

Lasso方法在基于行为决定因素的宫颈癌早期检测中的应用

黄登香, 卢春婷

广西金融职业技术学院, 广西 南宁

收稿日期: 2022年1月16日; 录用日期: 2022年2月14日; 发布日期: 2022年2月21日

摘要

宫颈癌是世界上严重危害女性健康的恶性肿瘤之一, 所幸的是, 这种疾病是可以预防的。预防或早期发现是一个具有挑战性的难题, 本文利用Lasso方法、Adaptive Lasso方法、Elastic net方法和Adaptive Elastic net方法通过宫颈癌行为风险数据集建立Logistic模型, 以帮助进行宫颈癌早期检测和筛查。从实验结果看, Lasso方法表现更优。

关键词

Lasso, Adaptive Lasso, Elastic net, Adaptive Elastic net, 宫颈癌早期检测

Application of Lasso Procedure for Behavior Determinant Based Cervical Cancer Early Detection

Dengxiang Huang, Chunting Lu

Guangxi Financial Vocational College, Nanning Guangxi

Received: Jan. 16th, 2022; accepted: Feb. 14th, 2022; published: Feb. 21st, 2022

Abstract

Cervical cancer is one of the malignant tumors that seriously endanger women's health in the world. Fortunately, this disease can be prevented. Prevention or early detection is a challenging problem. In this paper, in order to help early detection and screening of cervical cancer, we consider the Lasso, adaptive Lasso, elastic net and adaptive elastic net to establish a logistic model

through the behavioral risk data set of cervical cancer. From the experimental results, Lasso procedure has good performance.

Keywords

Lasso, Adaptive Lasso, Elastic Net, Adaptive Elastic net, Early Detection of Cervical Cancer

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

宫颈癌是世界上严重危害女性健康的恶性肿瘤之一, 在全世界女性恶性肿瘤中, 其发病率位居第 4 位[1]。宫颈癌高危发病因素有人乳头瘤病毒(HPV)感染、不良的饮食习惯、不健康的生殖及性行为相关因素、精神创伤等[2]。由于引发宫颈癌的病因清楚, 宫颈癌不同于其他癌症的特点是该疾病是可以预防的。目前, 宫颈癌的一级和二级预防技术已相当成熟。主要的一级预防手段是注射人乳头瘤病毒(HPV)疫苗以预防 HPV 感染, 运用安全有效的 HPV 疫苗可从根本上阻断 HPV 传播, 这是目前最有效的宫颈癌预防措施[3], 该预防方法已在世界范围内多个国家开展。在我国, 宫颈癌也在一定程度上引起人们的重视, 逐渐增强的防癌意识, 能够较早地进行宫颈癌的筛查和防治, 是我国近年来宫颈癌死亡率较以往已有显著下降趋势的原因[4]。在宫颈癌防治方面, 医学领域已经掌握了较为成熟的策略和技术手段。然而, 由于对该疾病的知情及认知不足、地域条件和经济条件的限制等因素, 宫颈癌与癌前病变的筛查及早期诊断与早期治疗仍然是宫颈癌防治的难题。

早期发现恶性肿瘤在降低癌症特异性死亡率方面具有成本效益[5]。迄今为止, 行为学在宫颈癌早期诊断或一级预防中的应用还很有限, 从宫颈癌的行为决定因素中寻找宫颈癌的预测因子, 确定宫颈癌的患病风险是一项有意义的研究。当响应变量是定性变量, 表现为“成功”或“失败”等类似的情形时, 这时候常用的模型便是 Logistic 回归模型, 它是广泛应用的广义线性模型之一。本文结合从 UCI 机器学习数据库获取的宫颈癌行为风险数据集[6]建立 Logistic 回归模型, 使用 Lasso 方法, Adaptive Lasso 方法、Elastic net 方法和 Adaptive Elastic net 方法做变量选择, 根据行为决定因素对宫颈疾病的安全或危险性进行分类, 作为宫颈癌的早期发现或一级预防。通过对比发现, 四种方法在模型的预测中, Lasso 方法表现最优。

