

p8106_hw1_jsg2145

Jared Garfinkel

2/18/2020

```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )

## See spec(...) for full column specifications.
```

Part a

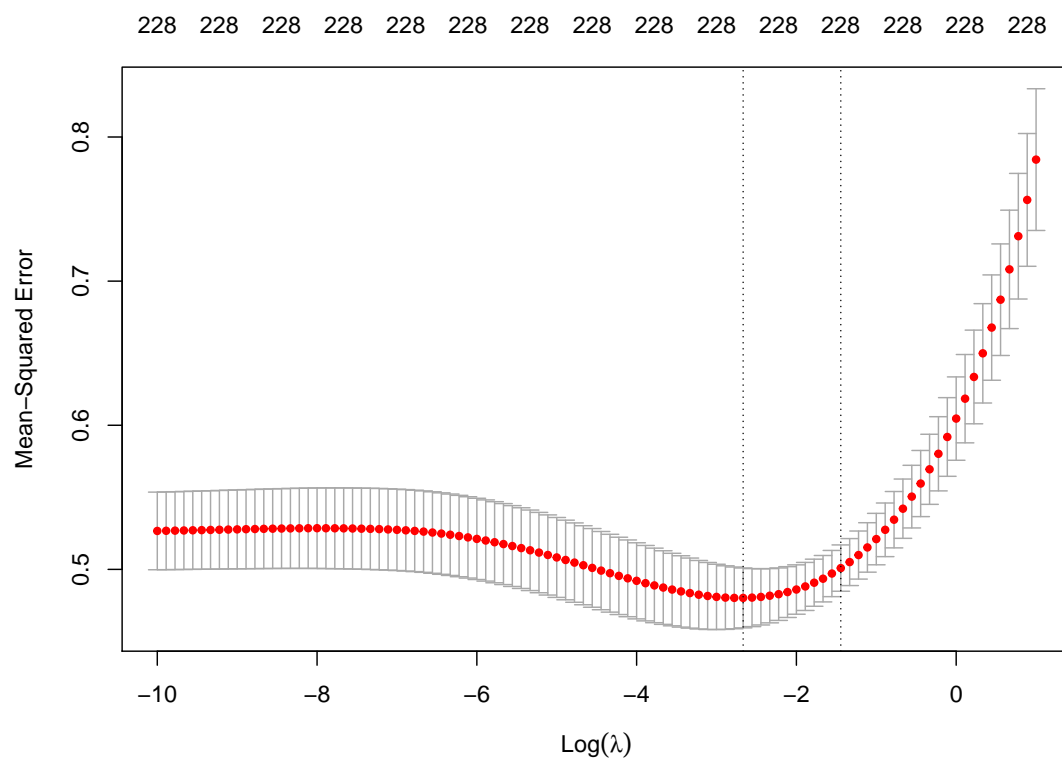
```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )

## See spec(...) for full column specifications.
```

The MSE for the linear model is 0.556.

