

# A simple explanation of the Lasso and Least Angle Regression

Give a set of input measurements  $x_1, x_2, \dots, x_p$  and an outcome measurement  $y$ , the lasso fits a linear model

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

The criterion it uses is:

$$\text{Minimize } \sum (y - \hat{y})^2 \text{ subject to } \sum [\text{absolute value}(b_j)] \leq s$$

The first sum is taken over observations (cases) in the dataset. The bound " $s$ " is a tuning parameter. When " $s$ " is large enough, the constraint has no effect and the solution is just the usual multiple linear least squares regression of  $y$  on  $x_1, x_2, \dots, x_p$ .

However when for smaller values of  $s$  ( $s \geq 0$ ) the solutions are shrunken versions of the least squares estimates. Often, some of the coefficients  $b_j$  are zero. Choosing " $s$ " is like choosing the number of predictors to use in a regression model, and cross-validation is a good tool for estimating the best value for " $s$ ".

## Computation of the Lasso solutions

**The computation of the lasso solutions is a quadratic programming problem, and can be tackled by standard numerical analysis algorithms. But the least angle regression procedure is a better approach. This algorithm exploits the special structure of the lasso problem, and provides an efficient way to compute the solutions simultaneously for all values of " $s$ ".**

**Least angle regression is like a more "democratic" version of forward stepwise regression. Recall how forward stepwise regression works:**

### **Forward stepwise regression algorithm:**

- Start with all coefficients  $b_j$  equal to zero.
- Find the predictor  $x_j$  most correlated with  $y$ , and add it into the model. Take residuals  $r = y - \hat{y}$ .
- Continue, at each stage adding to the model the predictor most correlated with  $r$ .
- Until: all predictors are in the model

The least angle regression procedure follows the same general scheme, but doesn't add a predictor fully into the model. The coefficient of that predictor is increased only until that predictor is no longer the one most correlated with the residual  $r$ . Then some other competing predictor is invited to "join the club".

### **Least angle regression algorithm:**

- Start with all coefficients  $b_j$  equal to zero.
- Find the predictor  $x_j$  most correlated with  $y$
- Increase the coefficient  $b_j$  in the direction of the sign of its correlation with  $y$ . Take residuals  $r = y - \hat{y}$  along the way. Stop when some other predictor  $x_k$  has as much correlation with  $r$  as  $x_j$  has.
- Increase  $(b_j, b_k)$  in their joint least squares direction, until some other predictor  $x_m$  has as much correlation with the residual  $r$ .
- Continue until: all predictors are in the model

Surprisingly it can be shown that, with one modification, this procedure gives the entire path of lasso solutions, as  $s$  is varied from 0 to infinity. The modification needed is: if a non-zero coefficient hits zero, remove it from the active set of predictors and recompute the joint direction.