ML 4375 – Intro to Machine Learning – Summer 2018 – Mazidi – Exam 1 Review

General information:

* Test format: 50 questions, mostly multiple (multiple) choice, some short answer
* You will not have to produce R code but should recognize what R code does
  + 1 – Introduction to Machine Learning
  + 2 – Learning R
  + 3 – Linear Regression
  + 4 – Logistic Regression
  + 8 – kNN
  + 9 - Clustering

Understand the algorithms learned so far (linear regression, logistic regression, knn, clustering):

* a conceptual understanding of the algorithm
* familiarity with the math underpinning the algorithms
* loss and cost functions
* tendencies towards bias or variance in different situations (for supervised learning)
* usage: classification, regression, both
* category: supervised or unsupervised learning
* which metrics are used for evaluation and why
* intepreting algorithm output
* interpreting summary(), anova(), residual plots

Understand the following machine learning terminology:

* Residuals
* Regression vs. classification
* Quantitative vs. qualitative (categorical) data
* Underfitting vs. overfitting
* Bias variance tradeoff
* loss function and cost function for different algorithms
* argmin, argmax
* metrics: mse, rmse, rse, R^2, accuracy
* causation vs. correlation
* confounding variable
* gradient descent and its alpha hyperparameter
* train and test sets; validation sets; cross validation
* instance == example == observation == row
* attribute == feature == predictor (not including target)
* factors in R and dummy variables
* curse of dimensionality
* Occam’s razor
* k-fold cross validation