**Homework 7. Lasso vs. Ridge Lending Dataset Modeling**

Author: Rohit V. Akole

UConn ID: 3133252

University of Connecticut

OPIM5604 Predictive Modeling

Professor: Pankaj Prakash

Abstract

This homework is about the Lending Dataset based on year 2014 data. I used the previously cleaned dataset to perform General regression methods (Lasso and Ridge) on the model with variables used in the homework 5 logistic regression model.

## Lasso Generalized Regression Model

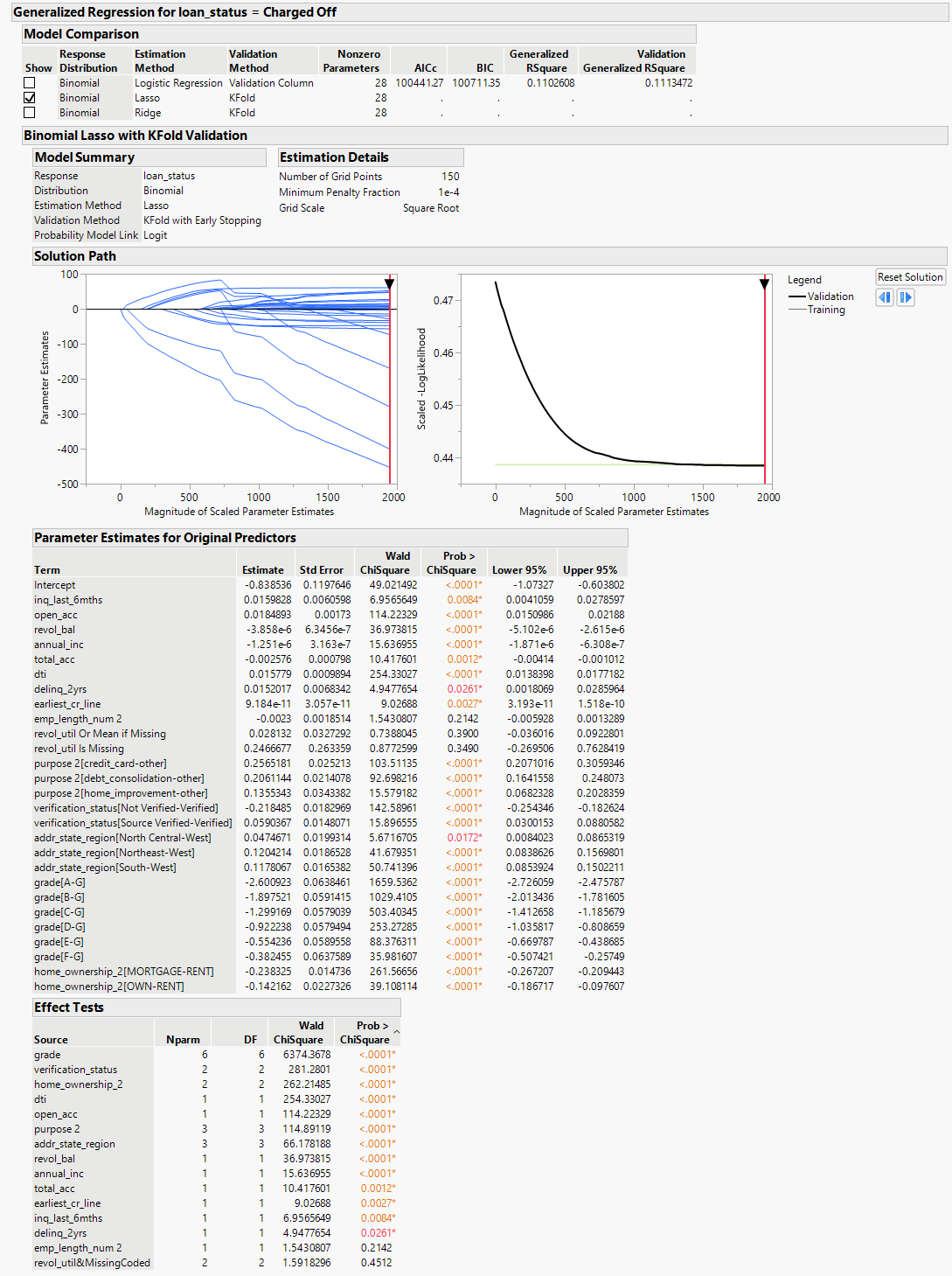


Figure 1. Penalized Lasso Generalized Regression Model

As per the video tutorial, I kept all the variables taken in the 1st iteration of the logistic regression model from homework 5. In Lasso, I kept the solution path threshold to default. However, in this model, I don’t see any of the variables with a 0 coefficient. 2 of the variables still have p-values more than 0.05.

## Ridge Generalized Regression Model



Figure 2. Penalized Ridge Generalized Regression Model

As per the video tutorial, I kept all the variables taken in the 1st iteration of the logistic regression model from homework 5. In Ridge, I kept the solution path threshold to default. The coefficient is minimized and almost all the variables have p-values less than 0.05 except emp\_length\_num 2.

