Statistical Learning for Multi-omics Analysis

16th Asian Institute in Statistical Genetics and Genomics Workshop
2022.07.21 ~ 2022.07.22

00 Index

01 Statistical Learning

02 Linear Regression

03 Penalized Regression

04 Shrinkage method

05 Elastic-net

06 Lasso vs Elastic-net

01 Statistical Learning

- Statistical learning is a set of tools for modeling and understanding complex datasets.
  - Supervised statistical learning
    - Regression : target variable(Y) is continuous
    - Classification : target variable(Y) is categorical
  - Unsupervised statistical learning : clustering

- Supervised Learning

  for a function(model) $Y = f(X) + \epsilon$ that represents the relationship between X and Y,

  figuring out $Y \approx \hat{f}(X)$

  1. Parametric : linear regression
  2. Non-parametric

02 Linear Regression

Assumption of β (regression coefficient) with f(X) defined as linear function

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

=> decide β which minimize error term. β also represent the effect of X to predict Y

: Ordinary Least Square Estimate (OLS)

- OLS

: unbiased estimates of β to have minimal variance

$$\mathrm{RSS}(\beta) = \sum \left(y_i - \beta_0 - \sum \beta_j x_{ij}\right)^2$$

cannot be computed when n (sample size) < p (feature size) (High dimensional data)

large variance when correlated variable is exist

=> Alternatives : penalized regression

03 Penalized Regression

can′t use linear regression to high-dimensional data such as gene expression data
=> add penalty term to linear regression

1. Lasso
2. Ridge

Shrinkage Methods

3. logistic likelihood : binary outcome
4. conditional logistic likelihood
   penalized likelihood for matched case-control outcome
   get case and control in one sample, usually used in epigenetics
5. Poisson likelihood : target variable is count data
6. Cox partial likelihood : censored survival outcome

04 Shrinkage method

- regularization method
- overcome disadvantage of OLS that cannot be adapted u high dimensional data
- reduce variance of β by regularizing the range of β to zero (shrinkage)

  => biased estimates.

  : whether β=0 or not is the only consideration rather than the real value of β
- Example

  1. Lasso (Least absolute shrinkage and selection operator)
  2. Ridge regression

04 Shrinkage Methods - Lasso

- use l1-norm as penalty term

$$RSS(\beta) + \lambda\|\beta\|_1 = \sum (y_i - \beta_0 - \sum \beta_j x_{ij})^2 + \lambda \sum |\beta_j|$$

- λ : tuning parameter. adjust proportion of penalty term of Lasso
  if λ = 0, Lasso = RSS(β)
  the number of degrees of freedom (DF) is different according to λ $(\lambda \propto \frac{1}{DF})$
  => feature selection
  when λ is large enough, β get closer to zero (shrinkage ↑)
  => decide λ with Cross-Validation (CV)

- optimal λ is defined according to One-Standard-Error Rule

- tend to select only one variable among all correlated variables and ignore the others
  => alternatives : Ridge

Shrinkage Methods – Lasso

- Cross Validation (CV)

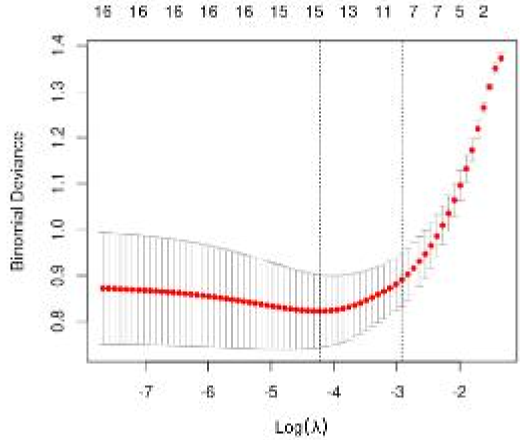
  : randomly separate n samples into K folds and compute β for each λ

