**Homework 3**

**Linear Regression, Ridge regression, Lasso regression and** **Elastic Regression**

**Name:** Krishna Prasad Palvai

**Student ID:** 922672795

Imported c1\_bdhs data into python and observations with missing values are removed.

Response variable in this regression analysis is RCA. Age is selected as it can be predicted using height and weight.

Predictor variables are SW, MOI, YOI, DOI\_CMC, RMOB, RYOB, RDOB\_CMC, Region, Has\_Radio, Has\_TV, Religion, WI, MOFB, YOB, DOB\_CMC, DOFB\_CMC, AOR, MTFBI, DSOUOM\_CMC, RW, RH and RBMI.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter Estimate | Linear Regression | Ridge Regression | Lasso Regression | Elastic Net |
| Intercept | 81.577318 | 81.58048 | 111.3244 | 111.5135 |
| SW | -5.61E-09 | -5.61E-09 | 1.86E-08 | 1.87E-08 |
| MOI | 0.037055436 | 0.037053 | 0 | 0 |
| YOI | -3.05E-16 | 0 | 0 | 0 |
| DOI\_CMC | 0.037055436 | 0.037053 | 0 | 0 |
| RMOB | 0.012759428 | 0.012759 | 0 | 0 |
| RYOB | -0.007772092 | -0.00777 | 0 | 0 |
| RDOB\_CMC | -0.080505673 | -0.08051 | -0.08252 | -0.0825 |
| Region | -0.002297346 | -0.0023 | 0 | 0 |
| Has\_Radio | -0.006725393 | -0.00669 | 0 | 0 |
| Has\_TV | 0.016039969 | 0.016001 | 0 | 0 |
| Religion | -0.015971 | -0.01593 | 0 | 0 |
| WI | -0.000817903 | -0.00081 | 0 | 0 |
| MOFB | -0.002057612 | -0.00206 | 0 | 0 |
| YOB | -0.002772388 | -0.00277 | 0 | 0 |
| DOB\_CMC | 0.000252167 | 0.000252 | 0 | 0 |
| DOFB\_CMC | -0.001982133 | -0.00198 | -0.00051 | -0.00054 |
| AOR | 0.018420877 | 0.018376 | 0 | 0 |
| MTFBI | -0.000176164 | -0.00018 | -0.00011 | -0.00017 |
| DSOUOM\_CMC | -0.000308255 | -0.00031 | -0.00012 | -0.00025 |
| RW | 2.28E-05 | 2.28E-05 | 2.11E-05 | 2.68E-05 |
| RH | -1.54E-05 | -1.54E-05 | -3.28E-05 | -4.12E-05 |
| RBMI | -1.23E-05 | -1.22E-05 | 5.98E-06 | 8.28E-06 |

* Observed that coefficients has shrinked after application of Ridge regularization compared to linear regression model. Religion and YOI estimates has shrunk.
* From application of lasso regression, dimensionality is reduced to 8 from 22 in linear regression model. Regularization parameters are selected based on the K-Fold Cross validation. Regularization parameters with lowest mean square error are chosen.
* From application of elastic net regression, dimensionality is reduced to 8 from 22 in linear regression model. Compared to lasso model, features in elastic net regression are the same. Comparatively coefficient estimates of elastic net is greater than coefficient estimates of Lasso model.