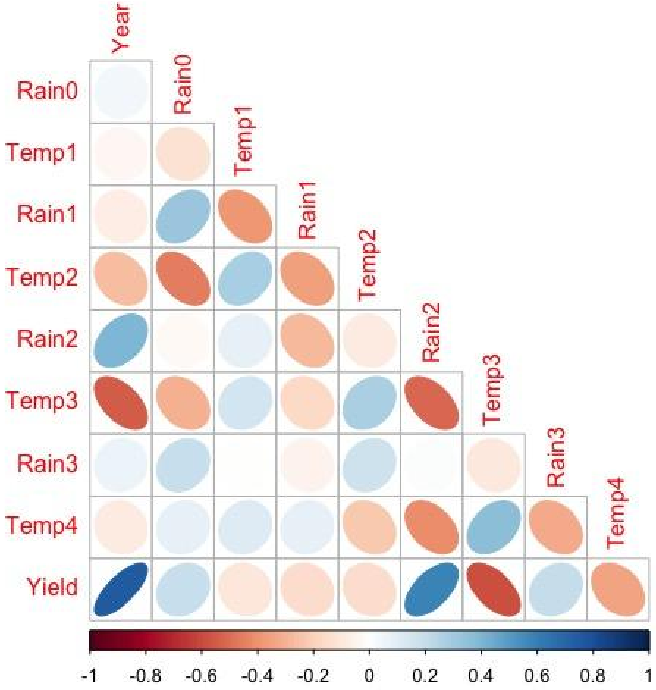
Mini project 3, Sparse Modeling

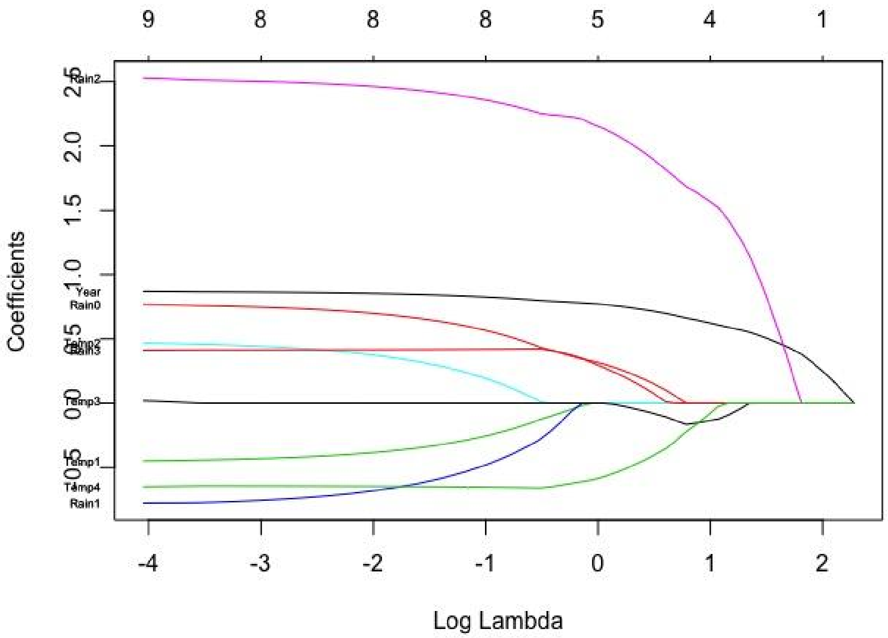
Magnus Lindström
Alexander Reinthal
Hampus Torén

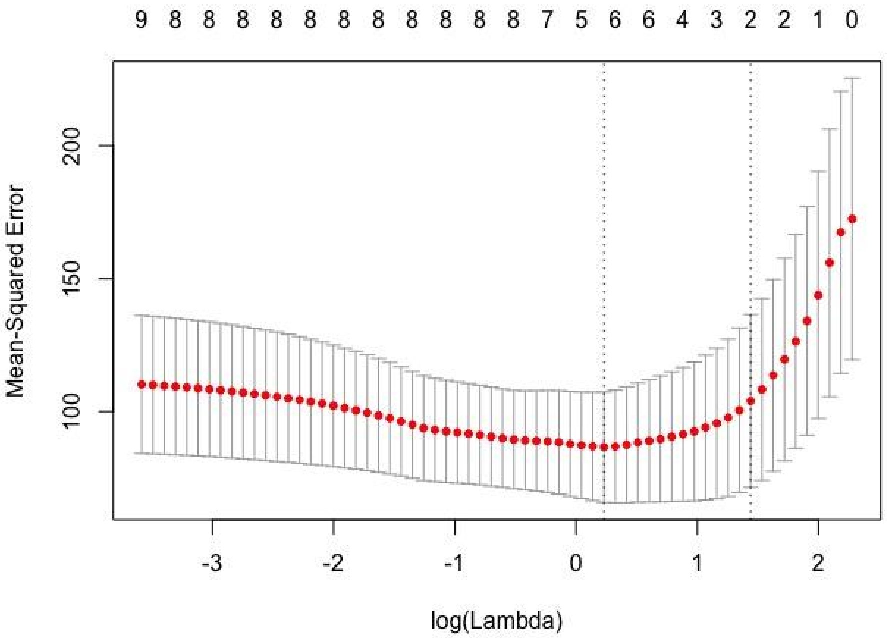
The Iowa dataset from lasso2 package

- n = 33, p = 10
- Info about rain, temperatures and total wheat harvest yield in Iowa during 33 years
- Response variable: Yield
- Some correlated predictors, will return to this



