# Non-negative Garrote

Ryan Hicks

Department of Statistics, Colorado State University

December 8, 2015

# Least Squares

- $\hat{\beta}_{LS} = \arg\min_{\beta \in \mathbb{R}^p} ||\mathbb{Y} - \mathbb{X}\beta||_2^2$
- Model selection through forward, backward, and all subsets regression, but not convex

# Ridge Regression

- $\hat{\beta}_{ridge,t} = \arg\min_{||\beta||_2^2 \leq t} ||\mathbb{Y} - \mathbb{X}\beta||_2^2$ for any $t \geq 0$.

- Convex, but does not help with model selection

- $\hat{\beta}_{ridge,\lambda} = \arg\min_{\beta} ||\mathbb{Y} - \mathbb{X}\beta||_2^2 + \lambda||\beta||_2^2$

- Stable solutions

# Lasso

- Lasso finds a middle ground in the penalization term that allows for model selection and optimization
- $\hat{\beta}_{lasso}(\lambda) = \arg\min_{\beta} ||\mathbb{Y} - \mathbb{X}\beta||_2^2 + \lambda||\beta||_1$

# Lasso

- Lasso is powerful but has problems
- For $p > n$, Lasso can select at most n features
- Does poorly with highly correlated features
- Not invariant to rescaling

# Non-negative Garrote

- $\hat{\beta}_{NNG} = \arg\min_{\beta} ||\mathbb{Y} - \mathbb{X}\beta||_2^2 + 2\lambda \sum_{j=1}^{p} u_j$, subject to $u_j \geq 0$, $j = 1, ..., p$
- $\hat{\beta}_{NNG} = (u_1\hat{\beta}_{LS_1}, ..., u_p\hat{\beta}_{LS_p})^T$

# Non-negative Garrote

- Stable solutions
- Shrinks and eliminates predictors
- Scale invariant
- Better predictive accuracy than subsets, comparable to ridge

# Non-negative Garrote

- NNG has its own problems
- Requires OLS solution*, sensitive to standard OLS assumptions
- Needs $n > p$
- Not popular, no R package

# References


- Breiman, Leo. "Better subset regression using the nonnegative garrote." Technometrics 37.4 (1995): 373-384.
- Xiong, Shifeng. "Some notes on the nonnegative garrote." Technometrics 52.3 (2010).
- Yuan, Ming, and Yi Lin. "On the non?negative garrotte estimator." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69.2 (2007): 143-161.

# The End

Thank You!

Questions?

# NNG Theory

- Piece-wise linear solution path
- Path-consistent
- Natural selection of tuning parameter: $\hat{\sigma}^2$, $\frac{\hat{\sigma}^2 ln(n)}{2}$
- MSE converges to 0.