STATS 503 Proposal, Team 14

Team members: Chen Xie, Xun Wang, Xinye Jiang

Dataset Description:

The dataset is about the recordings of human hand muscle activity corresponding to four different hand gestures. It has four classes of motions as the response variable and the rest 64 variables which show the 8 consecutive readings of all 8 sensors as explanatory variables.

Variables	Type	Description
V65	Categorical	Response: gesture classes (rock-0, scissors-1, paper-2, ok-3)
V1-V8	Continuous	Reading 1 Sensor 1-8
V9-V16	Continuous	Reading 2 Sensor 1-8
V17-V24	Continuous	Reading 3 Sensor 1-8
V25-V32	Continuous	Reading 4 Sensor 1-8
V33-V40	Continuous	Reading 5 Sensor 1-8
V41-V48	Continuous	Reading 6 Sensor 1-8
V49-V56	Continuous	Reading 7 Sensor 1-8
V57-V64	Continuous	Reading 8 Sensor 1-8

Proposed Algorithm:

1. Data management: Clean data and deal with missing values.
2. Data analysis: Summarize the data numerically and graphically. Data transformation (PCA / ...).
3. Variable Selection: Consider the importance of the variables in random forest and the AIC/BIC in the logistic regression comprehensively to choose variables to carry on the further analysis.
4. Classification and Prediction: Explore and visualize the relationship between the response variables and the predictors (classification methods: random forest / logistic regression / LDA / QDA / ...).
5. Comparison: Compare prediction performance among classification methods

Techniques may be applied: PCA, Validation, Variable Selection, classification methods (random forest / logistic regression / …)

Source: https://www.kaggle.com/kyr7plus/emg-4#0.csv