

In-class Assignment 9_1

October 27, 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 diagnostic of either 1= child has the disorder, or 0= child does not have the disorder.

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 likelihood that a child has the disorder? What are the somatic features most indicative of the disorder? The features used in predicting the likelihood of the disorder 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 (disorder status (1=present, 0=absent) and somatic markers) are presented in fdata.RData file, on Scholar/Resources/Data, and contains other demographic information as well.

Use a logistic regression scoring machine learning algorithm and the data from the study to answer our research questions from a data science perspective.

More specifically, do the following:

1. Build a first logistic model using only SOM1, SOM2, SOM3, SOM4 and SOM5. What are the important features in this model? What is the interpretation of the SOM1 coefficient?
2. Build a second model using only SOM6, SOM7, SOM8, SOM9. What are the most important features in this model?
3. Build a third model using only SOM10, SOM11, SOM12, SOM13, SOM14. What are the most important features in this model?
4. Compare the three models. Which one is better at predicting the likelihood of the disorder?

Note: Use R. Complete a pdf document with the R code, results and interpretation of results as Inclass9_1. It will be due by Monday at 1pm, with the rest of the week's assignments. Completion in class, including comments, and code, gives a 5 point bonus.