Lasso Performance

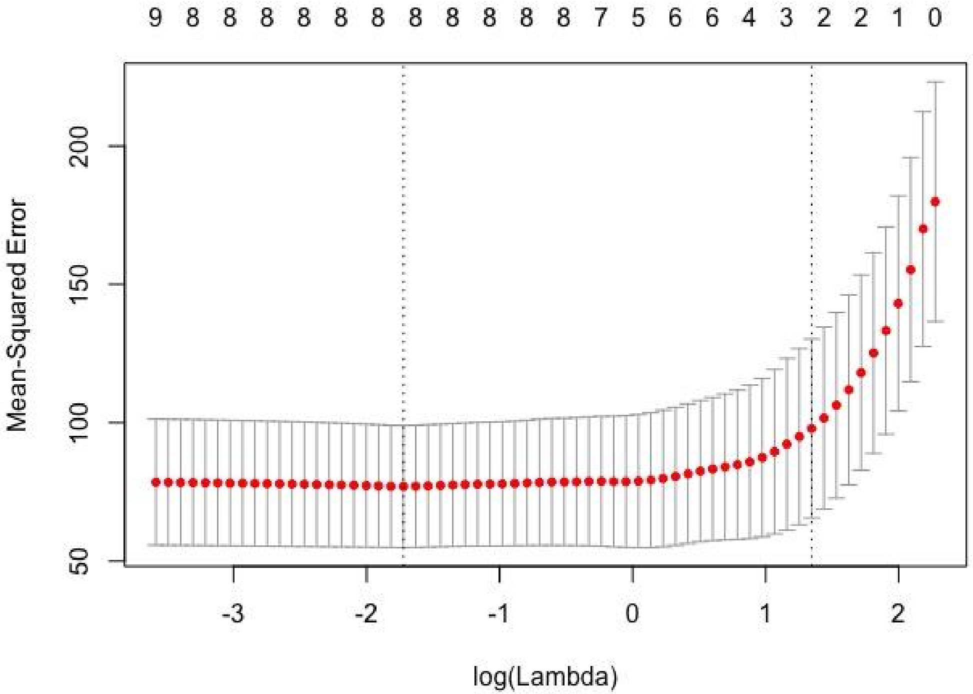




Min MSE: 86.6
MSE at one std: 104.0

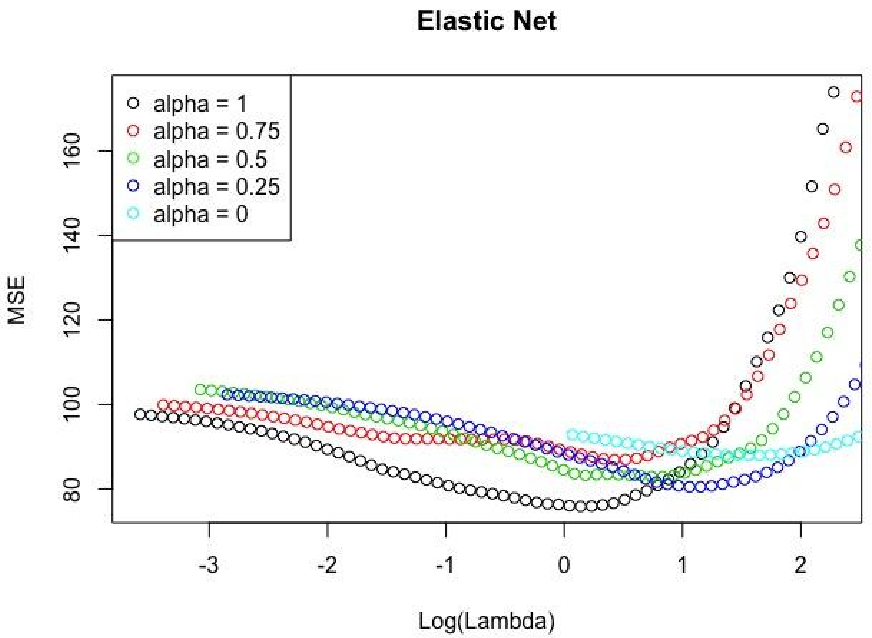
Adaptive Lasso Performance

- Used ridge regression weights to rescale
- Smallest MSE: 76.9
  one std away: 97.9



Elastic Net

- Some predictors are correlated, elastic net could be a good choice
- alpha = 1, pure LASSO, was still best

