R Lab : lars()

RAHUL DEBNATH
Under mentorship of Prof. Arnab Chakraborty

26 OCTOBER 2022

Indian Statistical Institute, Kolkata.

"THE FUTURE BELONGS TO THOSE WHO BELIEVE IN THE BEAUTY OF THEIR DREAMS."

Eleanor Roosevelt

Hello there, readers ! Yes, you. I need your eyes and attention here in 3 ...... 2 .... 1 ......

Today, what we are interested to discuss is about lars() using R.

The first thing that we shall try to learn is ridge regression for which we shall use a default installed package, we load the package MASS using the command :

$$\boxed{library(MASS)}$$

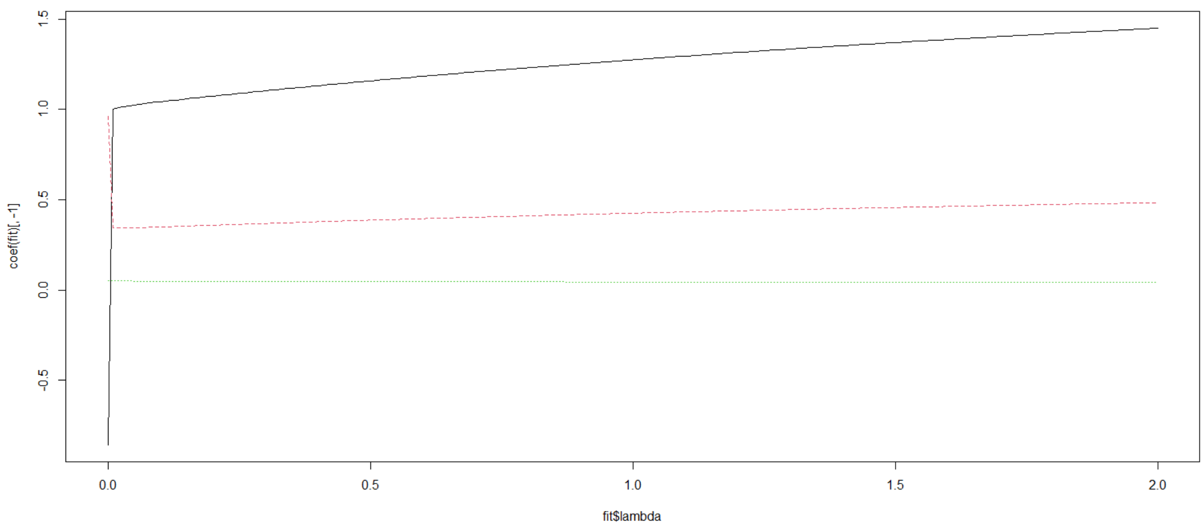
Now, we shall use the function lm.ridge.Here we specify the value of lambda which is the lambda of the adhoc formulation or we specify a list of lambda values if we donot know what value of lambda suits the most.So we use the following R command :

$$\boxed{fit = lm.ridge(y\ x1 + x2 + x3, lambda = seq(0, 2, len = 200))}$$

Here, we used a sequence from 0 to 2 of length 200 , so we get a list of values of lambda starting from 0 with an increament of some delta amounts each time. Now if look at the output using the commant fit , we see that it is a very large output data which is not a desiarable output.So, we find the coefficient using the coef function.We plot the different fitted things against lambda excluding the intercept term using the following R command:

$$\boxed{matplot(fit\$lambda, coef(fit)[, -1], ty =' l')}$$

We get the following output :



Here, the black line is the coefficient of x1, the red line is the coefficient of x2 and the green line is the coefficient of x3. From the plotted figure we observe the coefficient of x3 remains more or less flat throughout but coefficient of x1 and x2 changes drastically.Initially black line is less than -0.5 and the red line is more than +0.5 but the second we increased lambda by a small delta amount the values comes down to certain values and then they become stable throughout the increament of value of lambda.
The coefficient which is not locked in a multiple linear relationship i.e the green line is always

flat , whcih means it is stable throughout. This indicates the presence of multicolliearity, where the first two coefficients are the first two columns of our design matrix which are involved in multicollinear relationship.
Such plots known as Ridge Trace, is also a good diagniostic tool. For lambda greater than 0 where things get stabilized, we use those values with more confidence.

