Cogs 209 Mini-Project Proposal

Lead Author: Abbey Green

Co-authors: Carlos Escalante Vera, Anna McTigue, Miguel Avila-Morales

Title: **Using continuous motor activity data to classify depressive state**

Research question.

Globally, the incidence of depression is rising, and so there is a significant need to develop new tools to detect and treat depression. While surveys are canonically used to diagnose and track depression, they require motivation from the individual to complete, which is well known to be hindered in depression. We propose to use motor activity data to classify the depressive state of individuals. We predict that using a random forest algorithm may be able to predict depressive state based on motor activity, and we will also use a linear regression model for comparison. We expect that gender and age may impact motor activity, and so we will include them as covariates in the linear regression analysis. Given that motor activity can be recorded from widely available wearable devices, using motor activity to classify depression would make detecting and tracking depression easily accessible.

Data/materials.

We will use motor activity (actigraph data) (reference: Garcia-Ceja et al 2018). They collected data from 23 depressed individuals and 32 healthy controls. For each individual there are several days of motor activity data in addition to demographic data. Depressive state was determined using the Montgomery-Asberg Depression Rating Scale (MADRS). The data is available here.

It contains the following files:

File name	Format	Description
Condition	Folder with each file corresponding to an individual (each file being Comma separated values (CSV))	Actigraph data collected over time (motor activity data) for each individual in a data table. Columns are timestamp (minute by minute), date, and activity score. Row length varies depending on individual, but is typically 20000-50000 rows.
Control	Folder with each file corresponding to an individual (each file being Comma separated values (CSV))	Actigraph data collected over time (motor activity data) for each individual in a data table. Columns are timestamp (minute by minute), date, and activity score. Row length varies depending on individual, but is

		typically 20000-50000 rows.
scores.csv	csv	**Demographic data for each individual.** Columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1: working or studying, 2: unemployed/sick leave/pension), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped)

Course impact/relevance.

This project addresses regression, classification, cross-validation, supervised learning, and model selection.

Outcome(s). Provide examples of expected outcomes, e.g. analysis results or findings that may result from analyzing the data.

The outcome of this project will be the comparison of linear regression and random forest model performance on depression classification. We will generate a figure that shows the linear regression model performance including or not including age and gender as covariates. We will also compare model performance on training vs test data for cross-validation. We could also examine whether the model performs better on males compared to females, or whether model performance depends in some way on depression severity (i.e. performing better on severe depression compared to mild/moderate depression). It is also possible that binning motor activity across different time scales (15 min vs 1 hour vs 2 hours for example) could impact model performance.