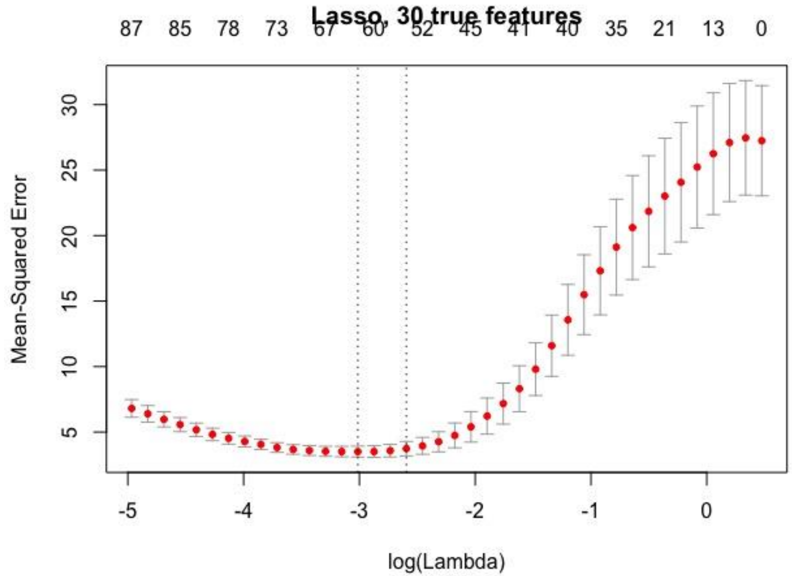


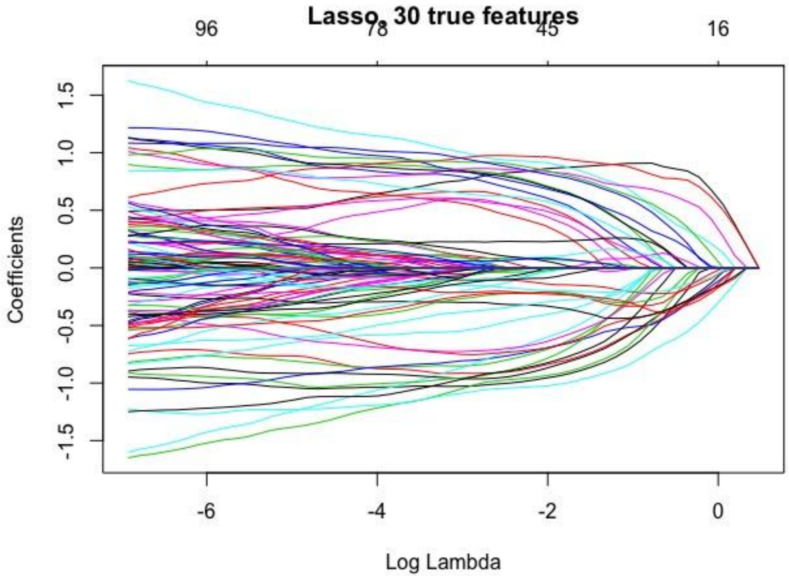
Non-Sparsity

- We wanted to see how the methods perform when the true model isn't sparse at all.
- Used simulated data with $n = 100$, $p = 100$, nrTrueVariables = 30,60,90.
- No correlation between predictor variables

Lasso performance, nTrueVariables = 30

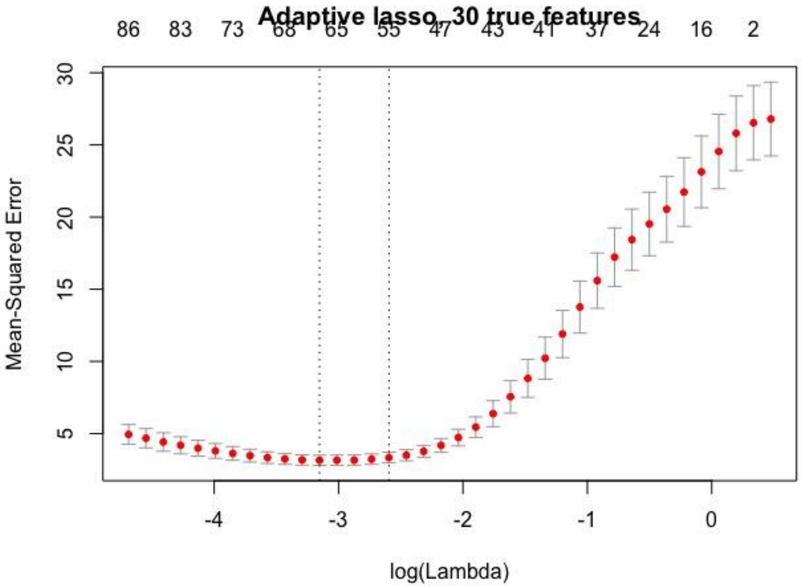
- LASSO improves performance, selects mostly the right predictors
- 2 true variables were set to 0 at lambda.1se

