

05-analysis.Rmd

Results

Before we look at regression performance, lets take a look at how the estimated regression coefficients differ for each regression:

```
knitr::kable(betas, digits = 3, caption = 'A table of coefficients', width=50)
```

Table 1: A table of coefficients

Coefficient Names	Coefficients for PLS Regression	Coefficients for Least Squares	Coefficients for Ridge Regression
Income	-0.605	-0.595	-0.545
Limit	0.677	1.060	0.676
Rating	0.671	0.279	0.612
Cards	0.048	0.053	0.043
Age	-0.019	-0.007	-0.029
Education	-0.001	-0.012	-0.002
Gender.Male	0.040	0.010	-0.057
Gender.Female	-0.040	NA	-0.073
Student.Yes	0.881	0.920	0.762
Married.Yes	-0.024	0.000	-0.034
Ethnicity.Asian	0.066	0.026	0.029
Ethnicity.Caucasian	0.004	0.013	0.010

As one can see with the above table and plot, Lasso Regression provides us a sparse estimator. In other words, Lasso Regression provides a solution in which multiple beta coefficients are zero. Although Ridge does not provide a sparse solution, it does tend to shrink the beta coefficients towards zero. This is why both Lasso and Ridge are called shrinkage methods.

Thanks to the above visual, it's also easy to tell that there are four variables that have a strong relationship with the response variable (Credit Card Balance). These variables have beta coefficients that are large in magnitude and they are Income, Limit, Rating, and Student. The Limit, Rating, and Student variables are strongly correlated with Credit Card Balance while the Income variable is significantly negatively correlated with Credit Card Balance.

Now that we have a better understanding of each of these regression methods, lets look at the data in order to see how each regression performed:

