List of sample R codes

This repo contains the R sample codes used in the BIOL672 Computational Statistics and Data Science Methods course taught at the Rochester Institute of Technology by Dr. Gregory A. Babbitt in 2020. Most of the scripts have a hard coded path to the iris_tab.txt data file supplied with the scripts. The path will need to be modified to match the users file structure. The scripts were tested on Win 10 and Linux Mint 19. Almost all the scripts use the same data as an example (The famous Fisher's iris data set for the morphometric of 150 specimens of 3 species of iris flowers). For some of the more hardware intensive scripts a much larger simulated set of this data is used (iris_tab_big.txt - 500000 specimens).

iris_code.R – a simple introductory script showing the essential anatomy (includes data input using readtable(), column extraction using subsetting operator, proper use of assignment operator, creation of dataframe, reshaping data with library(reshape2), one-way analysis of variance(ANOVA), exporting R console to a text file using sink(), plotting with library(ggplot2), and creating multipanel plots with library(grid).

RunMyRfromPERL.pl – a perl script demonstrating control of iris_code.R with a system call, and sending single line R commands via an open pipe.

A variety of multivariate stats functions are demonstrated
myIRIS_MLR.r – multiple linear regression
myIRIS_MLR_setosa.r - multiple linear regression on single species
myIRIS_ANCOVA.r – analysis of covariance
myIRIS_MANOVA.r – multivariate ANOVA
myIRIS_PCA.r – principal components analysis
myIRIS_FACT.r – factor analysis
myIRIS_KMEANS – k means clustering
myIRIS_Emclust.r – model based clustering with EM algorithm using library(mclust)
myTEACHER_FACT.r – exploratory factor analysis on survey data (GoodTeachSurvey_tab.txt)
myIRIS_GMM.r – fit Gaussian Mixture Model with EM algorithm
myIRIS_BIC.r – multi-model inference with Bayesian Information Criterion (includes GMM)
kernel_regression_iris.r – nonlinear regression

A variety of machine learning classification algorithms (to predict iris species)
myIRIS_KNN.r – K nearest neighbors
myIRIS_NBe1071.r – naive Bayes
myIRIS_LDA-QDA.r – linear and quadratic discriminant analysis
myIRIS_SVMe1071.r – support vector machine (untuned kernel function)
myIRIS_SVMkernlab.r - support vector machine (parameter tuning of kernel function)
myIRIS_SVMmultithread_cpu.r - support vector machine (distributed to multiple CPU using caret)
myIRIS_SVMsingle_gpu.r - support vector machine (distributed to GPU via LiquidSVM algorithm)
myIRIS_NeuralNet.r – simple neural network
myIRIS_NeuralNetBig.r – neural network (distributed to multiple CPU using caret)
myIRIS_randomforest.r – random forest
myIRIS_adaboost.r – adaptive gradient boosting

Two image classification scripts via deep learning (using keras and tensorflow libraries)
myFASHION_DeepNet.r – convolutional neural network (70,000 28x28 pixel grayscale images)
myCIFAR10_DeepNet.r - convolutional neural network (60,000 fullscale color images)