Variable Selection

1. Implement
2. Least Square Minimization

Please see the implementation: *least\_squares.m.* (Closed form solution is used)

With the full data (CASE1201.ASC.txt), the learned parameters are

W\_lsq =

-94.6591 (intercept term)

-0.4801 (takers)

-0.0082 (income)

22.6101 (years)

-0.4642 (public)

2.2120 (expend)

8.4762 (rank)

The Residual Sum of Squares (RSS) = 29842

1. Stepwise Selection

Please see the implementation *stepwise\_selection.m*

The selected covariates are all except *income*

i.e. rank, years, expend, public & takers (and by default the intercept term)

The Residual Sum of Squares (RSS) with the selected covariates is 29844

1. Lasso

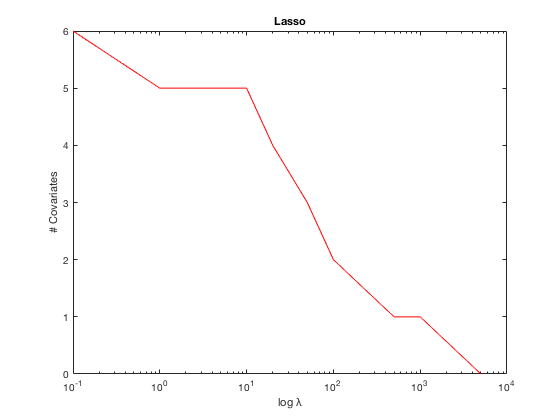
Implemented in *lasso.m*

Different values of regularization parameter lambda (λ) choose different number of covariates.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| λ | 0.1 | 1 | 5 | 10 | 20 | 50 | 100 | 500 | 1000 | 5000 |
| # of  Covariates | 6 | 5 | 5 | 5 | 4 | 3 | 2 | 1 | 1 | 0 |

(Note that when λ is very large (like 5000), all covariates except the intercept term becomes 0. The intercept term will be then equal to the average value of the predictions i.e. average of SAT scores (947.94))

(The x axis is plotted in log scale)



2. Compare these three methods in terms of cross validation errors. Choose K = 4 in cross validation.

(The test measure is RSS – Residual Sum of Squares)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| K=4 | Subset 1 | Subset 2 | Subset 3 | Subset 4 | Average |
| Least Squares | 9651.8 | 33672 | 3661.3 | 17609 | 16149 |
| Stepwise Selection | 9651.8 | 34413 | 3667.1 | 17609 | 16335 |
| Lasso | 9460 | 33277 | 3624.2 | 16822 | **15796** |

Lasso performs best out of the three methods!