November 6, 2020

Assignment 2

CS 624, Fall 2020

Due on November 13, 2020

Problem 1. We are interested in finding and assessing the effect sizes of the important predictors of coronary heart disease. Analyze the wcga data (available on Canvas) to answer this question. In particular, perform stepwise model building for a model with main effects only and one with two-way interactions using forward selection or backward elimination, find the area under the ROC curves and comment on the classification accuracy of the final models, use the Hosmer-Limeshaw test for goodness-of-fit and comment on the adequacy of the final models, give interpretation of all regression coefficients, find the sensitivity and specificity of the model for a range of thresholds spanning the interval [0,1], plot the ROC curves using these values, find the optimal thresholds that simultaneously maximizes sensitivity and specificity, add a triangular symbol to the ROC curve at the point with maximum sensitivity and specificity. Lastly, write a short paragraph summarizing your methods and results.