

LASSO and Linear Regression

Parissa Amin

2022-10-09

Loading the data

Loading and scaling.

```
library(coefplot)

## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 4.2.3
library(glmnet)

## Loading required package: Matrix
## Warning: package 'Matrix' was built under R version 4.2.3
## Loaded glmnet 4.1-4
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library(ggplot2)
library(reshape2)

x<-read.csv("D:/MultiOmics/Data/FeatVal.csv")
rownames(x) <- x[,2]; x <- x[,-1:-2]
x <- x[, c(names(x)[-which(names(x) == "Response")], "Response")]
x[x=="False"] <- 0
x[x=="True"] <- 1

x[] <- lapply(x, as.numeric)
x[,1:93] <- scale(x[,1:93] )
x[x=="NaN"] <- 0

#Linear
```

```
Model<-lm(x$Response~.,data=x[,1:93])
summary(Model)
```

```
##
## Call:
## lm(formula = x$Response ~ ., data = x[, 1:93])
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.6444	-0.2117	0.0000	0.1710	0.6713

```
##
## Coefficients: (11 not defined because of singularities)
##
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.268456	0.032579	8.240	9.85e-12 ***
cin	0.001776	0.076056	0.023	0.9814
purity	-0.066051	0.079447	-0.831	0.4088
ploidy	0.064407	0.072600	0.887	0.3782
n_total	0.128607	0.067468	1.906	0.0610 .
fraction_cancer	-0.123079	1.250121	-0.098	0.9219
fraction_lymph	0.028595	0.945607	0.030	0.9760
fraction_stromal	0.052567	1.814061	0.029	0.9770
median_cancer_KDE_knn_50	0.258553	0.142552	1.814	0.0743 .
median_lymph_KDE_knn_50	-0.103958	0.116264	-0.894	0.3745
median_stromal_KDE_knn_50	-0.035424	0.112473	-0.315	0.7538
HRD	-0.020617	0.065503	-0.315	0.7539
Telomeric.AI	-0.066257	0.120842	-0.548	0.5853
LST	0.172471	0.083977	2.054	0.0440 *
HRD.sum	NA	NA	NA	NA
B2M	-0.127164	0.132906	-0.957	0.3422
TAP1	-0.139368	0.200173	-0.696	0.4887
TAP2	-0.027736	0.145411	-0.191	0.8493
HLA.A	0.020614	0.186716	0.110	0.9124
HLA.B	0.343580	0.253804	1.354	0.1804
HLA.C	-0.171880	0.125085	-1.374	0.1741
HLA.DPA1	-0.022335	0.238324	-0.094	0.9256
HLA.DPB1	-0.295190	0.278421	-1.060	0.2929
HLA.E	0.159876	0.166916	0.958	0.3416
HLA.F	-0.128836	0.169668	-0.759	0.4504
PD.1	0.035026	0.165653	0.211	0.8332
CTLA.4	-0.080318	0.217669	-0.369	0.7133
LAG3	0.025098	0.115139	0.218	0.8281
TIGIT	0.325854	0.271769	1.199	0.2348
TIM3	-0.001335	0.158109	-0.008	0.9933
PD.L1	0.021935	0.130433	0.168	0.8670
PD.L2	0.022408	0.137800	0.163	0.8713
CD27	-0.105652	0.212951	-0.496	0.6214
ICOS	0.301583	0.249847	1.207	0.2317
IDO1	0.023971	0.140454	0.171	0.8650
Act.CD4	0.056437	0.135069	0.418	0.6774
Act.CD8	-0.222905	0.240955	-0.925	0.3583
Tem.CD4	0.007398	0.105569	0.070	0.9443
Tem.CD8	0.350307	0.226179	1.549	0.1262
MDSC	-0.264728	0.153932	-1.720	0.0902 .
Treg	-0.049154	0.151816	-0.324	0.7471