  1. CVE based on MSE
  2. CVE based on deviance
  3. CVE based on classification error

5-fold CV

DATASET

Estimation 1	Test	Train	Train	Train	Train
Estimation 2	Train	Test	Train	Train	Train
Estimation 3	Train	Train	Test	Train	Train
Estimation 4	Train	Train	Train	Test	Train
Estimation 5	Train	Train	Train	Train	Test



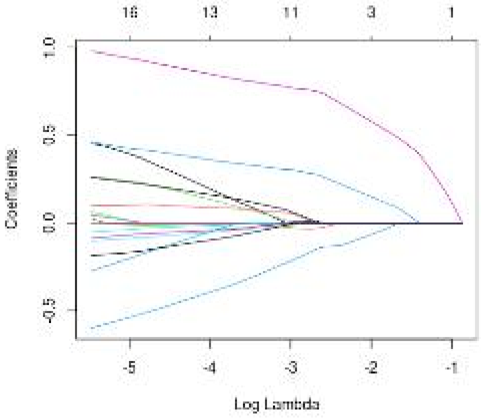
04 Shrinkage Methods - Ridge

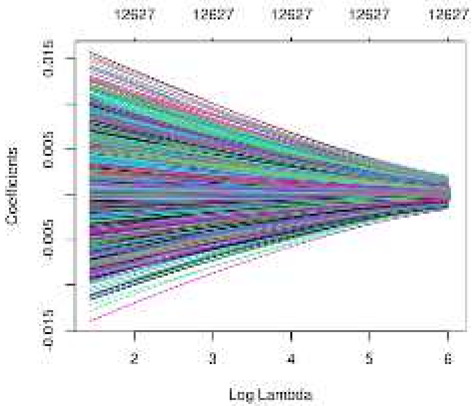
- use l2-norm as penalty term

$$RSS(\beta) + \lambda||\beta||_2 = \sum(y_i - \beta_0 - \sum \beta_j x_{ij})^2 + \lambda \sum \beta_j^2$$

- shrinkage β without making any β=0

  => feture selection impossible





05 Elastic-net

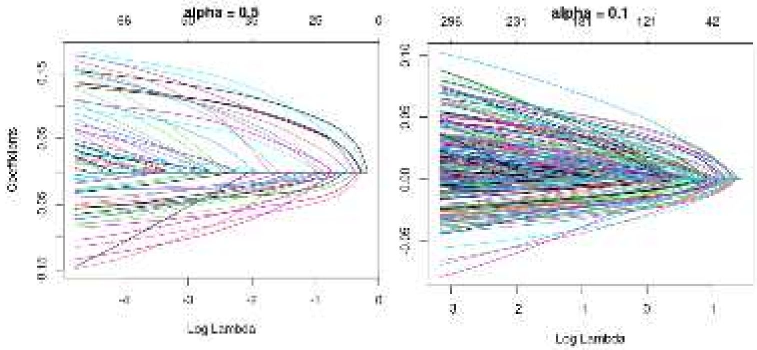
- use l1-norm and l2-norm together to take advantages of Lasso (feature selection) and Ridge (correlated variables)

$$RSS(\beta) + \lambda\alpha||\beta||_1 + \lambda(1-\alpha)||\beta||_2 = \sum(y_i - \beta_0 - \sum \beta_j x_{ij})^2 + \lambda\alpha\sum|\beta_j| + \lambda(1-\alpha)\sum\beta_j^2$$

- α : proportion of l1-norm and l2-norm

  the number of selected features increases according to α decreases.