2. Lasso 方法及其相关方法

1996 年 Tibshirani 提出 Lasso 方法, 既能对变量进行选择, 又能得出参数估计值, 从而得到了广泛的应用和推广。考虑如下 Logistic 回归模型

$$P = E(Y | X) = \frac{\exp\{X'\beta\}}{1 + \exp\{X'\beta\}} \quad (2.1)$$

其中, $Y = (y_1, y_2, \dots, y_p)'$ 为响应变量, $y_i, i=1, \dots, p$ 独立同分布, $X = (x_1, x_2, \dots, x_p)'$ 为预测变量, $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ 稀疏。对固定非负数 λ , Logistic 回归模型的 Lasso 方法定义如下:

$$\hat{\beta}_{(\text{Lasso})} = \arg \min_{\beta} \left\{ \sum_{i=1}^n [-y_i (x_i' \beta) + \ln(1 + \exp\{x_i' \beta\})] + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

从 Lasso 方法的定义可以看出, 它对每个系数做相同程度的压缩, 可能导致对某些系数造成过度压缩, 后来, Zou 提出了 Adaptive Lasso (ALasso)方法, 应用于 Logistic 回归模型:

$$\hat{\beta}_{(\text{ALasso})} = \arg \min_{\beta} \left\{ \sum_{i=1}^n [-y_i (x'_i \beta) + \ln(1 + \exp\{x'_i \beta\})] + \lambda \sum_{j=1}^p \hat{w}_j |\beta_j| \right\}$$

其中, $\hat{w}_j = \left(|\hat{\beta}_{j,(\text{Ols})}| \right)^{-\gamma}$, $\hat{\beta}_{j,(\text{Ols})}$ 是最小二乘估计系数, $\gamma > 0$ 。

Adaptive Lasso 方法对每个系数进行加权, 克服了 Lasso 方法对系数过度压缩的缺点[7]。

Lasso 以及 Adaptive Lasso 两种方法对高维数据的处理具有一定的局限性。Zou 和 Hastie 为解决这一问题提出了 Elastic Net 变量选择方法[8]。

针对 Logistic 回归模型(2.1), 对固定的参数 (λ_1, λ_2) , Elastic Net (Enet)变量选择方法[9]定义如下:

$$\hat{\beta}_{(\text{Enet})} = \arg \min_{\beta} \left\{ \sum_{i=1}^n [-y_i (x'_i \beta) + \ln(1 + \exp\{x'_i \beta\})] + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}$$

Logistic 回归模型的 Adaptive Elastic Net (AEnet)方法的定义如下:

$$\hat{\beta}_{(\text{AEnet})} = \arg \min_{\beta} \left\{ \sum_{i=1}^n [-y_i (x'_i \beta) + \ln(1 + \exp\{x'_i \beta\})] + \lambda_1^* \sum_{j=1}^p \hat{w}_j |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}$$

其中, $\hat{w}_j = \left(|\hat{\beta}_{j,(\text{Enet})}| \right)^{-\gamma}$, γ 是一个正常数。

对于上述四种方法的计算, 本文使用 R 语言调用 gcdnet 包来完成计算。

3. 宫颈癌行为风险分析

行为研究广泛应用于健康科学和心理学等社会科学理论中, 常见的行为相关理论或模型有健康信念模型、保护动机理论、计划行为理论、社会认知理论等, 从这些理论和模型中可知, 行为的决定因素有七个, 即感知、意图、动机、主观规范、态度、社会支持和授权。研究表明, 参与者的社会支持可以改善宫颈癌预防行为[10], 社会心理学和健康心理学的一些理论认为, 意图导致并决定行为[11]。因此, 这些行为决定因素从一定程度上影响和导致行为的发生。Sobar [6]等人在这项研究中, 将七个决定因素和行为本身转化为问卷, 然后将问卷分发给 72 名受访者, 其中 21 名受访者为宫颈癌患者, 51 名受访者为非宫颈癌患者。根据这七个决定因素和行为作为特征或属性来创建一个 Logistic 回归模型, 以此作为宫颈癌患病风险的早期检测。