Now we use the following R command to look into the coefficients :

$$\boxed{coef(fit)}$$

The output is

```
                                    x1        x2         x3
0.00000000   0.18546193 -0.8573656 0.9615618 0.04986391
0.01005025   0.09694658  0.9996151 0.3444963 0.04980413
0.02010050  -0.01086024  1.0109242 0.3429083 0.04973831
0.03015075  -0.11815329  1.0168772 0.3430956 0.04967278
0.04020101  -0.22489369  1.0214704 0.3437257 0.04960756
0.05025126  -0.33107684  1.0255088 0.3445302 0.04954264
0.06030151  -0.43670390  1.0292616 0.3454194 0.04947803
0.07035176  -0.54177763  1.0328444 0.3463548 0.04941373
0.08040201  -0.64630137  1.0363150 0.3473173 0.04934973
0.09045226  -0.75027871  1.0397057 0.3482960 0.04928602
0.10050251  -0.85371332  1.0430361 0.3492847 0.04922262
0.11055276  -0.95660894  1.0463183 0.3502792 0.04915951
0.12060302  -1.05896929  1.0495608 0.3512768 0.04909669
0.13065327  -1.16079809  1.0527693 0.3522758 0.04903417
0.14070352  -1.26209905  1.0559480 0.3532748 0.04897194
0.15075377  -1.36287585  1.0591000 0.3542728 0.04890999
0.16080402  -1.46313214  1.0622275 0.3552691 0.04884833
0.17085427  -1.56287154  1.0653325 0.3562632 0.04878695
0.18090452  -1.66209765  1.0684163 0.3572546 0.04872586
0.19095477  -1.76081404  1.0714801 0.3582432 0.04866504
0.20100503  -1.85902423  1.0745248 0.3592285 0.04860451
0.21105528  -1.95673174  1.0775512 0.3602105 0.04854424
0.22110553  -2.05394002  1.0805599 0.3611889 0.04848426
0.23115578  -2.15065254  1.0835516 0.3621637 0.04842454
0.24120603  -2.24687268  1.0865267 0.3631347 0.04836510
0.25125628  -2.34260384  1.0894855 0.3641019 0.04830592
0.26130653  -2.43784935  1.0924285 0.3650652 0.04824701
0.27135678  -2.53261255  1.0953561 0.3660246 0.04818836
0.28140704  -2.62689672  1.0982684 0.3669801 0.04812998
0.29145729  -2.72070512  1.1011657 0.3679315 0.04807186
0.30150754  -2.81404097  1.1040484 0.3688790 0.04801399
0.31155779  -2.90690749  1.1069166 0.3698224 0.04795639
0.32160804  -2.99930784  1.1097704 0.3707618 0.04789904
0.33165829  -3.09124517  1.1126102 0.3716971 0.04784195
0.34170854  -3.18272260  1.1154361 0.3726285 0.04778510
0.35175879  -3.27374320  1.1182482 0.3735558 0.04772851
0.36180905  -3.36431005  1.1210467 0.3744790 0.04767217
0.37185930  -3.45442618  1.1238317 0.3753983 0.04761607
0.38190955  -3.54409460  1.1266034 0.3763135 0.04756023
```

