Telemonitoring of Parkinson's disease progression

Comparing between machine learning techniques that can be used for improve the prediction of Parkinson's disease.

Rahaf Nasser Alghamdi

Abstract

The goal of this project is to improve prediction of Parkinson's disease (PD) progression, which is needed to support clinical decision-making and to accelerate research trials. I worked with data provided in this link (https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring), I will use machine learning models to achieve promising results for this problem.

Design

Unified Parkinson's Disease Rating Scale (UPDRS) is widely used for tracking PD symptom progression. Motor- and Total-UPDRS are two important clinical scales. it can be predict UPDRS scores through analyzing the speech signal properties. with machine learning algorithms it can be improve the prediction

Data

This dataset consists of audio medical measurements of 42 people with early-stage Parkinson's disease. These people were hired for a six-month trial of a remote-controlled device to remotely diagnose the disease. The data were automatically recorded in patients' homes.

The columns include an index number of each patient, his age, his gender, the duration from the initial date of admission, two motor indicators UPDRS and total UPDRS, and 16 other medical sound measurements. Each line relates to one of 5,875 different sound measurements from those of the volunteers.

Algorithm

It can be separate the data into train and test, the train set is 30% of our original data set. Then divide the features, and select subset of important features to work with it.

Models:

Linear Regression, Polynomial Regression, k-Nearest Neighbors, Gradient Boosting Regressor, Decision Tree Regressor and Random Forests. all models generally perform well, there is no one to give us bad results, but Random Forests have the best possible performance.

Tools

Numpy and Pandas for data manipulation

Matplotlib and Seaborn for plotting

Scikit-learn for modeling

- Feature Selection with sklearn and Pandas
https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b
- Parkinsons Telemonitoring Data Set
https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring
- K-nearest Neighbors (KNN) Classification Model
https://www.ritchieng.com/machine-learning-k-nearest-neighbors-knn/

- Polynomial Regression
https://towardsdatascience.com/polynomial-regression-bbe8b9d97491

- An Implementation and Explanation of the Random Forest in Python | by Will ... - Medium
https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest...

- Mean Square Error & R2 Score Clearly Explained
https://www.bmc.com/blogs/mean-squared-error-r2-and-variance-in-regression-analysis/

- Make your Data Talk!
https://towardsdatascience.com/make-your-data-talk-13072f84eeac

- How to Build and Train Linear and Logistic Regression ML Models in Python

https://www.freecodecamp.org/news/how-to-build-and-train-linear-and-logistic-regression-ml-models-in-python/

- Linear Regression Example
https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html

- Accurate telemonitoring of Parkinso's disease progression by non-invasive speech tests

https://www.nature.com/articles/npre.2009.3920.1.pdf

- An Improved Approach for Prediction of Parkinson’s Disease using Machine Learning Techniques

https://export.arxiv.org/pdf/1610.08250