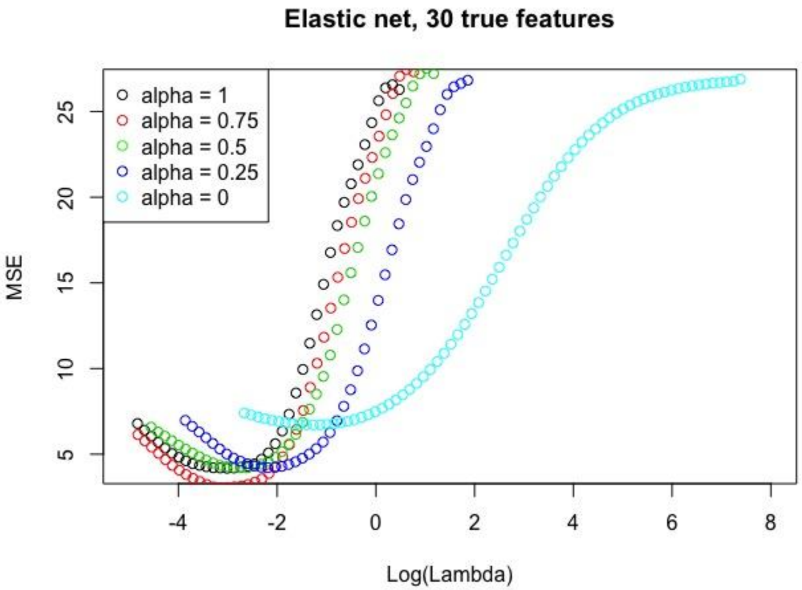




Adaptive Lasso, Elastic Net

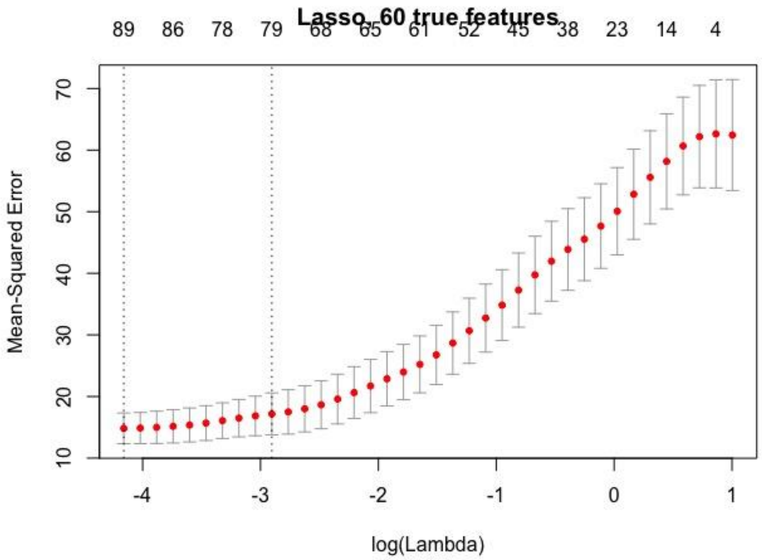
- Alpha = 0.75 was best
- 2 true variables were set to 0 at lambda.1se

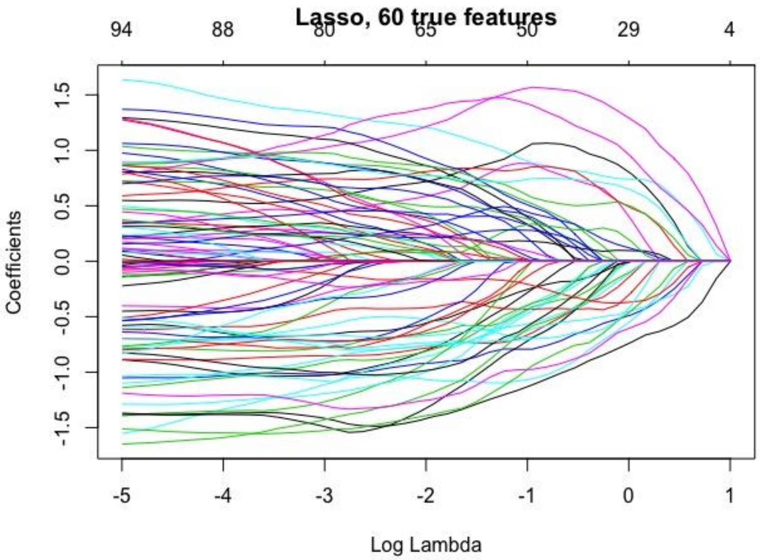




Lasso performance, nTrueVariables = 60

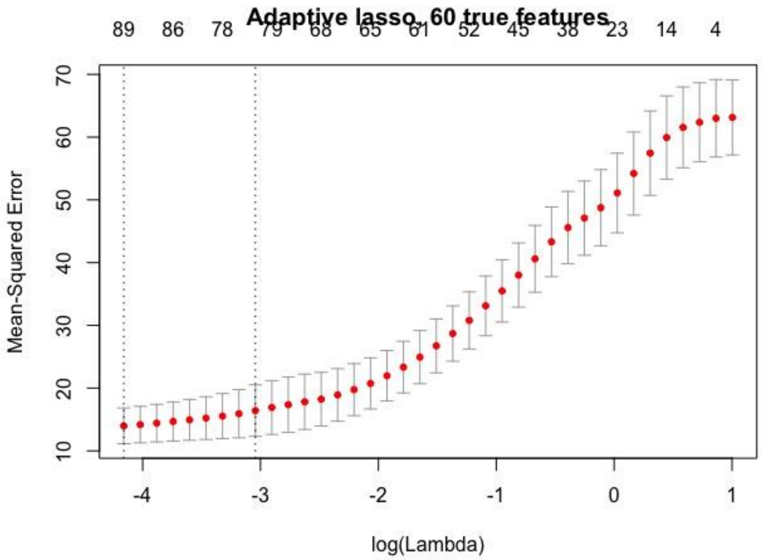
- LASSO now only makes things worse.
- No longer separates the right variables
- 3 true variables were set to 0 at lambda.1se