## Lasso Penalized Model vs. Ridge Penalized Model vs. Old Model (AUC Comparison)

A screen shot of a computer

Description automatically generated

Figure 4. Old Model from Homework 5

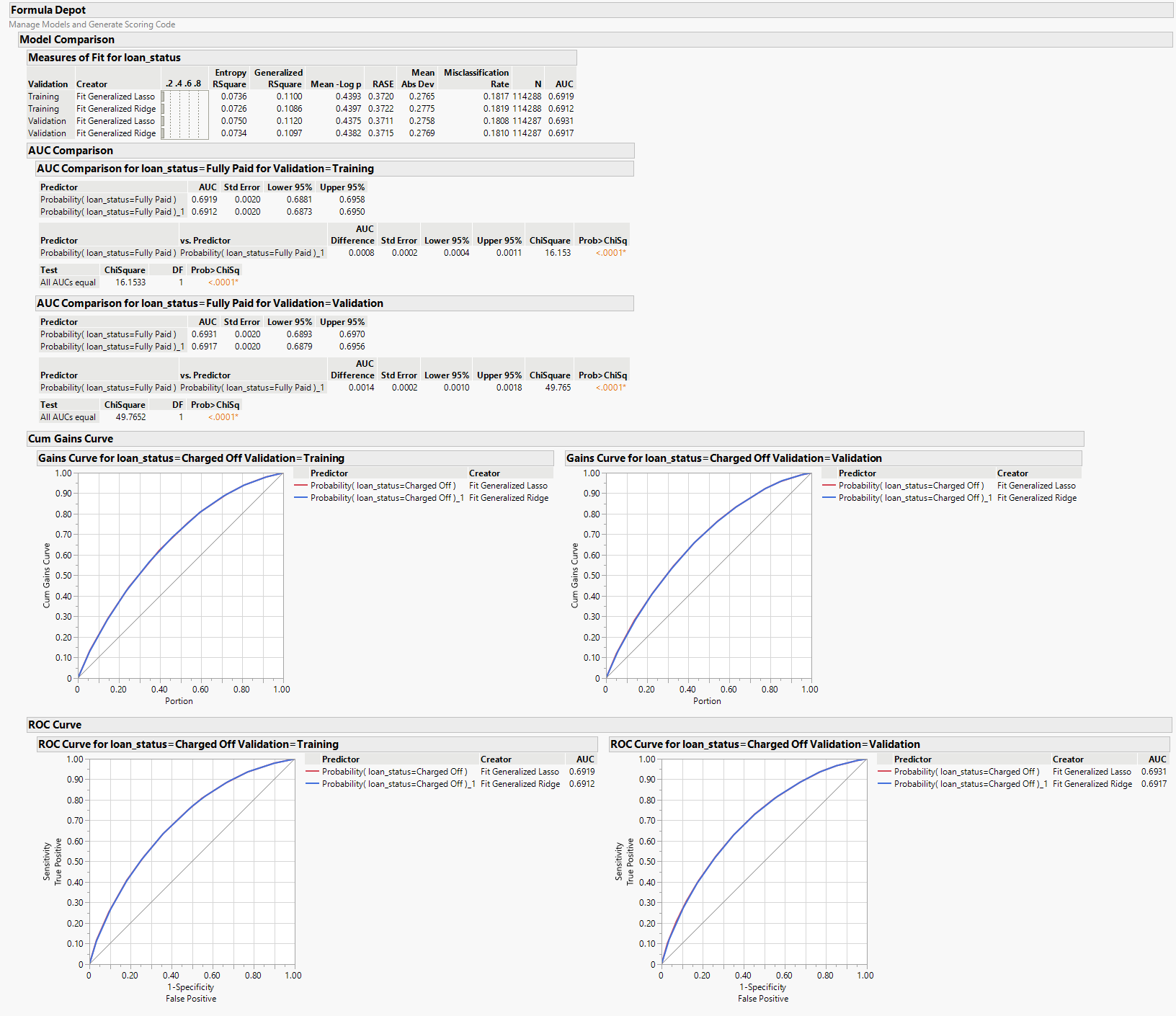


Figure 3. Penalized Model Comparison and AUC

|  |  |  |  |
| --- | --- | --- | --- |
| **Validation Dataset** | **Old Model** | **Lasso Penalized Model** | **Ridge Penalized Model** |
| **AUC** | 69.26% | 69.31% | 69.17% |
| **Entropy R2** | 7.45% | 7.50% | 7.34% |
| **Generalized R2** | 11.13% | 11.20% | 10.97% |
| **RMSE** | 37.12% | 37.11% | 37.15% |

According to the above table for the validation dataset, all 3 models present similar results. But as per the AUC, Lasso (69.31%) is performing a little better than Old Logistic Model (69.26%) and Ridge Penalized Model (69.17%). RMSE is also at lowest in the Lasso Model at 37.11%. R2 is also the highest for the Lasso Penalized Model. Overall, the Lasso Penalized Model is the best model of all 3.