INFO 7390
Advances in Data Sciences and Architecture
Assignment 2 – Linear Models

Professor: Nik Bear Brown
Due: October 11, 2019

In this assignment you will create predictive linear models.

Generate linear models for the following:

Find a significant linear relation of your choosing in your data. Create a multivariate linear model.
Find a significant logistic linear model of your choosing in your data. Create a logistic linear model. (

Answer the following questions for all of the models:

* Is the relationship significant?
* Are any model assumptions violated?
* Is there any multi-collinearity in the model?
* In the multivariate models are predictor variables independent of all the other predictor variables?
* In in multivariate models rank the most significant predictor variables and exclude insignificant ones from the model.
* Does the model make sense?
* Does regularization help?

Q1 (5 Points)

Pick a public dataset that can be used for linear and logistic regression. You MUST get approval for your dataset and post it on piazza.

Q2 (70 Points)

Create at least three multivariate linear models. One must use step-wise regression and one must use interaction variables and one must use feature extraction by creating a new variable not in the original data set. You must properly encode the variables. You must explain in words what the model means with at least one example with realistic input.

Q3 (25 Points)

Create at least three logistic linear models. One must use interaction variables and one must use feature extraction by creating a new variable not in the original data set. You must properly encode the variables. You must explain in words what the model means with at least one example with realistic input including the probability of the dependent variable.