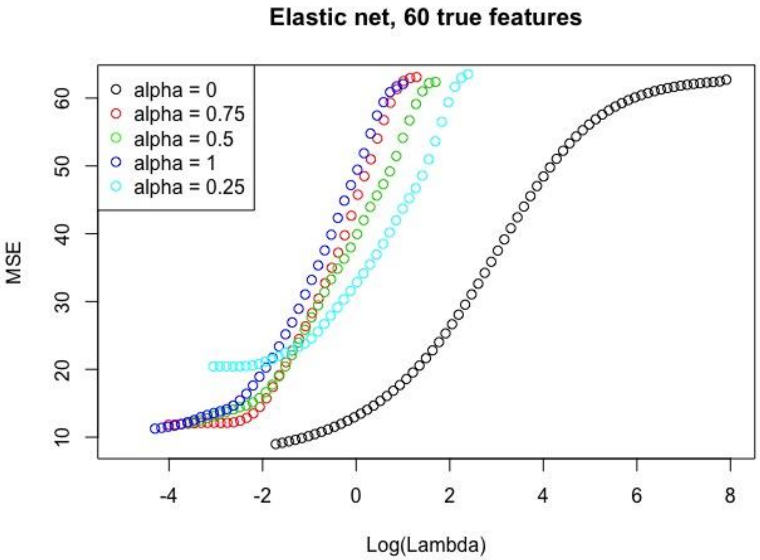




Adaptive Lasso, Elastic Net

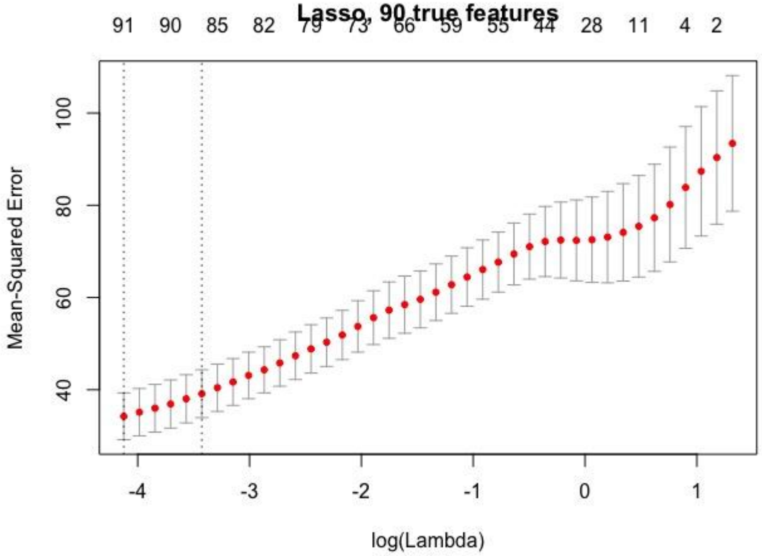
- Elastic net starts showing an interesting splitting of data

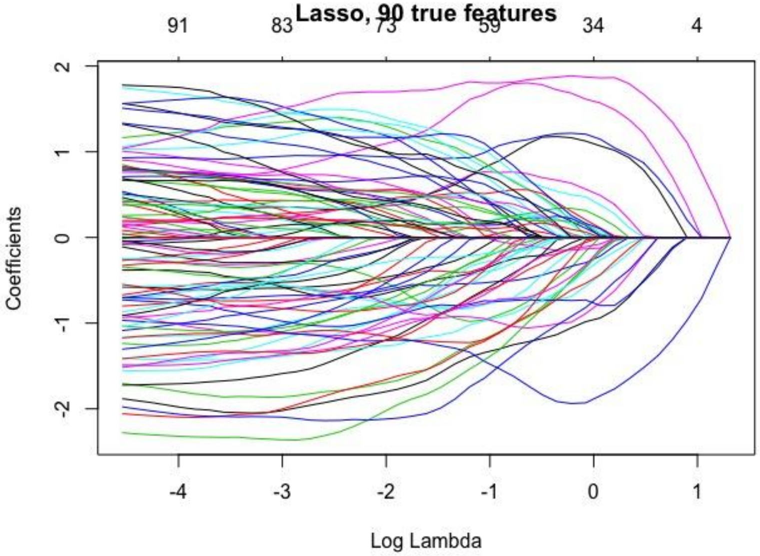




Lasso performance, nTrueVariables = 90

- LASSO is still pretty bad
- 10 true variables were set to 0 at lambda.1se

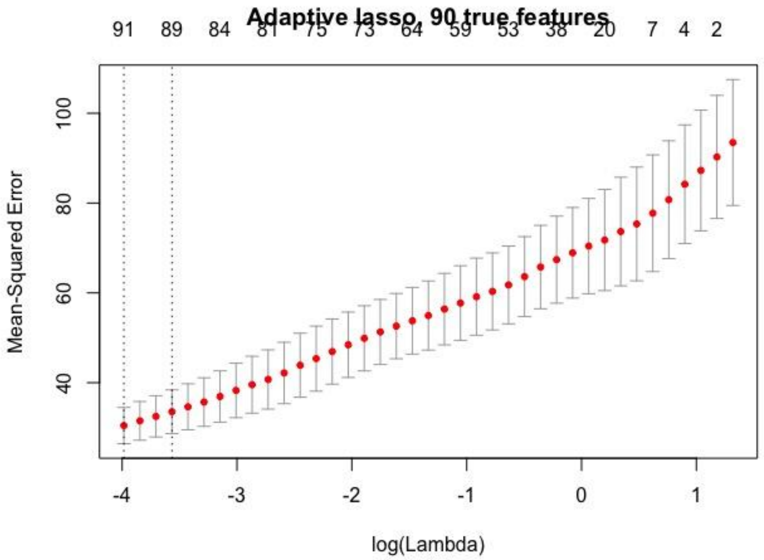


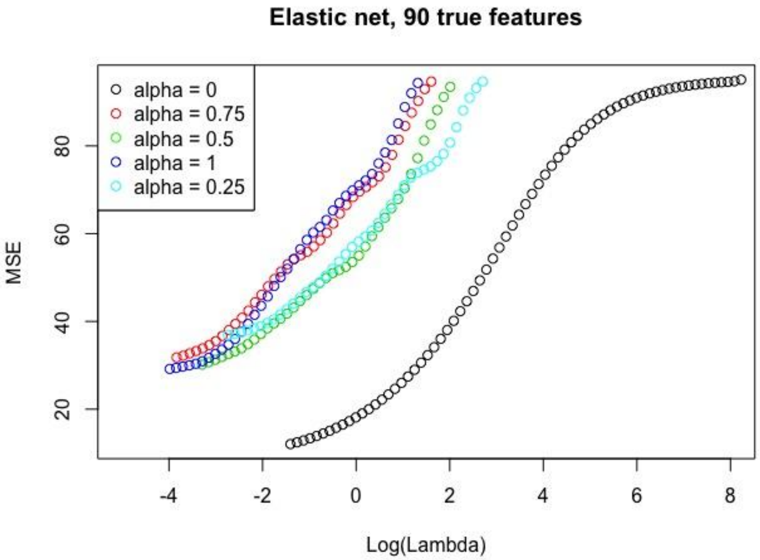


Adaptive Lasso, Elastic Net

- Elastic net showing strange results. lambda → 0 should give the same limit for all alpha