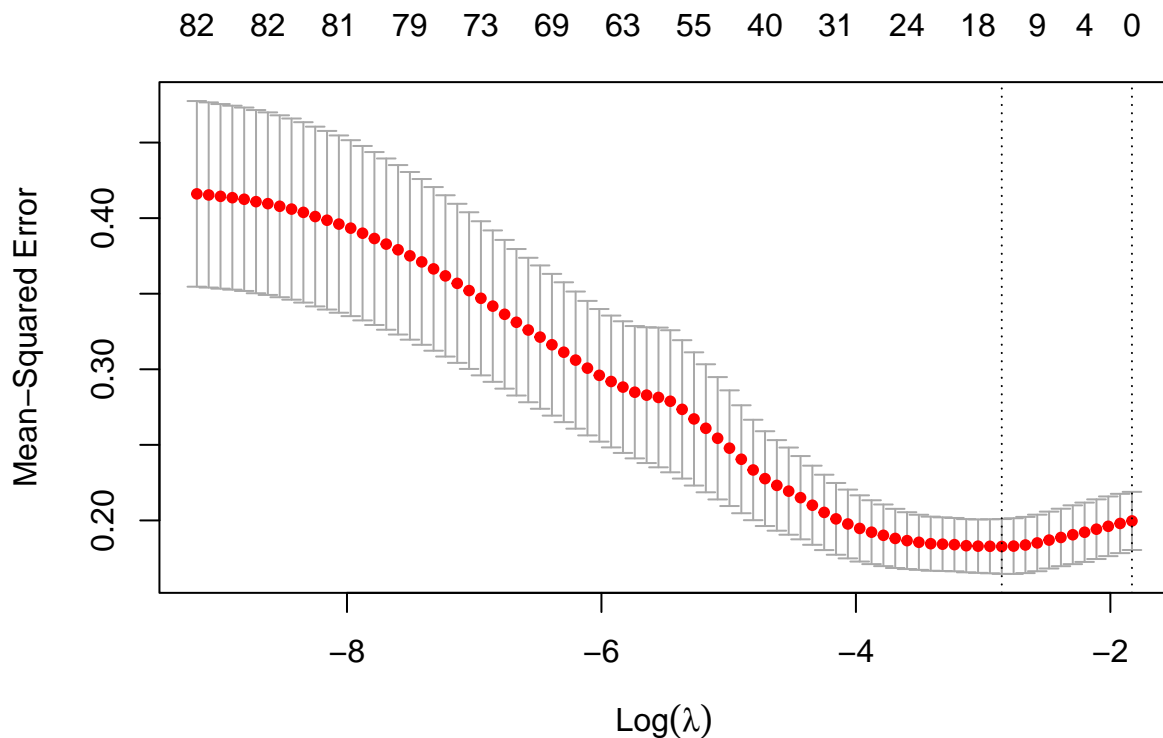
## T.cells	-0.149413	0.266184	-0.561	0.5765
## CD8.T.cells	0.047181	0.140722	0.335	0.7385
## Cytotoxic.lymphocytes	0.267646	0.193070	1.386	0.1703
## NK.cells	0.094826	0.126228	0.751	0.4552
## B.lineage	-0.029104	0.136213	-0.214	0.8315
## Monocytic.lineage	-0.226654	0.171172	-1.324	0.1900
## Myeloid.dendritic.cells	0.015437	0.079666	0.194	0.8469
## Neutrophils	-0.036579	0.067794	-0.540	0.5913
## Endothelial.cells	0.065313	0.123948	0.527	0.6000
## Fibroblasts	-0.175810	0.087599	-2.007	0.0488 *
## Signature.1	0.477342	0.337040	1.416	0.1614
## Signature.2	0.134591	0.096648	1.393	0.1684
## Signature.3	0.357294	0.218448	1.636	0.1067
## Signature.4	0.105934	0.071827	1.475	0.1450
## Signature.5	0.042791	0.069033	0.620	0.5375
## Signature.6	0.135720	0.082894	1.637	0.1063
## Signature.7	0.103558	0.068985	1.501	0.1381
## Signature.8	0.063032	0.072760	0.866	0.3895
## Signature.9	0.039605	0.093273	0.425	0.6725
## Signature.10	0.042678	0.086271	0.495	0.6225
## Signature.11	0.028530	0.061783	0.462	0.6458
## Signature.12	0.040741	0.058833	0.692	0.4911
## Signature.13	0.136636	0.115571	1.182	0.2413
## Signature.14	NA	NA	NA	NA
## Signature.15	0.112969	0.109160	1.035	0.3045
## Signature.16	0.026522	0.066103	0.401	0.6896
## Signature.17	NA	NA	NA	NA
## Signature.18	0.083046	0.079962	1.039	0.3028
## Signature.19	NA	NA	NA	NA
## Signature.20	0.017353	0.047038	0.369	0.7134
## Signature.21	0.002938	0.061131	0.048	0.9618
## Signature.22	NA	NA	NA	NA
## Signature.23	NA	NA	NA	NA
## Signature.24	0.109786	0.102697	1.069	0.2890
## Signature.25	0.068364	0.068664	0.996	0.3231
## Signature.26	0.079058	0.052971	1.492	0.1403
## Signature.27	0.006329	0.047552	0.133	0.8945
## Signature.28	NA	NA	NA	NA
## Signature.29	0.079463	0.105541	0.753	0.4542
## Signature.30	NA	NA	NA	NA
## Unknown	NA	NA	NA	NA
## NumberMutationsAnalysed	-0.020245	0.077584	-0.261	0.7950
## iC10	0.045166	0.052614	0.858	0.3937
## Stage	-0.124797	0.055196	-2.261	0.0271 *
## LNStatus	-0.036233	0.052840	-0.686	0.4953
## ER	-0.014188	0.077878	-0.182	0.8560
## HER2	NA	NA	NA	NA
## PAM50_Basal	-0.314313	0.310297	-1.013	0.3148
## PAM50_Her2	-0.173301	0.300364	-0.577	0.5659
## PAM50_LumA	-0.238659	0.245586	-0.972	0.3347
## PAM50_LumB	-0.301775	0.331042	-0.912	0.3653
## PAM50_Normal	-0.135567	0.150966	-0.898	0.3725
## PAM50_Unk	NA	NA	NA	NA
## ---				