Part b

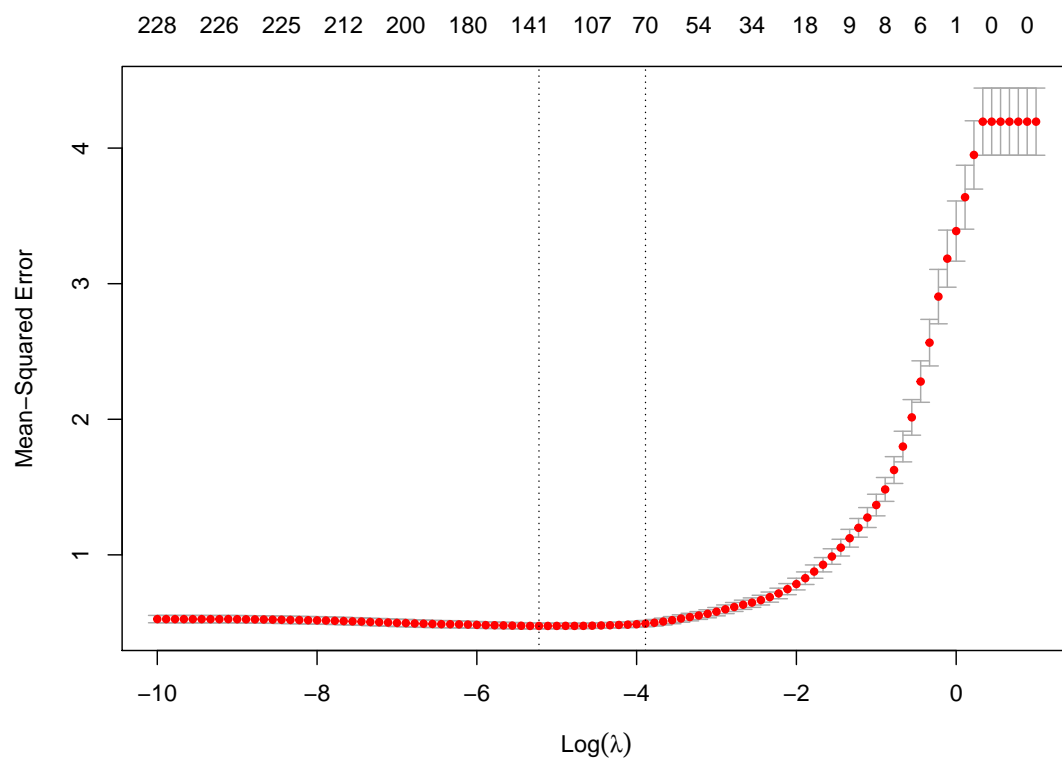
```
## [1] 229 100
```



The lambda from the ridge regression is 0.069.

The MSE from the ridge model is 0.512.

Part c

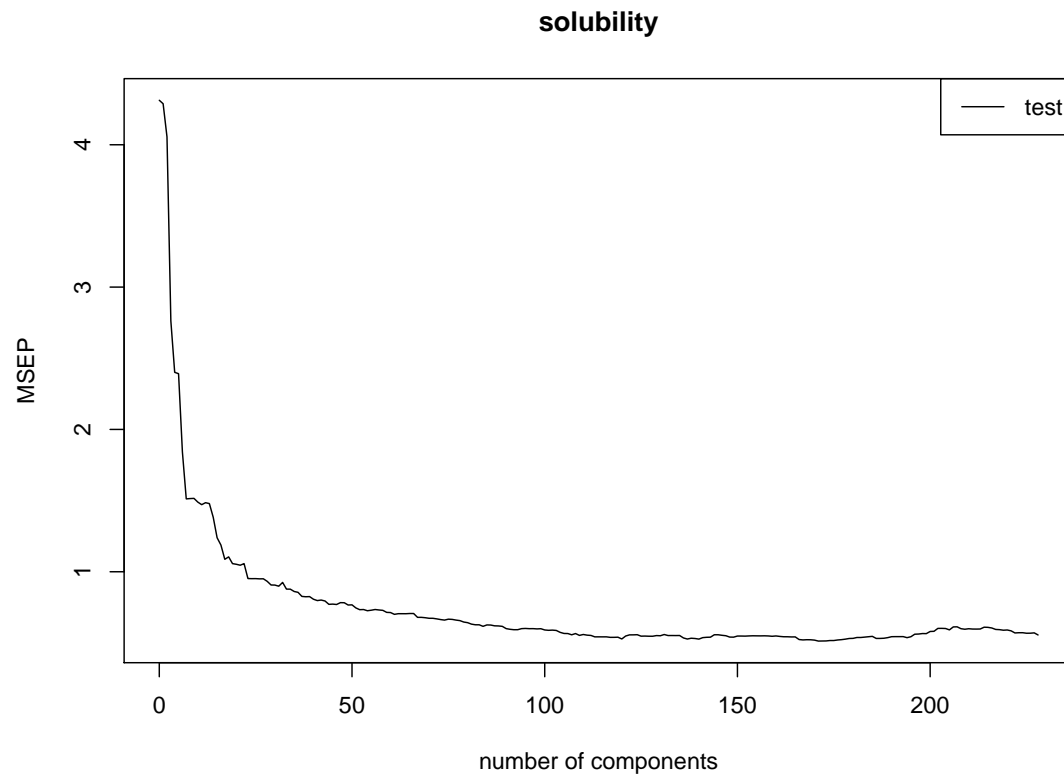


The lambda returned by the lasso regression is 0.0054.