  usually use α=0.1 or 0.5

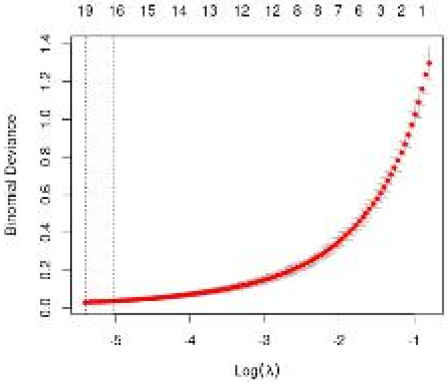


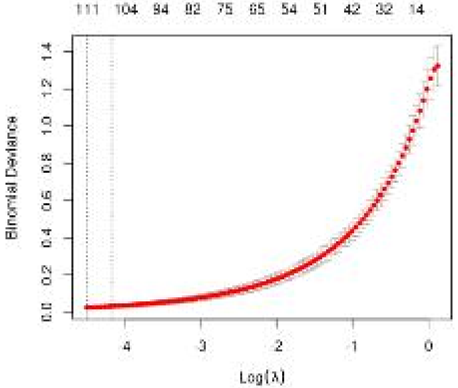
06 Lasso vs Elastic-net

```r
## Lasso
set.seed(1234)
gvc <- cv.glmnet(X_train, y_train, alpha=1, family="binomial", nfolds=5)
gvc$lambda.min
gvc$lambda.1se
plot(gvc) # left line : minimal CVE / right : one-standard-error rule

set.seed(111)
gvc <- cv.glmnet(X_train, y_train, alpha=0.4, family="binomial",nfolds=5)
plot(gvc)

EN <- assess.glmnet(gvc, newx=X_test, newy=y_test)
```






```
> lasso
$deviance
lambda.1se
0.02420614
attr(,"measure")
[1] "Binomial Deviance"

$class
lambda.1se
         0
attr(,"measure")
[1] "Misclassification Error"

$auc
[1] 1
attr(,"measure")
[1] "AUC"

$mse
 lambda.1se
0.001053166
attr(,"measure")
[1] "Mean-Squared Error"

$mae
lambda.1se
0.02364925
attr(,"measure")
[1] "Mean Absolute Error"
```

```
> EN
$deviance
lambda.1se
0.03487719
attr(,"measure")
[1] "Binomial Deviance"

$class
lambda.1se
         0
attr(,"measure")
[1] "Misclassification Error"

$auc
[1] 1
attr(,"measure")
[1] "AUC"

$mse
lambda.1se
0.00147683
attr(,"measure")
[1] "Mean-Squared Error"

$mae
lambda.1se
0.03410010
attr(,"measure")
[1] "Mean Absolute Error"
```

06 Lasso vs Elastic-net

- A predictive model, and predictors of under-five child malaria prevalence in Ghana: How do LASSO, Ridge and Elastic net regression approaches compare?
Aheto at el. Prev Med Rep. 2021 Jun 27;23:101475. doi: 10.1016/j.pmedr.2021.101475. PMID: 34306999; PMCID: PMC8258678.

Predictors (X)

child age, number of under-five children in a household,
has mosquito bed net for sleeping, sex of household head,
sex of a household member, household wealth,
dwelling sprayed against mosquito last 12 months,
sex of household head, child-anaemia status, has electricity in HH,
has a television in the household, place of residence,
the region of residence, number of household members,
number of children who slept under mosquito bed net previous night,
insecticide-treated net available in the household

