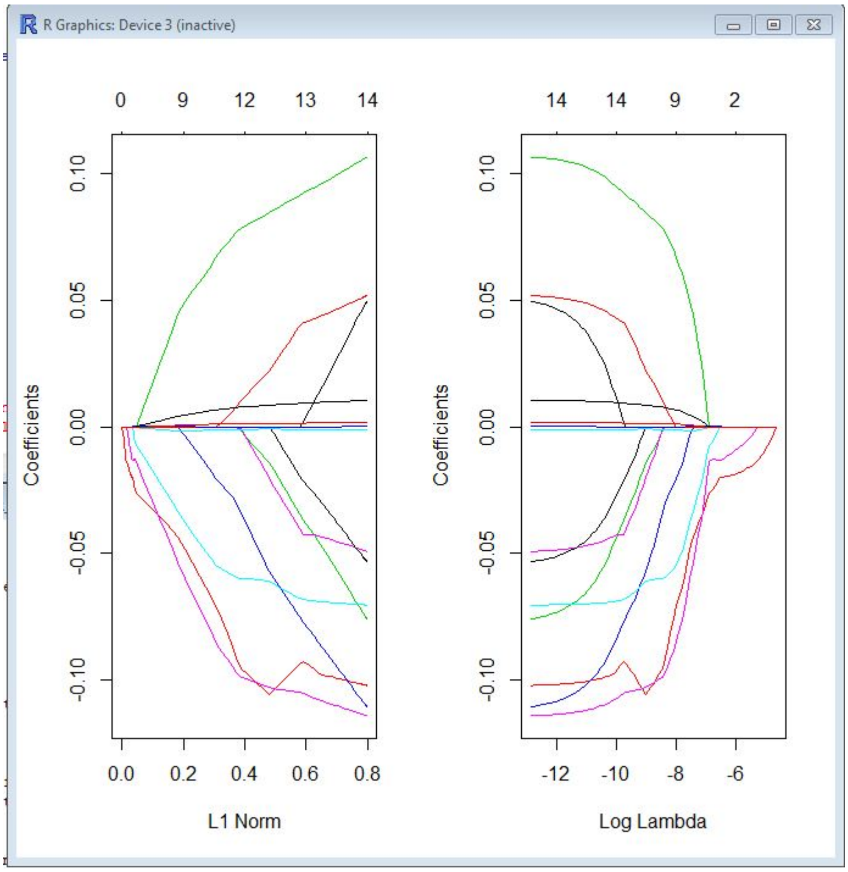
Interpretting LASSO variable trace plots

I am new to the `glmnet` package, and I am still unsure of how to interpret the results. Could anyone please help me read the following trace plot?

The graph was obtaining by running the following:

```
library(glmnet)
return <- matrix(ret.ff.zoo[which(index(ret.ff.zoo)==beta.df$date[2]), ])
data   <- matrix(unlist(beta.df[which(beta.df$date==beta.df$date[2]), ][ ,-1]),
                 ncol=num.factors)
model  <- cv.glmnet(data, return, standardize=TRUE)

op <- par(mfrow=c(1, 2))
plot(model$glmnet.fit, "norm",   label=TRUE)
plot(model$glmnet.fit, "lambda", label=TRUE)
par(op)
```



r data-visualization interpretation lasso glmnet

edited Dec 6 '13 at 4:56
gung
64.2k 18 140 264

asked Aug 27 '13 at 13:20
Mayou
197 3 17

1 Answer

In both plots, each colored line represents the value taken by a different coefficient in your model. **Lambda** is the weight given to the regularization term (the L1 norm), so as lambda approaches zero, the loss function of your model approaches the OLS loss function. Here's one way you could specify the LASSO loss function to make this concrete:

$$\beta_{lasso} = \text{argmin}\left[RSS(\beta) + \lambda * \text{L1-Norm}(\beta)\right]$$

Therefore, when lambda is very small, the LASSO solution should be very close to the OLS solution, and all of your coefficients are in the model. As lambda grows, the regularization term has greater effect and you will see fewer variables in your model (because more and more coefficients will be zero valued).

As I mentioned above, the **L1 norm** is the regularization term for LASSO. Perhaps a better way to look at it is that the x-axis is the *maximum permissible value the L1 norm can take*. So when you have a small L1 norm, you have a lot of regularization. Therefore, an L1 norm of zero gives an empty model, and as you increase the L1 norm, variables will "enter" the model as their coefficients take non-zero values.

The plot on the left and the plot on the right are basically showing you the same thing, just on different scales.

edited Aug 27 '13 at 15:00

answered Aug 27 '13 at 13:54

David Marx
3,749 7 25

1 Very neat answer, thanks! Is is possible to deduce the "best predictors" from the graphs above, i.e. a final model? – Mayou Aug 27 '13 at 13:56

3 No, you'll need to cross-validation or some other validation procedure for that; it'll tell you which value of the L1 norm (or equivalently, which log(lambda)) yields the model with best predictive ability. – JAW Aug 27 '13 at 14:55

5 If you're trying to determine your strongest predictors, you could interpret the plot as evidence that variables that enter the model early are the most predictive and variables that enter the model later are less important. If you want the "best model," generally this is found via cross validation. A common method for attaining this using the glmnet package was suggested to you here: stats.stackexchange.com/a/68350/8451 . I *strongly* recommend you read the short Lasso chapter in ESLII (3.4.2 and 3.4.3), which is free to download: www-stat.stanford.edu/~tibs/ElemStatLearn – David Marx Aug 27 '13 at 14:58