There are 141 non-zero coefficient estimates.

The MSE of the lasso model is 0.495.

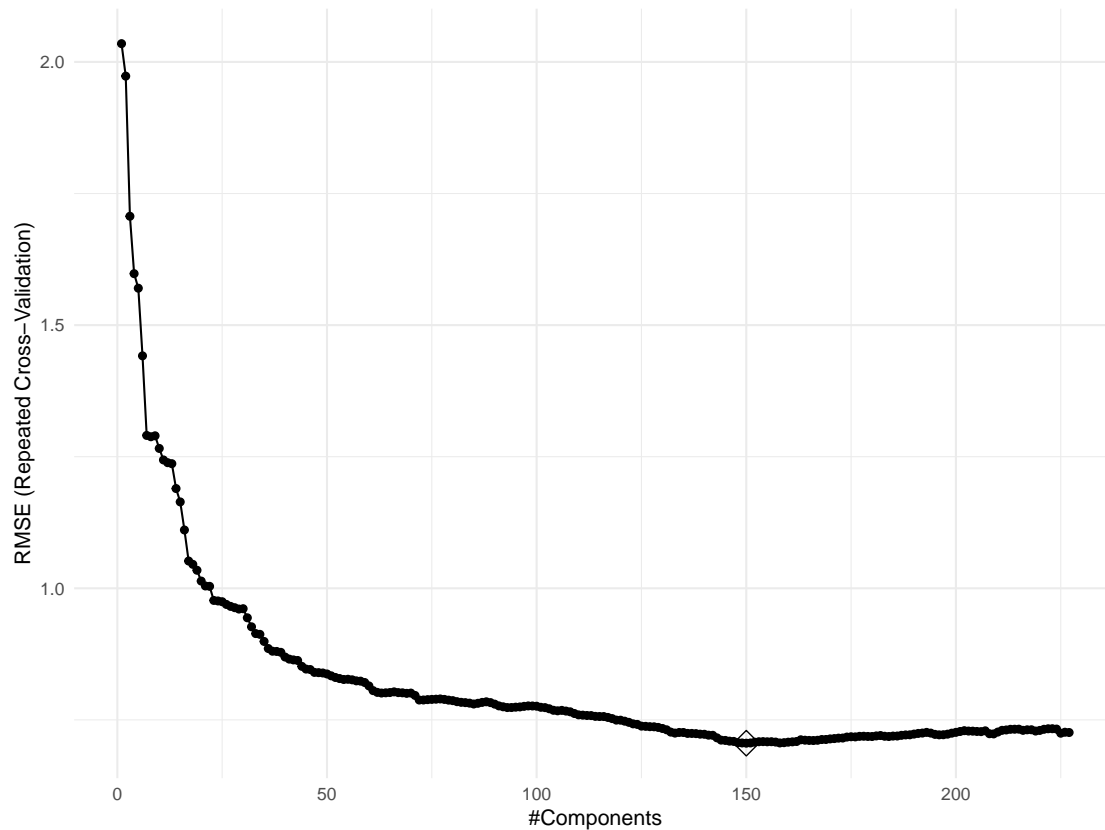
Part d, Principal Component Regression



```
## 151 comps
```

```
##      151
```

```
## [1] 0.5483361
```