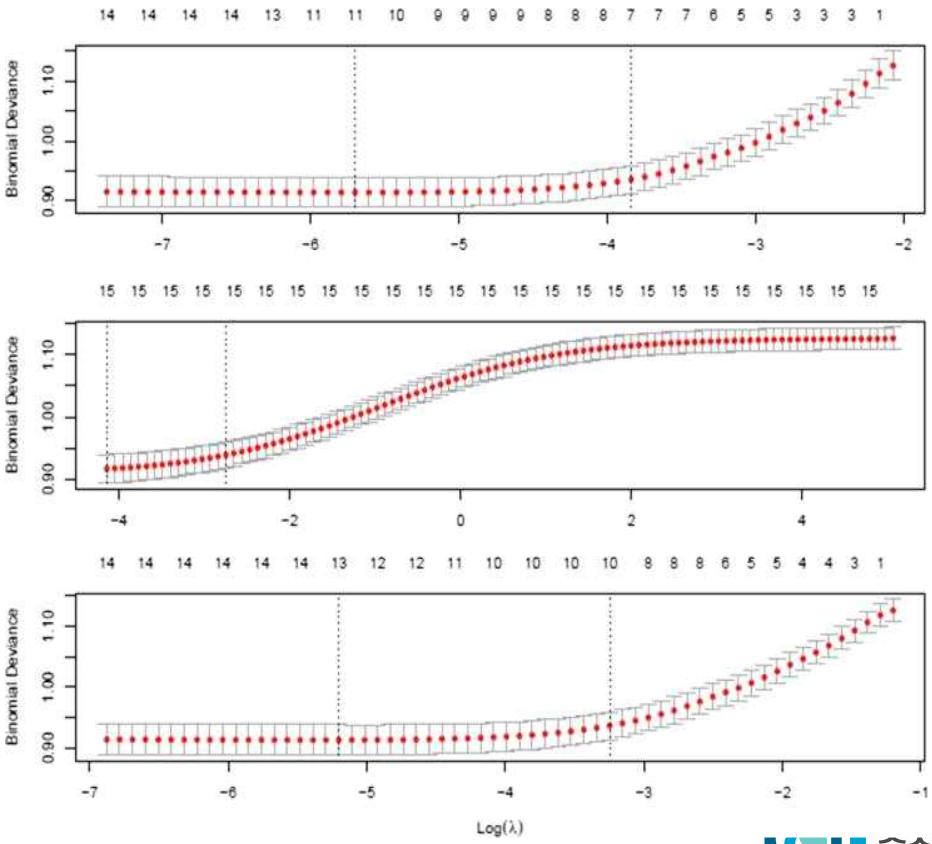
Lasso, Ridge,
Elastic-net

5-fold CV

Target (Y)

result of malaria rapid
diagnostic test (RDT) (0/1)

06 Lasso vs Elastic-net



Model	R-Square	RMSE (95% CI)	SD	AUC Value	Number
Lasso	0.196989	0.9489 (0.9286, 0.9691)	0.0202	81.20%	11
Ridge	0.1972966	1.0366 (1.0194, 1.0537)	0.0172	81.20%	15
Elastic net	0.1971316	0.9531 (0.9342, 0.9721)	0.0190	81.20%	13

Lasso

Ridge

Elastic Net (α = 0.4186508)

06 Lasso vs Elastic-net

	LASSO	RIDGE	ELASTIC NET
	alpha = 1	alpha = 0	alpha = 0.4186508
(Intercept)	1.043406015	0.81622002	0.957635393
Region	-0.125009418	-0.11751371	-0.125348706
Urban-rural residence	0.797182998	0.79393783	0.806272194
Has electricity in HH	-0.334558348	-0.35767631	-0.353889442
Has Television in HH	.	-0.06166143	-0.001048464
Sex of HH	0.019345178	0.07802045	0.047635449
Has mosquito bed net for sleeping	.	-0.08209747	.
Household wealth index	-0.356708586	-0.31547715	-0.352298541
sex of household member	.	-0.02806722	.
Anaemia level	-0.821035391	-0.77787099	-0.817498792
Dwelling sprayed against mosquito last 12 months	-0.350619294	-0.38498364	-0.367712726
Number of children who slept under mosquito bed net previous night	0.022397129	0.05179602	0.025362654
Number of U5C in household	0.003430091	0.03509552	0.02308751
Insecticide-treated net	0.136570517	0.18949137	0.157816524
Child Age	0.651001735	0.619984	0.652697693
number of household members	.	0.02590606	0.000677887

the age-related decline in malaria antibodies acquired from the mother during pregnancy as the child grows.

children from wealthy households are more likely to be living in affluent neighbourhoods with good drainage system and clean environments that decrease the breeding of mosquitoes thus decreasing the likelihood of mosquito bites and malaria (Dickinson et al., 2012)

Q & A