**CSCE 623 Spring 2020: Machine Learning. In-Class Work, Day 6**

From Chapter 3: Linear Regression Part 1: “Evaluation Criteria Cheat Sheet”

There are many ways of determining which machine learning model parameters to use and measuring the performance of a model. One of the struggles with digesting chapter 3 is the presentation of many diagnostic and decision-making measures. Build a dictionary for each of these measures. Include the following for each measure:

* Math Formula, English Name/Definition & page number in book;
* Purpose / how to use it for decision-making / how to interpret;
* Benefits / assumptions / limitations / sensitivities

Example purposes: Use to estimate coefficients; Estimate accuracy of the model; hypothesis testing

Example benefits/assumptions/limitations: Expressed in the units of Y: Sensitive to the number of predictors

RSS

MSE

TSS





RSE

*t*-statistic

*p*-value

R2

Correlation(X,Y)

F-statistic

Leverage statistic *hi*



From Chapter 3: Linear Regression Part 2: “Potential Problems”

For each of the problems listed below, explain a) what the problem is; b) how it would be detected in the model or the data and c) what should be done to resolve the problem.

1. Non-linearity of the response-predictor relationships
2. Correlation of Error Terms
3. Non-Constant Variance of error terms
4. Outliers
5. High leverage points
6. Collinearity