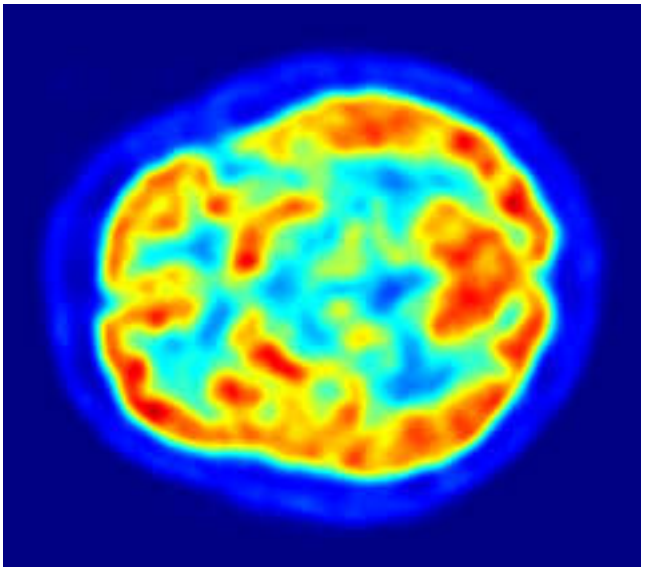
Initial practice

Statistical learning with deep artificial neural networks

Activity

Description

Parkinson's disease is a chronic and progressive neurodegenerative disorder. As dopamine-generating neurons in parts of the brain become damaged or die, people begin to experience difficulty speaking, writing, walking, or completing other simple tasks. These symptoms worsen over time, resulting in increased severity in patients.



In this activity, we propose the implementation of a dense deep neural network for the prediction of the severity of Parkinson's disease from data captured by telemonitoring of the voice of patients. Supplementary information about the data can be found at: https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring. .

This dataset is comprised of a variety of biomedical voice measurements from 42 people with early-stage Parkinson's disease recruited for a six-month trial of a telemonitoring device for remote monitoring of symptom progression. The recordings were automatically captured at the patients' homes.

The data set columns contain: `patient ID`, `age`, `sex`, `time interval since initial enrollment date`, `UPDRS_motor`, `UPDRS_total` and 16 voice biomedical measures. The variables that will be used in the PEC are: `total_UPDRS` and 16 biomedical voice measurements. Each row corresponds to one of the 5,875 voice recordings of these individuals.

The main objective of the activity is to predict the severity of Parkinson's from the 16 voice measures, where the severity of Parkinson's is based on a binary variable that is defined from the binarization of the variable `total_UPDRS`, indicating severity when `total_UPDRS> 25` and not otherwise.

Statement of the activity:

Create a dynamic report in R that performs:

1. Load the data file `parkinsons_updrs.data`.
2. Make a description of the dataset variables.
3. Create the binary variable of Parkinson's severity.
4. Normalize by means of the `min-max` transformation the variables of the 16 voice measurements.
5. Separate train and test.
6. Implement a dense deep neural network for severity prediction with two hidden layers and 10 nodes in each layer.
7. Provide the predictions of the test dataset.