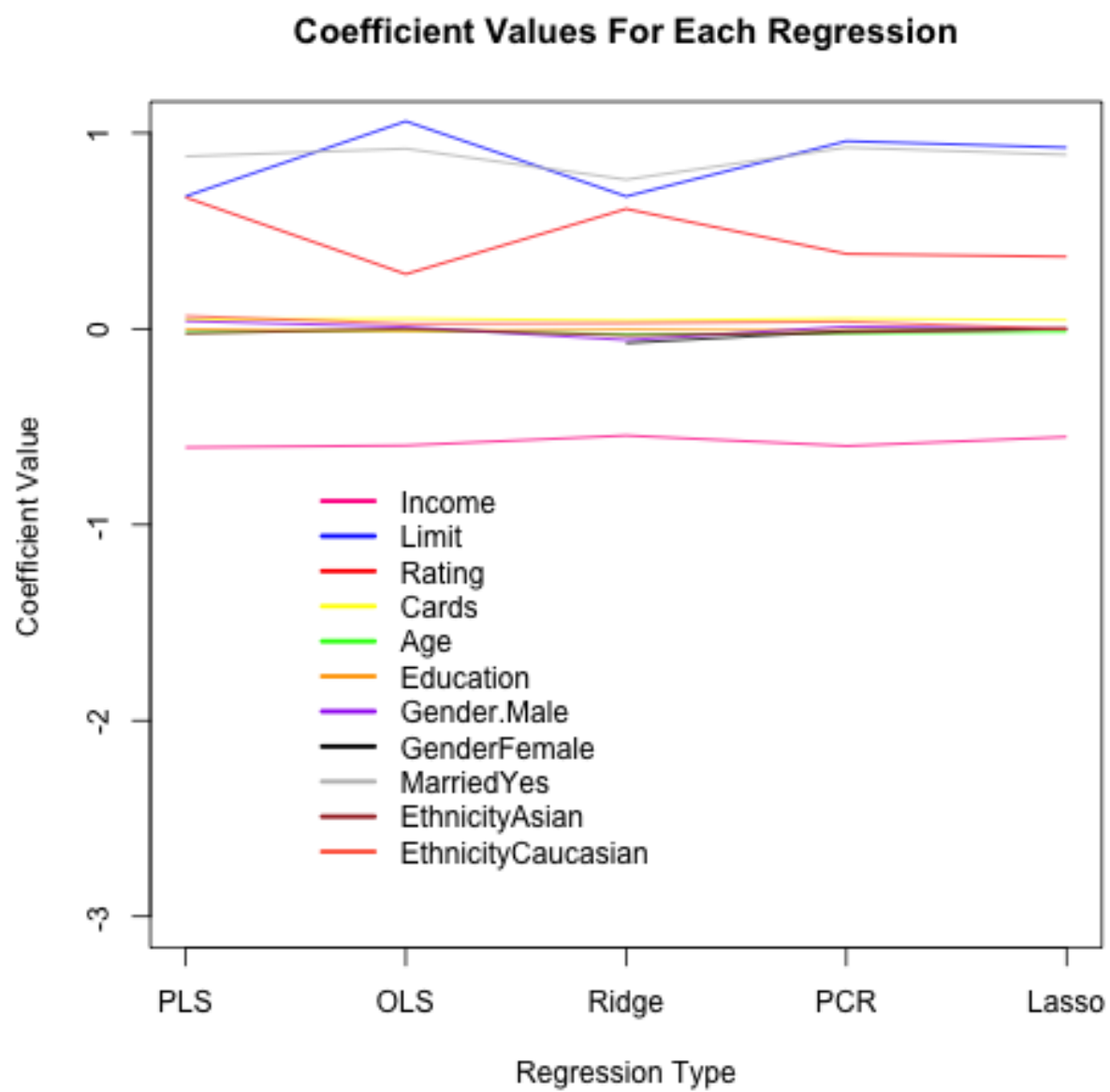


Figure 1: Coefficient Plot

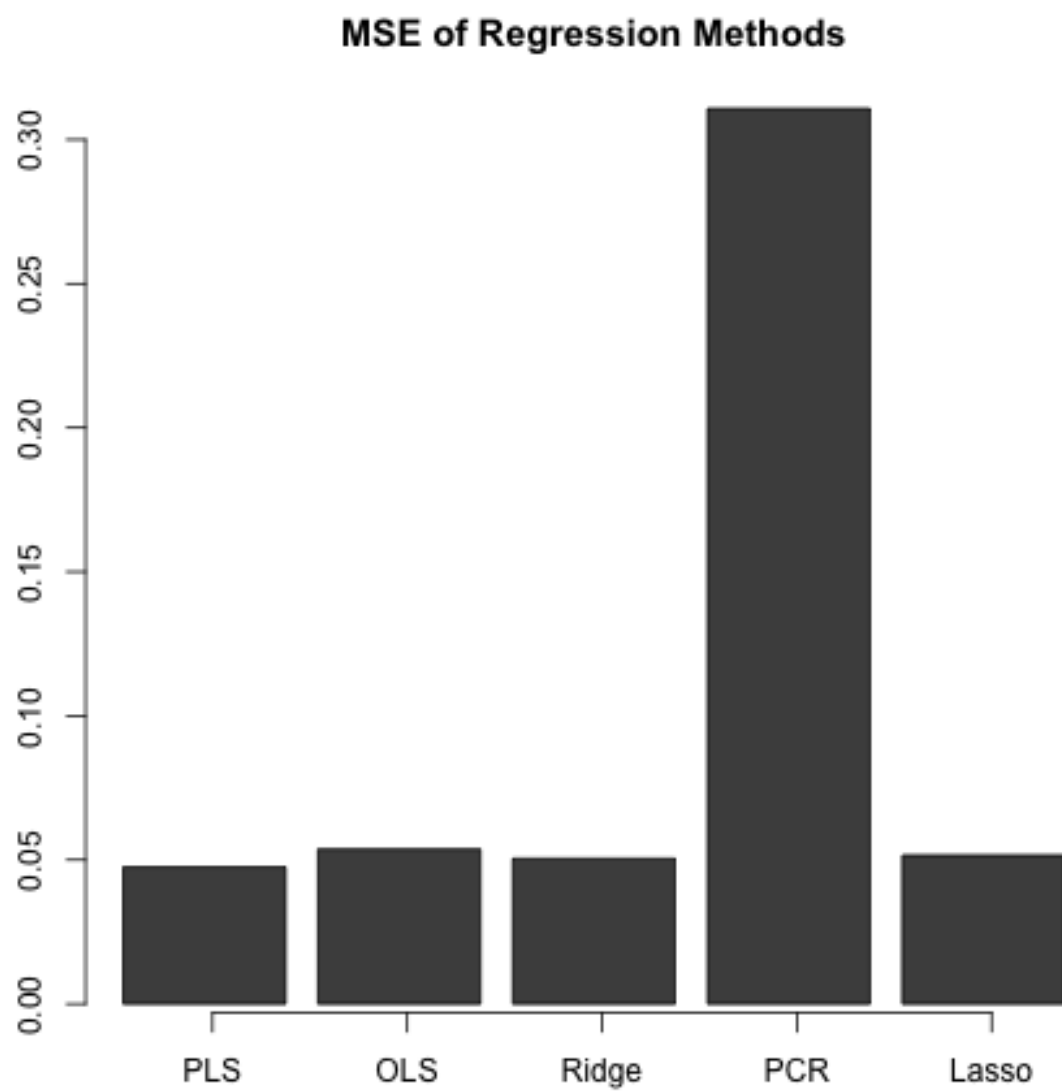


Figure 2: MSE of Regressions