FINAL TERM PROJECT, BIOSTAT620

Due: Monday, April 27, 2020

Do this project with a group (e.g. an existing group with max 5 people per group) or individually

The primary objective of the final term project:
To establish a prediction model for sleep/wake using summary features derived from the Empatica E4 wearable data collected from the entire class of twenty-five people. This project includes both aspects of training and **validation** on the prediction model.

Potential Publication/Creativity:
You are encouraged to submit your final project for a publication in a peer-reviewed journal during the spring/summer semester. In order to publish your project, I hope to see some creativity in your final project. The creativity may be shown in creation of new features (principal components), fast algorithms (e.g. divide-and-conquer), sparse feature selection, design a study for supervised learning, nonlinear prediction, prediction other types of outcomes, heterogeneity across groups, data integration, etc.

Data Preparation (Due by 11:59PM, April 13):
This time you first need to synchronize individual data, ACC (vector magnitude), TEMP, HR, EDC into mean, median per epoch (5 minutes), and label of sleep/wake. Then, you need to merge twenty-five data sets together as needed. Individual dataset should be prepared as an excel file, including the following variables and E4 summary statistics per 5-minute epoch:

E4_ID, Sleep, Sex, Age, VE, D.VE, Day, W.hours, ACC_Mean, ACC_SD, ACC_Q1, ACC_Q2, ACC_Q3, TEMP_Mean, TEMP_SD, TEMP_Q1, TEMP_Q2, TEMP_Q3, HR_Mean, HR_SD, HR_Q1, HR_Q2, HR_Q3, EDC_Mean, EDC_SD, EDC_Q1, EDC_Q2, EDC_Q3

Sex: Female (1), Male (0)
Age: in years
VE: Yes (1, having vigorous exercise at least 30 minutes for one day prior to sleep), No (0).
D.VE: Duration of vigorous exercise in minutes per day
Day: Mon(1), Tue(2), Wed(3), Thu(4), Fri(5), Sat(6), Sun(7)
W.Hours: number of work hours, including study, homework, and job for living, per day.

If you don't feel to include some of the personal information (e.g. sex, age), please leave them as NA.

Contents of the final term project

As usual, this final term project should include the following components. Once again, creativity is the thing that I would value the most.

Introduction: objective, hypothesis, motivation, etc.

Data description: background, potential event(s) of interest or label(s) of interest from personal diary.

Data preprocessing: data merging, and extract relevant data to event of interest.

Data analysis: descriptive statistics, modeling, supervised learning, unsupervised learning, prediction, or others, whichever methods you think are appropriate to achieve the study objective and provide answers to the hypotheses.

Conclusion & Discussion: Strength, novelty, innovation, and limitation of the project. What was your experience of data analysis? What have you found to be most interesting and surprising? What is the future work?

Each project should be prepared with one-inch margins, in 12-point size letters and no more than 25 lines per page, double-spaced throughout. The first page should include a title, authorship, key phrases, and a one-paragraph abstract. The abstract should not exceed 200 words. Each project should not exceed **10 pages**, including title, authors, abstract, key phrases, figures and tables as well as references.

future works:

1. creativities:

    Logistic regression:

        sleep efficiency score,

        AUC-ROC,

        interaction terms

    Non-supervised method: subjects clustering: select typical features < 10 counts

        features: ….

        Apply K-means, tSNE

To dos

2. ACF autocorrelation correction
3. Normalization by centering