Elements of Machine Learning
Assigment 3 - Problem 5

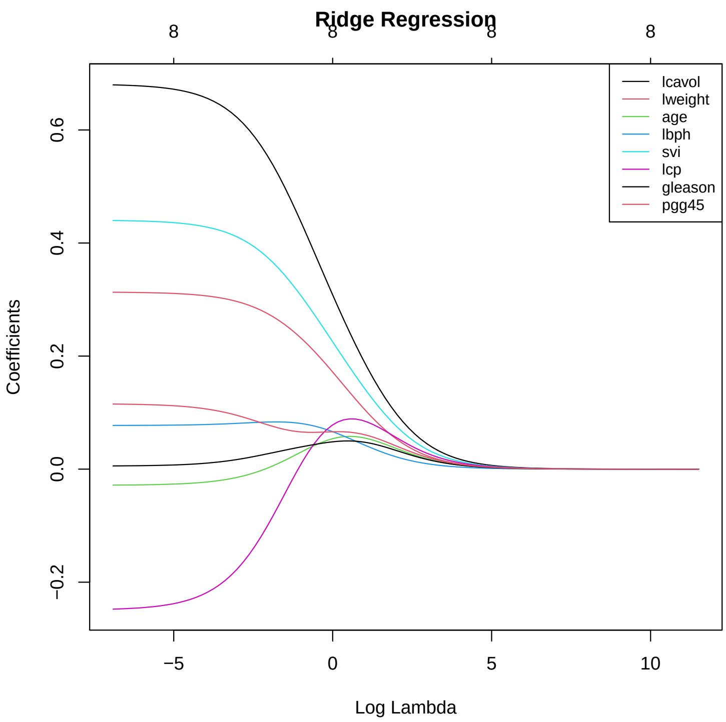
Sangeet Sagar(7009050), Philipp Schuhmacher(7010127)
{sasa00001,phsc00003}@stud.uni-saarland.de

January 7, 2022

Problem 6 (P, 20 Points)

(4P) Use the training set to fit ridge regression models and generate a plot showing the values of the coefficients in relation to the parameter (cf. Figure 6.4, p. 238, ISLR). What can you observe?

It can be observed that when λ is very small close to 0 (to the extreme left), none of the predictors are penalized. Put in simple words, they are same as least squares coefficient when $\lambda = 0$. But as λ (or $\log(\lambda)$) increases, the ridge regression penalization starts to come into effect and the coefficients start to shrink. But note that they are never exactly 0, rather they almost tend to 0.

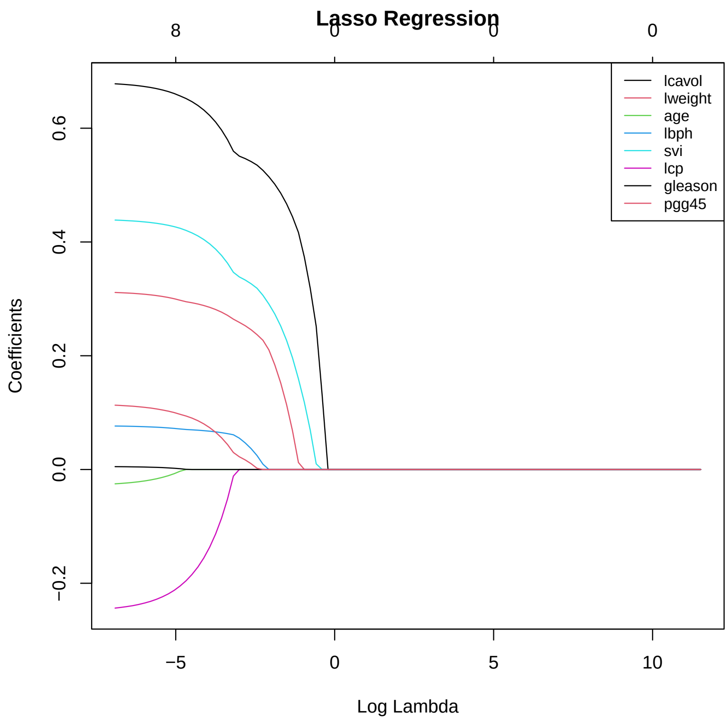


(4P) Perform 10-fold cross-validation on the training set to determine the optimal value for for the ridge regression model. Report train and test error measured in MSE for this .

```
Optimal value of lambda:  0.3853529
MSE on train set: 0.5221277
MSE on test set: 0.4621843
```

(4P) Use the training set to fit lasso models and generate a plot showing the values of the coefficients in relation to the parameter (cf. Figure 6.6, p. 242, ISLR). What can you observe in comparison to the plot generated in 2. (make at least 2 observations)?

1. When λ (or $\log(\lambda)$) is small the coefficients are almost same as least square coefficient.
2. As λ increases the magnitudes of the coefficients reduces.
3. With $\lambda \sim 1$ all predictors shrink to exactly 0 because of lasso penalty.
4. These regularization technique don't perform feature selection rather they shrink magnitudes of coefficients to reduce their contribution towards the response.



(5P) Perform 10-fold cross-validation on the training set to determine the optimal value for in the lasso. Report train and test error measured in MSE for this . How many and which features are used? Compare this to the coefficients determined for ridge regression in 3.

```
Optimal value of lambda:  0.01963041
MSE on train set: 0.4798182
MSE on test set: 0.4495713
```

(2P) Compare the performance of the best models generated in 3. and 5. . Which model would you choose and why? What alternative model could have been used?
The selection of best model should be based on test MSE. For Ridge, we observe the test MSE to be 0.46 while for lasso, test MSE is 0.44. Hence I would prefer lasso due to its lower MSE.
Few other alternative models could be forward, backward selection or bayesian regression where the penalty coefficient λ is a random variables and is estimated from the data rather than setting it manually. It has the advantage that it adapts data at hand.

References