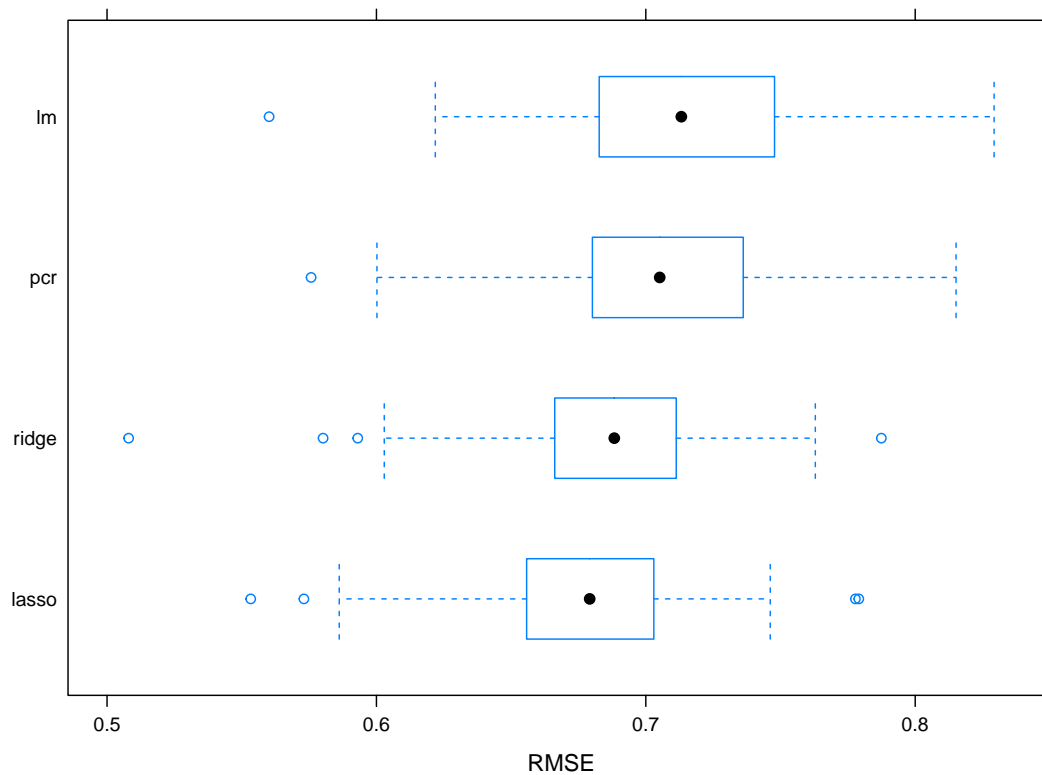
The PCR model chooses 150 components with an MSE of 0.548.

Now do it in caret

Part e, Results

```
##
## Call:
## summary.resamples(object = resamp)
##
## Models: lasso, ridge, pcr, lm
## Number of resamples: 50
##
## MAE
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max. NA's
## lasso 0.4455531 0.4981123 0.5204810 0.5159017 0.5377883 0.5820059    0
## ridge 0.3984352 0.4986790 0.5260025 0.5213705 0.5409572 0.5946389    0
## pcr   0.4541985 0.5175942 0.5402287 0.5413205 0.5609868 0.6359704    0
## lm    0.4313246 0.5131961 0.5286140 0.5307464 0.5511538 0.6293257    0
##
## RMSE
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max. NA's
## lasso 0.5533301 0.6568415 0.6791697 0.6756089 0.7026340 0.7790598    0
## ridge 0.5080045 0.6670191 0.6883045 0.6842593 0.7102808 0.7874402    0
## pcr   0.5757491 0.6804902 0.7051608 0.7057170 0.7358013 0.8151700    0
## lm    0.5601616 0.6837552 0.7131899 0.7132717 0.7474847 0.8292929    0
```

```
##
## Rsquared
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max. NA's
## lasso 0.8448545 0.8798899 0.8874942 0.8917332 0.9054675 0.9316582 0
## ridge 0.8362614 0.8725424 0.8895076 0.8887679 0.9029347 0.9312101 0
## pcr   0.8258269 0.8692911 0.8838365 0.8823444 0.8950950 0.9307879 0
## lm    0.8110752 0.8700942 0.8842094 0.8802278 0.8993033 0.9145587 0
```



While the MSE for a linear model is 0.556, this test error can be reduced using further regression methods. Of the ridge, lasso, and principal component regression methods, the lasso regression returns the lowest MSE, suggesting a model with 141 components is the best model of the three.

Part f, Discussion

As mentioned above, the model with the lowest MSE is considered the best predictive model, being the lasso model.