagnose the disease, and doctors use brain scans and the medical history to base the diagnosis. The project amounts to build a prediction model for some continuous indicators associated with the disease.

1 Introduction

The Parkinson disease is a neurodegenerative disorder of the central nervous system that causes partial or full loss in motor reflexes, speech, behaviour, mental processing, and other vital functions. It is generally observed in elderly people and causes disorders in speech and motor abilities (writing, balance, etc.) in 90% of the patients. It is the second more common neurological health problem in elder ages and its incidence is estimated on about 10 million people all around the world. Currently there is no known cure for the disease, although there is a significant amount of drug therapies to decrease difficulties caused by the disorder. The disease is usually diagnosed and treated using invasive methods, based on symptoms like slowness in movement, poor balance, trembling, or stiffness of some body parts, but especially in voice problems. The main reason behind diagnosis from speech impairments is that telediagnosis and telemonitoring systems based on speech signals are easy to use, enable an early diagnosis and also lessen the workload of medical personnel.

In this problem we will try to model how the Parkinson disease evolves in a series of patients using a non-invasive test. The tests consist in periodic voice recordings at the patients homes, and we are given the task to predict how two indicators of Parkinsons disease relate to these recordings. These indicators are scores in the Unified Parkinsons Disease Rating Scale (UPDRS) and are divided in different categories. In our case we want to predict the motor UPDRS and the total UPDRS for the patients.

The data you need to used can be found at:

https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring

The data spans several measurements of 42 different patients along a 6 month period, for a total of 5,875 rows, 20 variables (19 not counting the ID). For each of the subjects (individuals) we know their age and sex and for each measurement we have 16 biometrical voice measures, accompanied by the motor and total UPDRS scores.

Take a moment to examine the data. Do you notice anything special Pick a single patient and check the test times). The measurements are grouped by patient and accompanied by a time measure; there are two basic ways to handle this situation, which in turn have an influence on how to partition the data for training/test or cross-validation purposes:

1. consider each row as an individual (independent) observation
2. consider individuals as a whole and never split the information in single individuals

In any case, you might want to investigate whether *centering* individuals (i.e., subtracting the mean of every whole individuals from every data row) could be benefitial for performance.

In order to make predictions you might want to use different techniques. Please consider using different neural networks and/or classical statistical methods (like multivariate regression). It will also be interesting to consider dimensionality reduction techniques, like PCA or PCR.