Group 6

Summary:

The project compares 4 kinds of model, 1. best subset selection with AIC 2. best subset selection with BIC 3. Ridge regression with CV 4. Lasso regression with CV, based on the Titanic dataset from Kaggle. Based on the accuracy, the project analyzed different properties of the method mentioned above. For the final models, it finds that Pclass, Sex and Title appear in all 4 models. Also, it suggests that female and people who hold higher class tickets are more likely to survive.

Strength:

This project discusses various kinds of feature selection methods and their characteristics separately. For example, the results suggest that BIC penalizes more heavily for additional parameters and LASSO often shrinks coefficients to be identically 0. Moreover, even though the methods are different, the final models for each method appears to obtain similar features in the dataset.

Weakness:

The project is unclear in some parts. For example, it only applies cross validation to logistic regression with Ridge and Lasso regularization yet does not apply cross validation to models selected by AIC and BIC. The report doesn't explain why. Also, the author converts some numerical variables to categorical without giving specific reasons, say does it improve prediction accuracy or reduce computational complexity? Converting a numerical variable to a categorical variable with several levels does not necessarily reduce computational complexity.

Evaluation on quality of writing (1-5): 3

The poster presents the logic and results clearly with graphs that straightly show the properties of certain models. However, the introduction part seems less relevant to the project. It makes the reader confuse about the major goal of the project.

Evaluation on presentation (1-5): 5

I think the speaker did great in slides design and presentation delivery. The PowerPoint is well organized, and the speaker talks clearly about his project.

Evaluation on creativity (1-5): 3

In the data processing part, the project discusses the information contained in the attribute, Name, which is often dropped in many other projects on this dataset. In addition, the author tried categorization on several numerical variables.

Confidence on your assessment (1-3) :3