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3977 on 66 degrees of freedom
## Multiple R-squared:  0.6433, Adjusted R-squared:  0.2001
## F-statistic: 1.452 on 82 and 66 DF,  p-value: 0.05886

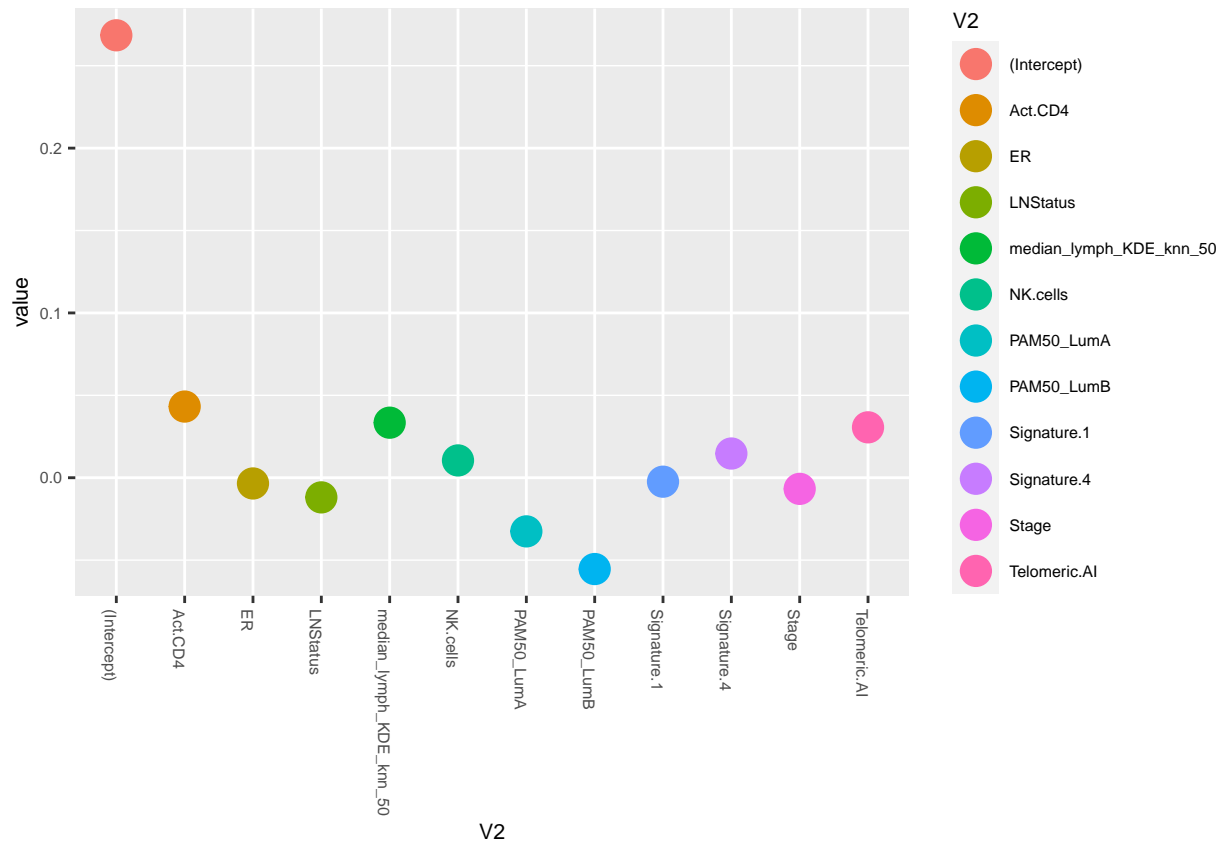
#LASSO
y1<-data.matrix(x[,1:93])
y2<-data.matrix(x$Response)

cv_model <- cv.glmnet(y1,y2, alpha = 1)
best_lambda <- cv_model$lambda.min
lasso_coef=predict(cv_model,type="coefficients",s=best_lambda)[1:93,]
nonzeroCoef<-lasso_coef[which(lasso_coef!=0)]
nonzeroCoef<- as.data.frame(nonzeroCoef)
nonzeroCoef[,2]<- rownames(nonzeroCoef)
nonzeroCoef.m<- melt(nonzeroCoef)

## Using V2 as id variables
Cross validation LASSO:
plot(cv_model)
```



```
ggplot(nonzeroCoef.m,aes(x=V2,y=value,color=V2))+geom_point(size=5)+
  theme(text = element_text(size =8))+
  theme(axis.text.x=element_text(angle = -90, hjust = 0))
```



Conlusion

Along side the ML and XAI model, I wanted to try two more traditional approaches to fit features to response to NAC. I used lineasr regression and and LASSO. While linear regression doesnt identify many significant features, lasso gives me 8 features that are relatively similar to the features I found to be important using SHAP values.