在本研究中, 为做变量选择建立 Logistic 回归模型, 需对数据进行预处理, 将 72 个数据集中的前 19 个变量做标准化处理, 依次用符号 V_1, V_2, \dots, V_{19} 表示, 最后一个变量 ca_cervix 作为分类变量, 表示受访者是否为宫颈癌患者, 其中宫颈癌患者记为 1, 非宫颈癌患者记为-1。利用 Lasso、ALasso、Enet、AEnet 四种方法分别做变量选择, 分析和比较四种方法在模型选择中的表现。为检测四种方法选出模型的预测精度, 在计算过程中, 将 72 个数据随机分取 3/4 作为训练集, 即随机选取 54 个数据作为训练集, 用于变量选择。余下的 1/4 共 18 个数据作为测试集, 用于测试各种方法的预测准确率。在这个过程中, 因数据案例较少, 只有 72 个, 本文将通过 10 折交叉验证法分别选出四种方法对应的最优模型, 在建立最优模型后, 使用测试集来检测模型预测精度。在这里, 模型预测精度(YCJD)是指用最优模型对测试集进行预测的准确率, 即模型预测精度 = 1 - 误分类误差(Misclassification Error), 而误分类误差指的是在使用测试集进行预测时被错误分类的比率。

本文将使用 R 语言来完成数据处理, 通过调用 gcdnet 包来进行计算。分别使用 Lasso、ALasso、Enet、

AEnet 四种变量选择方法来做变量选择, 并使用 10 折交叉验证选出各种方法对应的最优模型。分别使用四种方法计算得到各自变量的回归系数变化情况(系数路径图)和误分类误差结果如图 1~图 4 所示。对应的最优模型各参数值和系数结果如图 5 所示。

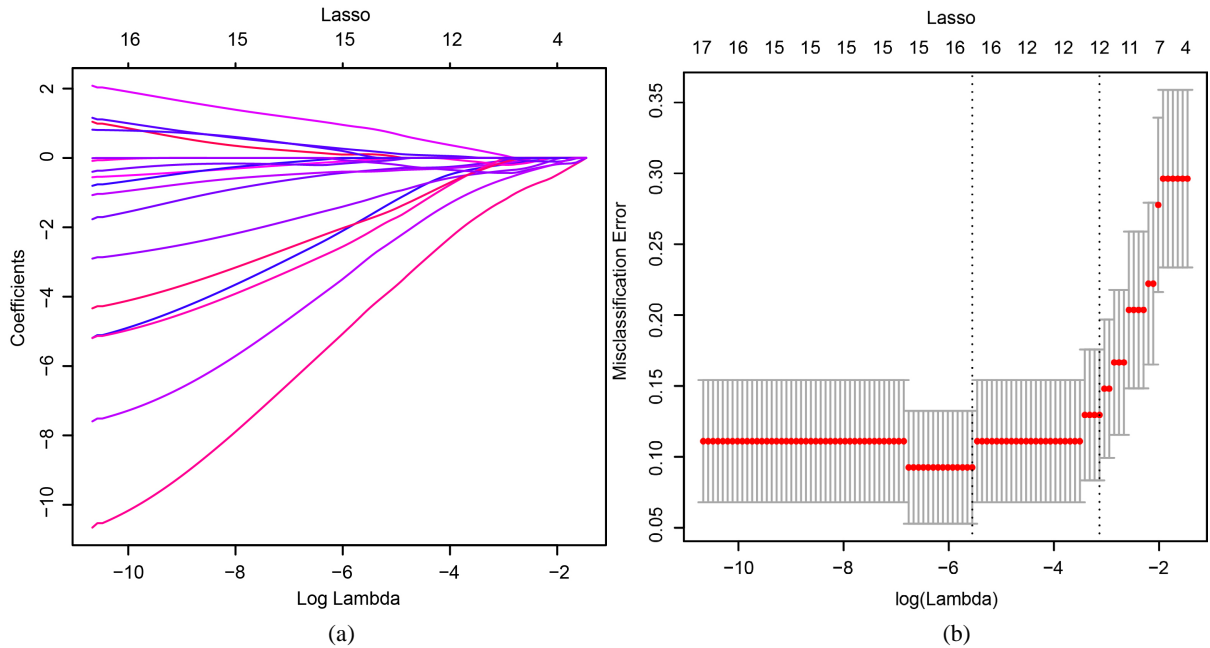


Figure 1. The Cross-Validation results of Lasso. (a) The coefficients paths at each step of Lasso; (b) The Misclassification Error of Lasso