0.39195980	-3.63331827	1.1293620	0.3772248	0.04750462
0.40201005	-3.72210017	1.1321075	0.3781321	0.04744926
0.41206030	-3.81044322	1.1348400	0.3790354	0.04739414
0.42211055	-3.89835032	1.1375598	0.3799347	0.04733925
0.43216080	-3.98582434	1.1402669	0.3808302	0.04728461
0.44221106	-4.07286814	1.1429613	0.3817217	0.04723020
0.45226131	-4.15948455	1.1456433	0.3826093	0.04717603
0.46231156	-4.24567636	1.1483129	0.3834930	0.04712208
0.47236181	-4.33144636	1.1509703	0.3843729	0.04706837
0.48241206	-4.41679730	1.1536154	0.3852489	0.04701490
0.49246231	-4.50173190	1.1562485	0.3861211	0.04696164
0.50251256	-4.58625288	1.1588696	0.3869894	0.04690862
0.51256281	-4.67036292	1.1614788	0.3878540	0.04685582
0.52261307	-4.75406468	1.1640762	0.3887148	0.04680325
0.53266332	-4.83736079	1.1666619	0.3895719	0.04675090
0.54271357	-4.92025386	1.1692360	0.3904253	0.04669877
0.55276382	-5.00274650	1.1717985	0.3912749	0.04664686
0.56281407	-5.08484127	1.1743496	0.3921209	0.04659517
0.57286432	-5.16654071	1.1768893	0.3929632	0.04654369
0.58291457	-5.24784736	1.1794177	0.3938018	0.04649243
0.59296482	-5.32876372	1.1819349	0.3946369	0.04644139
0.60301508	-5.40929227	1.1844409	0.3954683	0.04639055
0.61306533	-5.48943547	1.1869359	0.3962961	0.04633993
0.62311558	-5.56919577	1.1894199	0.3971204	0.04628952
0.63316583	-5.64857559	1.1918930	0.3979412	0.04623932
0.64321608	-5.72757732	1.1943552	0.3987584	0.04618933
0.65326633	-5.80620336	1.1968067	0.3995721	0.04613954
0.66331658	-5.88445605	1.1992476	0.4003823	0.04608996
0.67336683	-5.96233775	1.2016777	0.4011891	0.04604058
0.68341709	-6.03985077	1.2040974	0.4019925	0.04599141
0.69346734	-6.11699742	1.2065066	0.4027924	0.04594243
0.70351759	-6.19377997	1.2089053	0.4035889	0.04589366
0.71356784	-6.27020070	1.2112938	0.4043820	0.04584508
0.72361809	-6.34626184	1.2136719	0.4051718	0.04579670
0.73366834	-6.42196564	1.2160399	0.4059583	0.04574852
0.74371859	-6.49731429	1.2183977	0.4067414	0.04570053
0.75376884	-6.57230998	1.2207454	0.4075212	0.04565274
0.76381910	-6.64695489	1.2230832	0.4082978	0.04560514
0.77386935	-6.72125118	1.2254110	0.4090710	0.04555773
0.78391960	-6.79520097	1.2277289	0.4098411	0.04551051
0.79396985	-6.86880640	1.2300370	0.4106079	0.04546348
0.80402010	-6.94206956	1.2323353	0.4113715	0.04541664
0.81407035	-7.01499254	1.2346239	0.4121319	0.04536998
0.82412060	-7.08757742	1.2369029	0.4128892	0.04532351
0.83417085	-7.15982623	1.2391724	0.4136433	0.04527723
0.84422111	-7.23174102	1.2414323	0.4143942	0.04523112
0.85427136	-7.30332380	1.2436827	0.4151421	0.04518520
0.86432161	-7.37457658	1.2459238	0.4158869	0.04513947
0.87437186	-7.44550135	1.2481555	0.4166285	0.04509391
0.88442211	-7.51610008	1.2503780	0.4173672	0.04504853