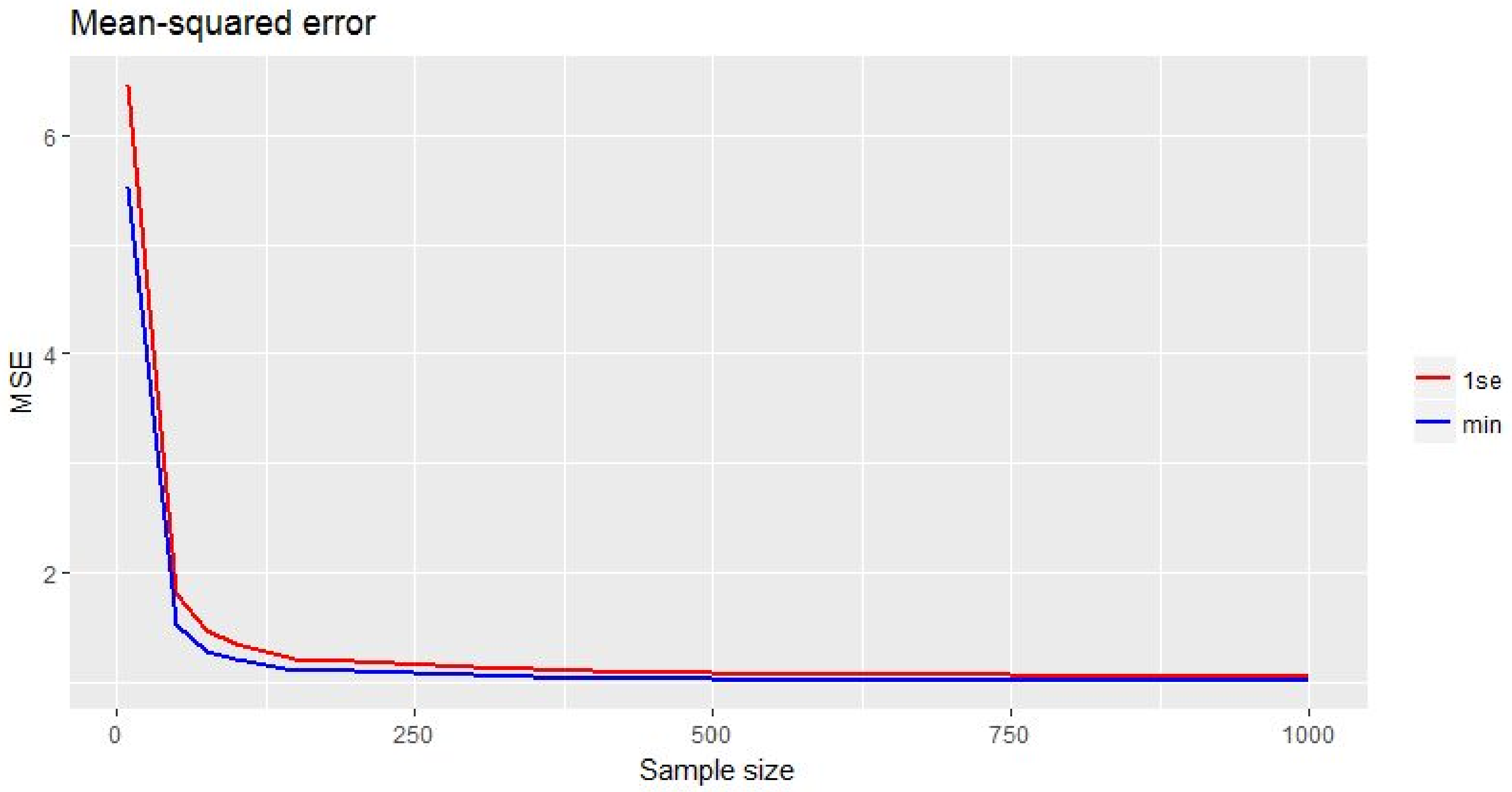
$$\frac{1}{2}||y - X\beta||^2 + (1-\alpha)\lambda||\beta||_2^2 + \alpha\lambda||\beta||_1$$

