

Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization. It's 100% free, no registration required.

Sign up

Here's how it works:

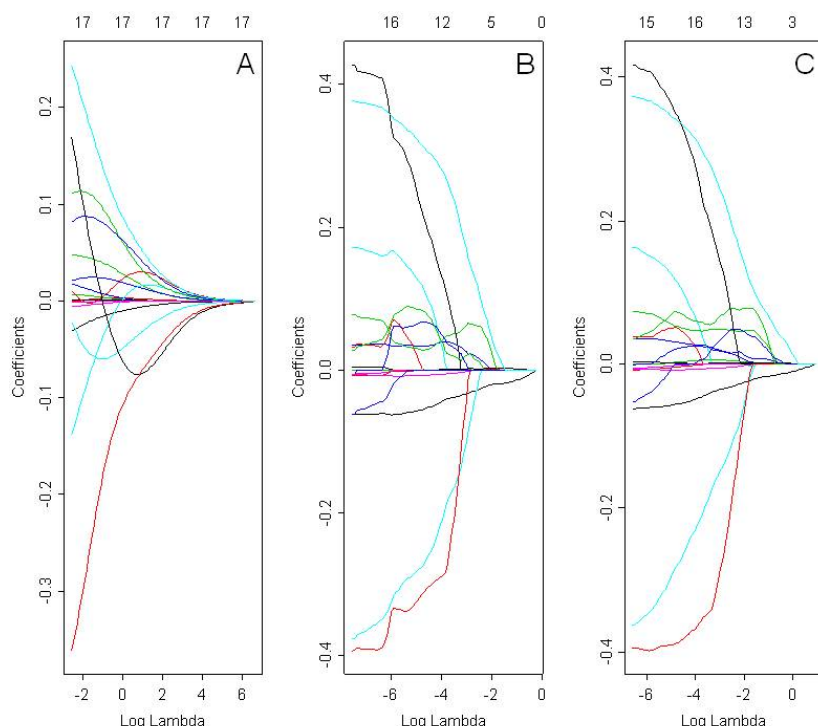
Anybody can ask a question

Anybody can answer

The best answers are voted up and rise to the top

Coefficients paths – comparison of ridge, lasso and elastic net regression

I would like to compare models selected with ridge, lasso and elastic net. Fig. below shows coefficients paths using all 3 methods: ridge (Fig A, $\alpha=0$), lasso (Fig B; $\alpha=1$) and elastic net (Fig C; $\alpha=0.5$). The optimal solution depends on the selected value of λ , which is chosen based on cross validation.



When looking at these plots, I would expect the elastic net (Fig C) to exhibit a grouping effect. However it is not clear in the presented case. The coefficients path for lasso and elastic net are very similar. What could be the reason for this? Is it just a coding mistake? I used the following code in R:

```
library(glmnet)
X<- as.matrix(mydata[,2:22])
Y<- mydata[,23]
par(mfrow=c(1,3))
ans1<-cv.glmnet(X, Y, alpha=0) # ridge
plot(ans1$glmnet.fit, "lambda", label=FALSE)
text(6, 0.4, "A", cex=1.8, font=1)
ans2<-cv.glmnet(X, Y, alpha=1) # lasso
plot(ans2$glmnet.fit, "lambda", label=FALSE)
text(-0.8, 0.48, "B", cex=1.8, font=1)
ans3<-cv.glmnet(X, Y, alpha=0.5) # elastic net
plot(ans3$glmnet.fit, "lambda", label=FALSE)
text(0, 0.62, "C", cex=1.8, font=1)
```

The code used to plot elastic net coefficients paths is exactly the same as for ridge and lasso. The only difference is in the value of α . α parameter for elastic net regression was selected based on the lowest MSE (mean squared error) for corresponding λ values.

Thank you for your help !

r multiple-regression modeling regularization

edited Jun 5 '13 at 12:39



Roland
2,133 8 22

asked Jun 5 '13 at 12:26



beginner
181 4 11

1 Answer

In the $p < n$ case (p number of coefficients, n number of samples, which by the number of coefficients you show in the plots I guess it is the case here), the only real "problem" with the Lasso model is that when multiple features are correlated it tends to select one of them somewhat randomly.

If the original features are not very correlated, I would say that it is reasonable that Lasso performs similar to Elastic Net in terms of coefficients path. Looking at the documentation for [glmnet](#) package, I also can't see any error in your code.

edited Jun 7 '13 at 11:30



COOLSerdash

8,963 4 27 52

answered Jun 7 '13 at 11:11



skd

194 5

Thank you very much for your comment. I started to think about regularization approach to select the best model, because I have observed a severe multicollinearity between my variables ($VIF \gg 10$). Many of them were correlated at the level of > 0.8 . Thus I was expecting elastic net to perform differently to lasso and show a grouping effect (in case of correlated variables). – [beginner](#) Jun 7 '13 at 15:07

You may just have to use a stronger ridge penalty? – [dcl](#) Jun 12 '13 at 0:12

did you check df adjusted multicollinearity? factor variables have normally multicollinearity because they're mutually exclusive. I don't know whether this kind of multicollinearity gives problems with lasso. – [Bakaburg](#) Feb 15 '15 at 14:09