0.89447236	-7.58637472	1.2525912	0.4181028	0.04500333
0.90452261	-7.65632721	1.2547952	0.4188353	0.04495830
0.91457286	-7.72595948	1.2569902	0.4195649	0.04491346
0.92462312	-7.79527343	1.2591761	0.4202915	0.04486878
0.93467337	-7.86427097	1.2613529	0.4210151	0.04482428
0.94472362	-7.93295396	1.2635209	0.4217358	0.04477996
0.95477387	-8.00132428	1.2656799	0.4224535	0.04473580
0.96482412	-8.06938377	1.2678301	0.4231683	0.04469182
0.97487437	-8.13713427	1.2699715	0.4238802	0.04464801
0.98492462	-8.20457759	1.2721042	0.4245892	0.04460436
0.99497487	-8.27171556	1.2742282	0.4252954	0.04456088
1.00502513	-8.33854995	1.2763435	0.4259986	0.04451757
1.01507538	-8.40508255	1.2784503	0.4266991	0.04447443
1.02512563	-8.47131511	1.2805486	0.4273968	0.04443145
1.03517588	-8.53724940	1.2826384	0.4280916	0.04438863
1.04522613	-8.60288715	1.2847197	0.4287836	0.04434598
1.05527638	-8.66823008	1.2867927	0.4294729	0.04430348
1.06532663	-8.73327991	1.2888573	0.4301595	0.04426115
1.07537688	-8.79803832	1.2909137	0.4308432	0.04421898
1.08542714	-8.86250701	1.2929618	0.4315243	0.04417697
1.09547739	-8.92668765	1.2950018	0.4322027	0.04413512
1.10552764	-8.99058189	1.2970336	0.4328783	0.04409342
1.11557789	-9.05419138	1.2990573	0.4335513	0.04405188
1.12562814	-9.11751776	1.3010730	0.4342217	0.04401050
1.13567839	-9.18056264	1.3030807	0.4348893	0.04396927
1.14572864	-9.24332763	1.3050804	0.4355544	0.04392820
1.15577889	-9.30581434	1.3070723	0.4362168	0.04388727
1.16582915	-9.36802434	1.3090562	0.4368766	0.04384650
1.17587940	-9.42995921	1.3110324	0.4375339	0.04380589
1.18592965	-9.49162050	1.3130008	0.4381885	0.04376542
1.19597990	-9.55300977	1.3149614	0.4388406	0.04372510
1.20603015	-9.61412856	1.3169144	0.4394902	0.04368493
1.21608040	-9.67497838	1.3188598	0.4401372	0.04364491
1.22613065	-9.73556076	1.3207975	0.4407817	0.04360503
1.23618090	-9.79587719	1.3227277	0.4414237	0.04356530
1.24623116	-9.85592917	1.3246504	0.4420633	0.04352572
1.25628141	-9.91571817	1.3265656	0.4427003	0.04348628
1.26633166	-9.97524568	1.3284734	0.4433349	0.04344698
1.27638191	-10.03451314	1.3303739	0.4439670	0.04340783
1.28643216	-10.09352200	1.3322669	0.4445967	0.04336882
1.29648241	-10.15227371	1.3341527	0.4452240	0.04332995
1.30653266	-10.21076968	1.3360312	0.4458489	0.04329122
1.31658291	-10.26901134	1.3379025	0.4464714	0.04325263
1.32663317	-10.32700009	1.3397667	0.4470915	0.04321418
1.33668342	-10.38473732	1.3416236	0.4477092	0.04317587
1.34673367	-10.44222442	1.3434735	0.4483246	0.04313769
1.35678392	-10.49946277	1.3453164	0.4489376	0.04309965
1.36683417	-10.55645372	1.3471522	0.4495483	0.04306175
1.37688442	-10.61319864	1.3489810	0.4501567	0.04302399
1.38693467	-10.66969887	1.3508028	0.4507628	0.04298635

1.39698492	-10.72595574	1.3526178	0.4513666	0.04294886
1.40703518	-10.78197059	1.3544259	0.4519681	0.04291149
1.41708543	-10.83774472	1.3562272	0.4525674	0.04287426
1.42713568	-10.89327945	1.3580216	0.4531644	0.04283716
1.43718593	-10.94857606	1.3598093	0.4537591	0.04280019
1.44723618	-11.00363586	1.3615903	0.4543516	0.04276335
1.45728643	-11.05846011	1.3633646	0.4549419	0.04272664
1.46733668	-11.11305009	1.3651323	0.4555300	0.04269006
1.47738693	-11.16740705	1.3668933	0.4561159	0.04265360
1.48743719	-11.22153226	1.3686478	0.4566996	0.04261728
1.49748744	-11.27542695	1.3703957	0.4572812	0.04258108
1.50753769	-11.32909235	1.3721372	0.4578606	0.04254500
1.51758794	-11.38252969	1.3738721	0.4584378	0.04250905
1.52763819	-11.43574018	1.3756006	0.4590129	0.04247323
1.53768844	-11.48872504	1.3773227	0.4595859	0.04243753
1.54773869	-11.54148547	1.3790385	0.4601567	0.04240196
1.55778894	-11.59402264	1.3807479	0.4607255	0.04236650
1.56783920	-11.64633775	1.3824510	0.4612921	0.04233117
1.57788945	-11.69843197	1.3841478	0.4618567	0.04229596
1.58793970	-11.75030647	1.3858384	0.4624192	0.04226088
1.59798995	-11.80196240	1.3875228	0.4629797	0.04222591
1.60804020	-11.85340091	1.3892010	0.4635381	0.04219106
1.61809045	-11.90462314	1.3908731	0.4640944	0.04215633
1.62814070	-11.95563023	1.3925391	0.4646488	0.04212172
1.63819095	-12.00642330	1.3941990	0.4652011	0.04208722
1.64824121	-12.05700348	1.3958529	0.4657514	0.04205285
1.65829146	-12.10737186	1.3975008	0.4662997	0.04201859
1.66834171	-12.15752955	1.3991427	0.4668460	0.04198444
1.67839196	-12.20747766	1.4007786	0.4673904	0.04195041
1.68844221	-12.25721726	1.4024086	0.4679328	0.04191650
1.69849246	-12.30674943	1.4040328	0.4684732	0.04188270
1.70854271	-12.35607525	1.4056511	0.4690117	0.04184901
1.71859296	-12.40519578	1.4072635	0.4695482	0.04181544
1.72864322	-12.45411208	1.4088702	0.4700829	0.04178198
1.73869347	-12.50282520	1.4104711	0.4706156	0.04174863
1.74874372	-12.55133618	1.4120662	0.4711464	0.04171539
1.75879397	-12.59964606	1.4136557	0.4716753	0.04168226
1.76884422	-12.64775587	1.4152395	0.4722023	0.04164924
1.77889447	-12.69566663	1.4168176	0.4727275	0.04161633
1.78894472	-12.74337936	1.4183901	0.4732507	0.04158353
1.79899497	-12.79089505	1.4199570	0.4737722	0.04155084
1.80904523	-12.83821473	1.4215184	0.4742917	0.04151826
1.81909548	-12.88533937	1.4230742	0.4748095	0.04148578
1.82914573	-12.93226998	1.4246245	0.4753254	0.04145341
1.83919598	-12.97900752	1.4261694	0.4758394	0.04142115
1.84924623	-13.02555298	1.4277088	0.4763517	0.04138899
1.85929648	-13.07190732	1.4292427	0.4768622	0.04135694
1.86934673	-13.11807150	1.4307713	0.4773709	0.04132500
1.87939698	-13.16404648	1.4322945	0.4778777	0.04129315
1.88944724	-13.20983322	1.4338124	0.4783829	0.04126141

```
1.89949749 -13.25543264  1.4353249 0.4788862 0.04122978
1.90954774 -13.30084570  1.4368322 0.4793878 0.04119824
1.91959799 -13.34607331  1.4383342 0.4798876 0.04116681
1.92964824 -13.39111640  1.4398309 0.4803857 0.04113548
1.93969849 -13.43597590  1.4413225 0.4808821 0.04110425
1.94974874 -13.48065271  1.4428089 0.4813767 0.04107312
1.95979899 -13.52514773  1.4442901 0.4818697 0.04104209
1.96984925 -13.56946188  1.4457661 0.4823609 0.04101116
1.97989950 -13.61359604  1.4472371 0.4828504 0.04098033
1.98994975 -13.65755110  1.4487030 0.4833383 0.04094960
2.00000000 -13.70132794  1.4501638 0.4838244 0.04091897
```