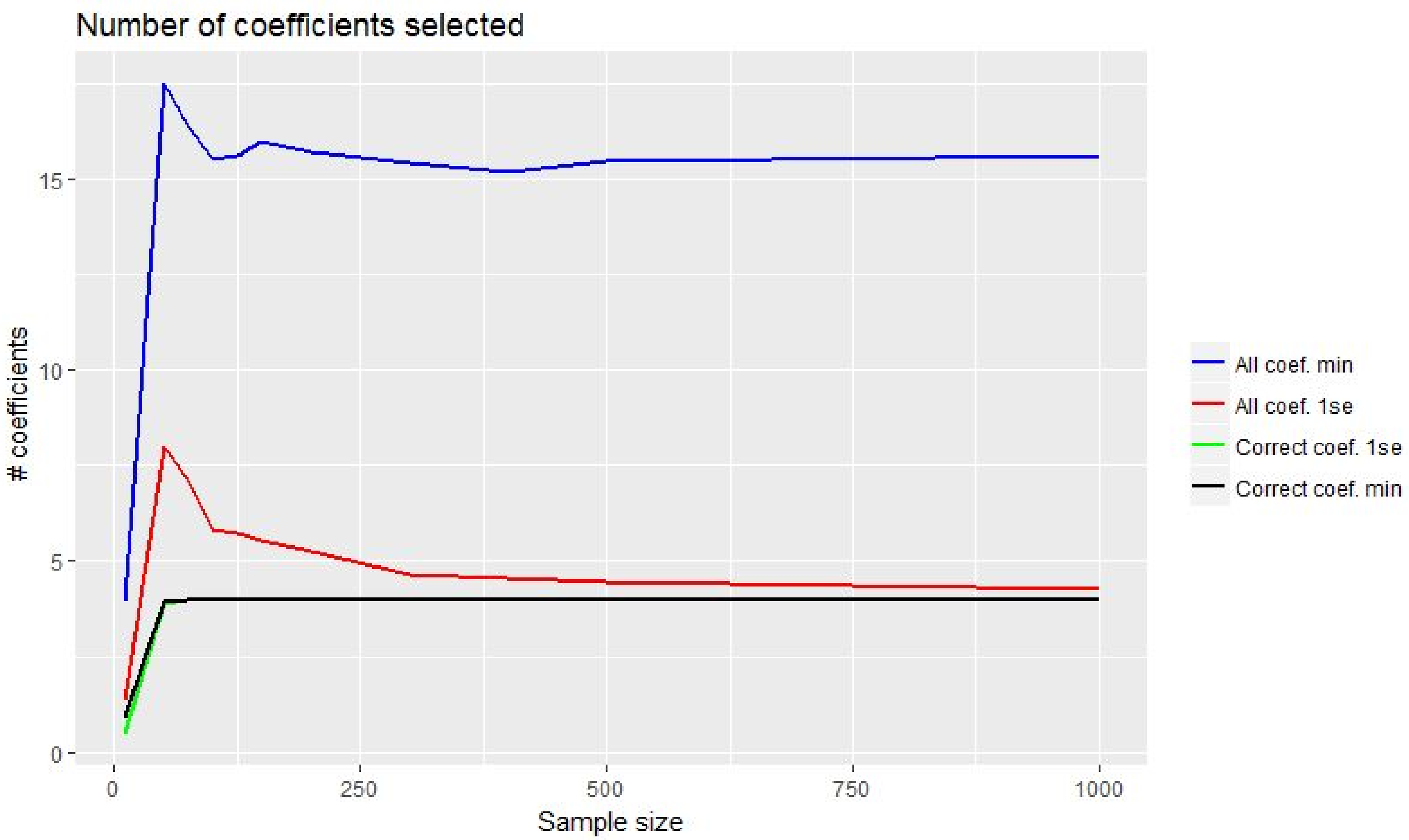




LASSO - Effect of sample size

- Two pairs of correlated predictors with correlation 0.9 and 100 unrelated predictors.
- The response variable consisted of three correlated predictors, one uncorrelated predictor and some noise. (same as Rebeckas code)
- For each sample size that was tested an average over 200 runs was taken.


Mean-squared error
MSE
6
4
2
0
250
500
750
1000
Sample size
1se
min


Number of coefficients selected
coefficients
Sample size
0
5
10
15
0
250
500
750
1000
All coef. min
All coef. 1se
Correct coef. 1se
Correct coef. min

Stability Of Model

Given bootstrap samples of smaller and smaller size does LASSO select the correct variables?

Stability Of LASSO

Synthetic data was generated using Rebecca's Code

$$\mathbf{x}_{1,2}, \mathbf{x}_{3,4} \sim N(\mu = \mathbf{0}, \Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix})$$

$$\mathbf{x}_j \sim N_p(0, I_p) \qquad j = 5...100$$

$$\mathbf{y} \sim \sum_i^q \mathbf{x}_i \beta_i + \epsilon \qquad \beta_i \in [-1, 1]$$

Stability Of LASSO

Performance measure:

"Oracle squared error"

$$H = \sum_i (\beta_i - \hat{\beta}_i)^2$$

and

Number of variables selected

Stability Of LASSO

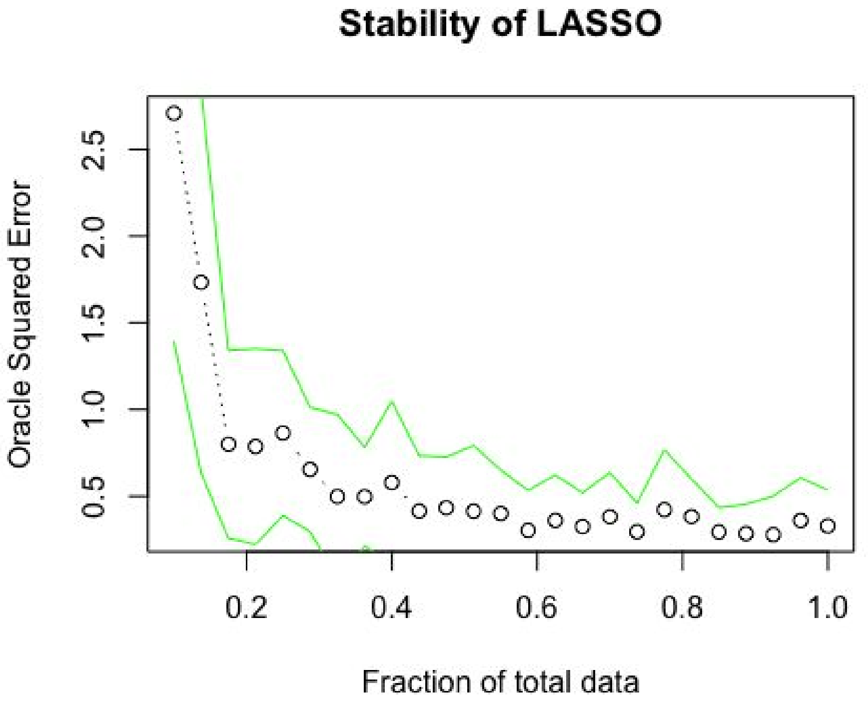
Simulation setup:

