VIA 507E Mini Exam 4
Due Date Jan 5th

You are given a dataset (MiniExam4Dataset.csv) that includes 14 features that represents clinical conditions of 500 ICU patients and target variable death that represents whether the patient died (=1) in the ICU or discharged alive (=0).
(a) Split the data set into a training set and a test set (80% Training, 20% Test)
(b) Standardize your features.
(b) Use cross-validation to select the best method and the best set of parameters. Try Regularized Logistic Regression (both L1 and L2 penalties and different C values), KNN classifier (different numbers of neighbors you believe to be reasonable). BE CAREFUL that the best model should be selected using cross validation hence you should never evaluate different methods using the test set. Also, be very careful that the standardization needs to be carefully done during cross validation not to end up with data snooping (recall the pipe approach discussed in the class).
(c) Once you decide on the final method and the set of best parameters, refit your model on the standardized training set and evaluate the performance (accuracy) on the standardized test set.
(d) Provide the test confusion matrix.