Conclusion

Method	Est. RMSLE	Actual RMSLE	Notes
Forward stepwise	0.14826	0.29415	125 dimension; selected using adjusted R^2
Forward stepwise	0.18468	0.22928	71 dimension; selected using Mallow's Cp
Forward stepwise	0.24152	0.22727	39 dimension; selected using Bayes' IC
Forward stepwise	0.14963	0.22282	199 dimension; selected using cross validation
Backward stepwise	0.14081	0.21879	133 dimension; selected using adjusted R^2
Backward stepwise	0.16272	0.21734	87 dimension; selected using Mallow's Cp
Backward stepwise	0.20075	0.23554	43 dimension; selected using Bayes' IC
Backward stepwise	0.14970	0.22234	184 dimension; selected using cross validation
Mixed stepwise	0.15639	0.27081	99 dimension; selected using adjusted R^2
Mixed stepwise	0.17950	0.20876	41 dimension; selected using Mallow's Cp
Mixed stepwise	0.20814	0.23318	79 dimension; selected using Bayes' IC
Mixed stepwise	0.20407	0.22880	71 dimension; selected using cross validation
Ridge regression	0.13922	0.18548	Lambda=15921.10313
Lasso	0.13168	0.19808	Lambda=376.45280
Principal components regression	0.55844	0.19129	161 components
Partial least squares	0.56133	0.18627	26 components
K-nearest neighbors	0.22252	0.38399	K=6; using all predictors
K-nearest neighbors	0.39730	0.37080	K=14; best predictor
K-nearest neighbors	0.39528	0.36765	K=16; best 2 predictors
K-nearest neighbors	0.39357	0.35547	K=100; best 3 predictors

The following table is sorted by estimated test error. Recall that all estimated test errors were calculated using 5-fold cross validation.

Method	Est. RMSLE	Actual RMSLE	Notes
Lasso	0.13168	0.19808	Lambda=376.45280
Ridge regression	0.13922	0.18548	Lambda=15921.10313
Backward stepwise	0.14081	0.21879	133 dimension; selected using adjusted R^2

Method	Est. RMSLE	Actual RMSLE	Notes
Forward stepwise	0.14826	0.29415	125 dimension; selected using adjusted R^2
Forward stepwise	0.14963	0.22282	199 dimension; selected using cross validation
Backward stepwise	0.14970	0.22234	184 dimension; selected using cross validation
Mixed stepwise	0.15639	0.27081	99 dimension; selected using adjusted R^2
Backward stepwise	0.16272	0.21734	87 dimension; selected using Mallow's Cp
Mixed stepwise	0.17950	0.20876	41 dimension; selected using Mallow's Cp
Forward stepwise	0.18468	0.22928	71 dimension; selected using Mallow's Cp
Backward stepwise	0.20075	0.23554	43 dimension; selected using Bayes' IC
Mixed stepwise	0.20407	0.22880	71 dimension; selected using cross validation
Mixed stepwise	0.20814	0.23318	79 dimension; selected using Bayes' IC
K-nearest neighbors	0.22252	0.38399	K=6; using all predictors
Forward stepwise	0.24152	0.22727	39 dimension; selected using Bayes' IC
K-nearest neighbors	0.39357	0.35547	K=100; best 3 predictors
K-nearest neighbors	0.39528	0.36765	K=16; best 2 predictors
K-nearest neighbors	0.39730	0.37080	K=14; best predictor
Principal components regression	0.55844	0.19129	161 components
Partial least squares	0.56133	0.18627	26 components

The following table is sorted by actual test error.

Method	Est. RMSLE	Actual RMSLE	Notes
Ridge regression	0.13922	0.18548	Lambda=15921.10313
Partial least squares	0.56133	0.18627	26 components
Principal components regression	0.55844	0.19129	161 components
Lasso	0.13168	0.19808	Lambda=376.45280
Mixed stepwise	0.17950	0.20876	41 dimension; selected using Mallow's Cp
Backward stepwise	0.16272	0.21734	87 dimension; selected using Mallow's Cp
Backward stepwise	0.14081	0.21879	133 dimension; selected using adjusted R^2
Backward stepwise	0.14970	0.22234	184 dimension; selected using cross validation
Forward stepwise	0.14963	0.22282	199 dimension; selected using cross validation

Method	Est. RMSLE	Actual RMSLE	Notes
Forward stepwise	0.24152	0.22727	39 dimension; selected using Bayes' IC
Mixed stepwise	0.20407	0.22880	71 dimension; selected using cross validation
Forward stepwise	0.18468	0.22928	71 dimension; selected using Mallow's Cp
Mixed stepwise	0.20814	0.23318	79 dimension; selected using Bayes' IC
Backward stepwise	0.20075	0.23554	43 dimension; selected using Bayes' IC
Mixed stepwise	0.15639	0.27081	99 dimension; selected using adjusted R^2
Forward stepwise	0.14826	0.29415	125 dimension; selected using adjusted R^2
K-nearest neighbors	0.39357	0.35547	K=100; best 3 predictors
K-nearest neighbors	0.39528	0.36765	K=16; best 2 predictors
K-nearest neighbors	0.39730	0.37080	K=14; best predictor
K-nearest neighbors	0.22252	0.38399	K=6; using all predictors

Shrinkage

The Lasso method had the lowest estimated test error and the 4th lowest actual test error. This model wasn't as heavily penalized as the ridge method (it had a lower λ), meaning it had fewer coefficients that were equal to or near zero. The ridge regression had the second lowest estimated test error and the lowest actual test error. It was much more heavily penalized than the Lasso, but because its coefficients never reach zero, it was still a high dimensional model. Overall, the shrinkage methods performed the best in terms of estimated test error and best or second-best in terms of actual test error. In the case where I wouldn't be able to see the true test error, I would be highly inclined to pick either of these models for their interpretability and low estimated test error.

Subset Selection

The subset selection methods all performed moderately well in terms of both estimated and actual test error. It is notable that the higher dimensional methods had lower estimated test errors but there isn't a very clear pattern between dimensionality and actual test error. This method performed the second best in terms of estimated test error and the third best in terms of actual test error. These methods have good interpretability and are easy to run, making them good contenders as final models.

Dimension Reduction

The dimension reduction methods performed very poorly in terms of estimated test error but were the 2nd and 3rd best in terms of actual test error; I am not sure why this happened. Dimension reduction can be hard to interpret and they showed poor estimated errors, so I would be disinclined to use these methods as a final model if I didn't know that their true test errors were so good.

K-Nearest Neighbors

While the estimated test errors for the KNN methods were mediocre, their actual test errors were abysmal; for the K=6 case with all predictors, this is expected because $p > 4$. However, even with 1, 2, or 3 predictors, this method still performed poorly.