图 1. Lasso 方法交叉验证结果。(a) Lasso 方法的系数路径图; (b) Lasso 方法的误分类误差结果图

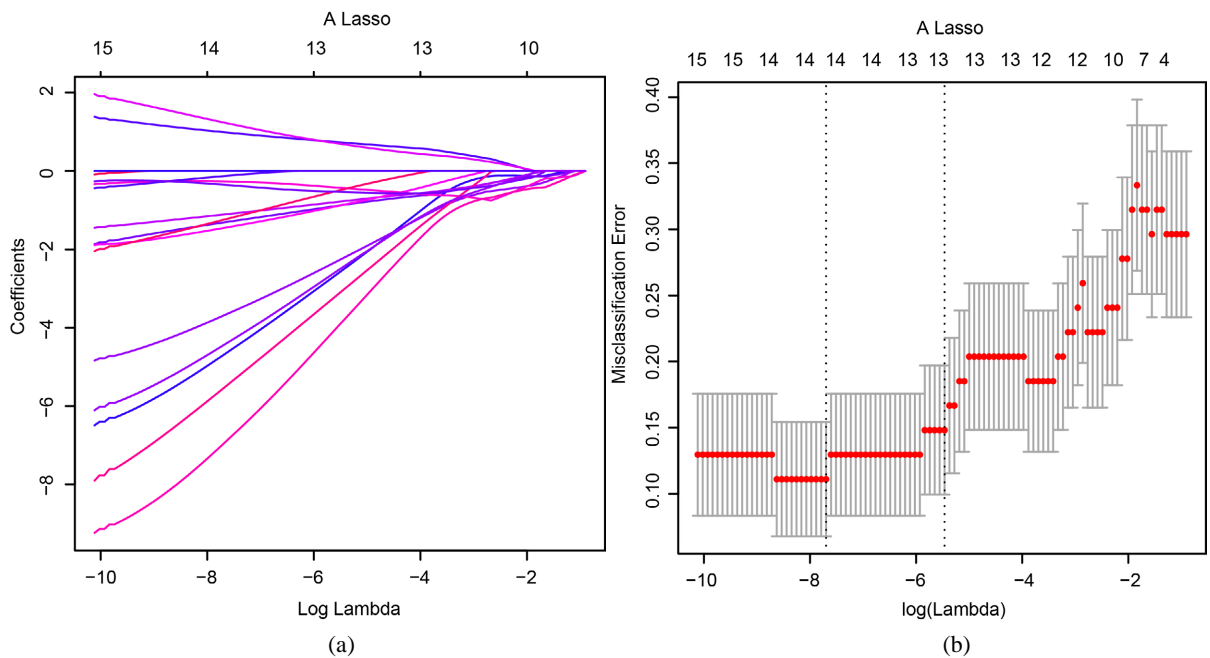


Figure 2. The Cross-Validation results of ALasso. (a) The coefficients paths at each step of ALasso; (b) The Misclassification Error of ALasso

