Healthcare Case Studies
Project Description

Instructions

In this project, you will create a linear regression model for predicting total post-index cost and a logistic regression model for predicting the pdc flag for diabetic patients.

The first step in creating the models is variable selection. As you select your variables, keep in mind that smaller models are easier to interpret and less prone to overfitting.

Utilize scatter and box-plots, statistics tests, and Stepwise, Ridge, and Lasso regression techniques to for your model implementation. You should also compare the performance of different models on the validation sets and assess the final performance by measuring the out-of-sample error.

For the linear regression model, you will need to perform residual analysis to check model assumptions.

Paper

Your paper should contain the following sections:

1. Introduction (Outline the problem and its importance. Provide any required domain-specific background. Provide a literature review. Summarize previous approaches to this problem or problems similar to this, their results and limitations)
2. Dataset (Describe the dataset and its origin. Explain any pre-prosessing done by you on the dataset)
3. Methodology (Describe how you employed R/Python for exploratory data analysis to identify relationships between variables in your data set, how you applied linear and nonlinear regression models, how you assessed the validity of a proposed models and adjust for the violation of key model assumptions or the presence of unimportant variables, etc.)
4. Results (Report the results of your analysis, providing appropriate tables/plots/figures. Compare your results with previous work.
5. Conclusion (Interpret your results. What have you learned and contributed. Suggest topics for further work.)

Make sure to spell-check your paper before submitting.

WebEx Demo

Your group will do a demo via WebEx where you'll share your screen and explain your methodology and results, and answer instructor's questions. Please see schedule on Canvas for your presentation time.