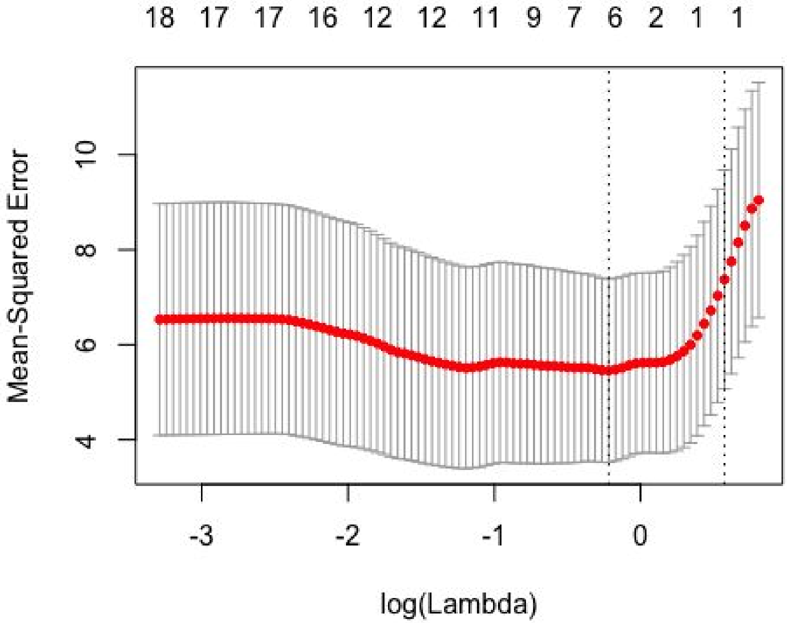
- The LASSO method was fed smaller and smaller bootstrap samples **with** replacement

footnote:

- For some reason performance was really bad, a lot of variables were selected.
- anyone think of why?

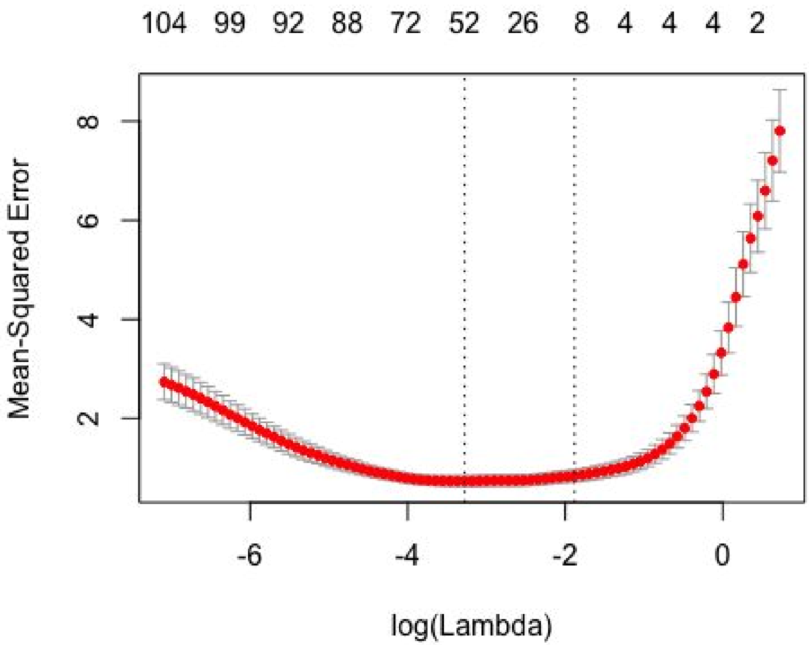
Stability Of LASSO - Bootstrap sample, 30 averages





Example - n = 20, p = 100 bootstrap sample

Stability Of LASSO - Bootstrap sample, n = 200



Stability of LASSO - Certainty

Problem:

- Measuring oracle squared error might not be a good measure of performance

Solution:

- Measure stability in terms of p-values on beta hats.

Stability of LASSO - Certainty

Planned simulation setup:

- Run 50 simulations for 30 sizes of bootstraps
- n = 200, p = 100, nrTrueVariables = 4
- Calculate p-values for the true variables

Failure:

- Algorithm (lasso.proj()) hits a wall when not given enough data.
- When algorithm does not converge fast (< 60 sec) the p-values are usually very high.
- Did not have large enough sample size to present plots :(

Conclusions

- **Iowa dataset:** We saw that sparse modeling improved the predictive performance since the mean-squared error decreased. Adaptive LASSO was the best method.
- **Non-sparsity:** Up to approximately 30 true variables among 100 , LASSO improved the prediction. However as the dataset grow less sparse LASSO did not improve the prediction at all.

Conclusions

- **Sample size:** The MSE converged when $n \approx p$. The correct predictors were selected already at $n \approx p/2$. The 1se-limit converge towards the correct amount of predictors for large n. The min-limit converged to approximately 16 predictors for large n.
- **Stability of LASSO:** Smaller bootstraps samples give better variable selection.

The End

Thank you for listening :)