BMTRY 790, Spring 2023 Assignment 4

Exercises

1. Evaluation of risk of death among patients with hepatitis may help physicians manage a patient's treatment regimen. Hepatitis is a disease that primarily impacts the liver and thus different measure of liver function, including lab tests and results of clinical examination can often provide information about how "sick" a patient is. Additionally information about gender, age, wellness, and therapies also may play a role in the disease. Due to the number of factors that may determine a person's risk of death due to hepatitis, it is of interest to develop a predictive model that would estimate risk of death within one year. A clinician has collected a data set of clinical and demographic characteristics he believes may be associated with a person's risk of death. Download the Hepatitis data from the class website and split the data into 2/3 training and 1/3 test.

   i. Build prediction models on the training data for the response "Alive1Year" using the following approaches: GAM using loess, GAM using splines, and CART. Provide a discussion of your approach to tuning each model. Specifically, describe the approach used for tuning each model and final parameters selected for fitting the final model.

   ii. Evaluate and compare the prediction performance on both training and test data for each model. Which model appears to be the most generalizable.

   iii. For the different models, provide any graphical information you deem relevant to helping your clinical investigator understand these models. Please provide discussion of that your graphics say about your model or the relationship between your predictors and outcome.

2. Systemic lupus erythamatosus (SLE) is a multisystem autoimmune disorder with 9:1 greater prevelance in females to males and 4:1 prevelance in African Americans (AA) to whites. The incidence of SLE among Gullah AA from the sea coast islands of North Carolina, South Carolina and Georgia is even higher that other AA poulations in the US. The Gullah AA have relatively low geneitc admixture with whites and therefore are still genetically similar to Africans in Sierra Leone, the predominant ancestral home. However, despite the genetic simialrity of the two populations, Sierra Leoneans have little to no indicence of SLE. It has been hypothesizes that a combination of geneic and environmental exposures may partially explain the higher than expected rates of SLE in the Gullah AA. A rheumatologists conducted a case-control study in Gullah AA with and without SLE to evaluate potential interactions between single nucleotide polymorphisms (SNPs) on the ITGAM gene and smoke exposure. Information collected on the participants included 67 SNPs on the ITGAM gene , smoking status (curent, former, never), passive smoke exposure as a child, and passive smoke exposure as an adult.

   i. Fit a logic regression model using the classification model setting with SLE status as the outcome. Describe how the model was tuned and what parameters where used in fitting the model. Also provide a plot of the final model and describe any potential interactions between SNPs on the ITGAM gene and different smoke exposures that were identified.

   ii Fit a logic regression model using the logistic/deviance model setting with SLE status as the outcome. Describe how the model was tuned and what parameters where used in fitting the model. Also provide a plot of the final model and describe any potential interactions between SNPs on the ITGAM gene and different smoke exposures that were identified. (HINT: LR under the logistic setting models the probability the outcome is a 0. Be sure to change the outcome to mdoel the probability of SLE.)

   iii Compare the models fitted in (i) and (ii) and comment on how they differ in terms of interpretation.