In-class Assignment 9_2

October 29, 2014

Problem.

Psychological research indicates that an affliction called disorganized attachment disorder, developed during infancy, might be prevalent in rampage killers, and might be one of the critical factors in the development of a rampage killer. Unfortunately this disorder remains largely undiagnosed due to the difficulty of the deployment of the current diagnostic tool. The current diagnostic tool is successful in detecting the disorder based on an extensive evaluation of the 15 month old infant by highly specialized doctors over a long period of time. The evaluation results in a score, SSC, with values 0 through 10, with scores over 6 being indicative of the disorder being present.

A new diagnostic tool, cheaper, faster and easier to deploy is being proposed. The new diagnostic tool proposes the observation of some specific physical characteristics over a 60 minute period. A study was conducted to validate the diagnostic tool. Are the 14 somatic markers (physical characteristics) successful in predicting the disorder score? What are the somatic features most indicative of the disorder? The features used in predicting the SSC response are:

SOM1: Infant throws herself backwards with no awareness of support or lack thereof behind her. Infant then seeks out parent.
SOM2: Infant's hand(s) or leg(s) gesture for the caretaker; infant then/suddenly looks away and moves away
SOM3: Infant thrusts her/himself back while also seeking proximity to parent
SOM4: Infant reaches for parent while looking away, or while legs in one direction and arms in another
SOM5: When distressed, infant pushes parent away with one of extremities
SOM6: Infant runs toward parent, but falls down during approach
SOM7: Contorted facial movements
SOM8: Jerking sequential extremities movements

SOM9: Infant freezes, could be face, or extremities, or whole body.
SOM10: Distant gaze with open mouth
SOM11: Jerking or pulling away/back from parent while being comforted or in reunion
SOM12: Infant hits parent with one hand, but holds parent with the other
SOM13: Child appears confused; body movement is one of "no completion"
SOM14: Infant cries and rolls away from caretaker on her/his side

The data (SSC score and somatic markers) is presented in fdata.RData file, on Scholar/Resources/Data, and contains other demographic information as well.

Tasks:

1. Load your data
2. Split data into 90% training and 10% test sets using the method learned in Chapter 4 of Practical Data Science with R.
3. Fit the full linear regression model. Include as features, besides Som1-Som14, demographic features: age, gender, location, ethnicity, and coder. Your target variable (that you will be predicting) is SSC score.
4. Apply stepwise regression.
5. Fit the model selected by the stepwise regression.
6. Validate the model using the test set. (First predict SSC values, and then visualize using ggplot function with “predicted” along x axis and “actual” along y axis). Comment on how well the model predicts SSC scores.

Note: Use the Lect28.R and lecture notes as guidance. Use R to implement the scoring algorithm. Submit a pdf with the R code, results and interpretation of results as Inclass9_2. It will be due by Monday at 1pm, with the rest of the week’s assignments. Correct completion in class, including comments, and code, gives a 5 point bonus.