INFO 7390
Advances in Data Sciences and Architecture
Assignment 1 – ML Models

Professor: Nik Bear Brown
Due: January 27, 2020

TAs

Manogna Mantripragada <mantripragada.m@husky.neu.edu>
Nikunj Lad <lad.n@husky.neu.edu>
Akshaykumar Ishwarbhai Patel <patel.ak@husky.neu.edu>

ML Models

In this assignment you will use three different algorithms to create predictive models and interpret them. Find a significant relation for each algorithm of your choosing in your data. Create multivariate models.

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?
* Which independent variables are significant?
* Which hyperparameters are important?

Scoring Rubric

Did I explain my idea clearly? (10 Points)

How effective are you at explaining what you are doing? You MUST write an abstract and a conclusion.

Did I explain my evaluation clearly? (15 Points)

Just saying "accuracy" is not a clear explanation of an evaluation scheme. Clearly explain the evaluation scheme. Do the metrics make sense? You MUST explain how you are preventing overfitting.

Tables, graphs and charts must support your evaluation.

It MUST run. (5 Points)

The code must run on a laptop other than yours. There MUST be a clear README on how to run it.

Public dataset *(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.

What code is yours and what have you adapted? (10 Points)

You must explain what code you wrote and what you have done that is different. Failure to cite ANY code will result in a zero for this section.

Did I explain my code clearly? (15 Points) Your code review score will be scaled to a range of 0 to 10 and be used for this score.

Did I explain my licensing clearly? (5 Points) Failure to cite a clear license will result in a zero for this section.

Answers to listed questions (35 Points)

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?
* Which independent variables are significant?
* Which hyperparameters are important?