The first column gives lambda, the second column gives intercept values while the third, fourth and fifth columns give values of intercepts of x1, x2 and x3 respectively.

But how to choose the value of lambda?

The best way to choose is by loking at the digits as it gives a subjective understanding. But if we want an objective way of choosing it, we use the command :

select(fit)

We get the output as

```
modified HKB estimator is 0.00080552
modified L-W estimator is 0.0005637402
smallest value of GCV  at 0.01005025
```

So, we notice that thare are certain methods of estimating the tuning parameter such as HKB, L-W, GCV(generalised cross validation) methods. Though the values for different methods are quite different but it doesnot matter much because when lambda is slightly more than 0, things start getting stabilised. So, this is known as ridge regression.

That's all for this session,we shall continue with Lasso in R lab in our next session.

Thank You for reading with patience.

NOTE : The whole R code that we used for this session is given below for reference purpose:

```
x1=1:100
x2=3*x1 + rnorm(100)/10
x3=x1*x1
tmp=c()
for(i in 1:100){
y = 1+2*x1+0.05*x3+4*rnorm(100)
tmp=rbind(tmp, lm(y~x1+x2+x3)$coef)}
library(MASS)
fit = lm.ridge(y~x1+x2+x3,lambda=seq(0,2,len=200))
matplot(fit$lambda,coef(fit)[,-1],ty='l')
coef(fit)
select(fit)
```

$\boxed{\textbf{NOTE :}}$ WHEN THE R CODE THAT IS BEING USED ABOVE IS RUN A NUMBER OF TIMES , YOU MAY NOTICE THAT THE PLOT CHANGES AND ALSO THE COEF RANGE IN THE PLOT CHANGES BUT THE BASIC IDEA OF ATTAINING STABILITY DOESNOT CHANGES WITH A SLIGHT INCREAMENT OF VALUE OF LAMBDA FROM 0. SO, WHEN YOU RUN THE CODE THE PLOT MAY NOT RESEMBLE THA ONE ATTACHED HERE. ALSO THE VALUES OF coef(fit) MAY NOT RESEMBLE THE SAME AS HERE.