图 2. ALasso 方法交叉验证结果。(a) ALasso 方法的系数路径图; (b) ALasso 方法的误分类误差结果图

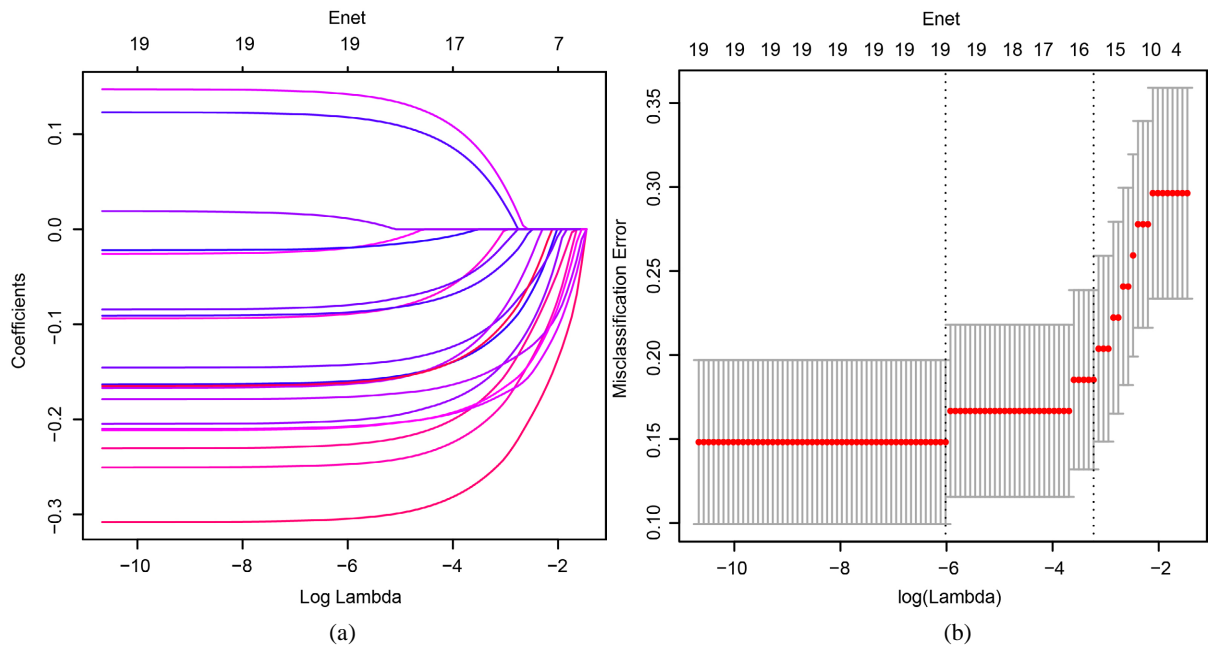


Figure 3. The Cross-Validation results of Enet. (a) The coefficients paths at each step of Enet; (b) The Misclassification Error of Enet

图 3. Enet 方法交叉验证结果。(a) Enet 方法的系数路径图; (b) Enet 方法的误分类误差结果图

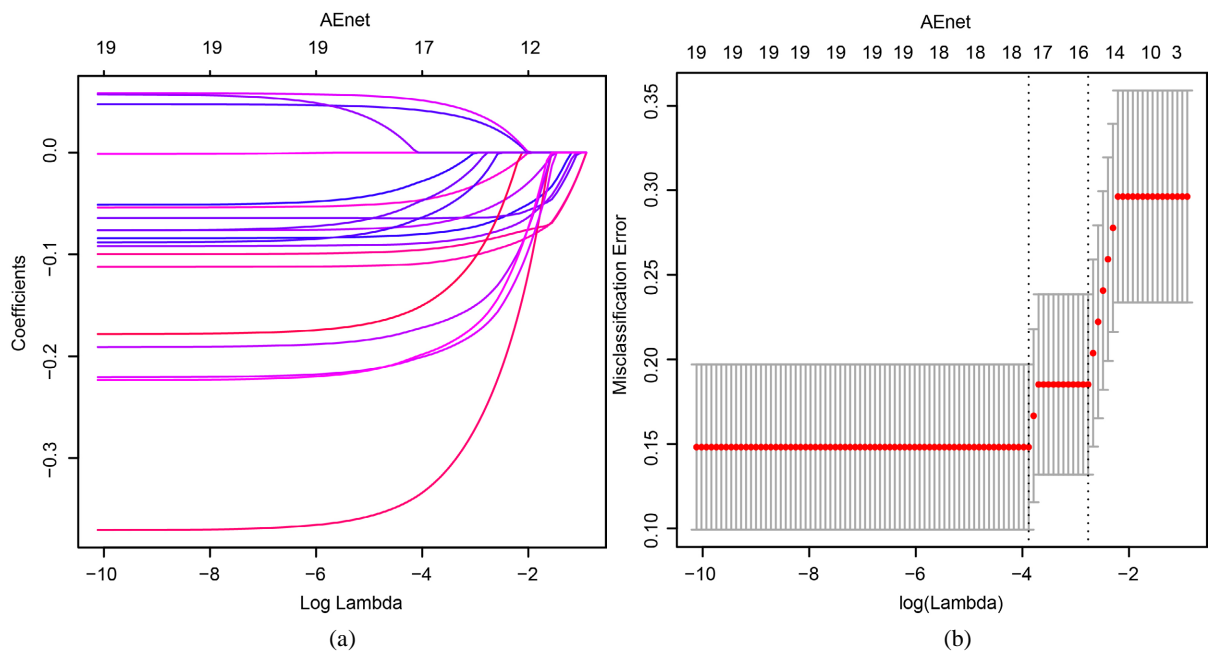


Figure 4. The Cross-Validation results of AEnet. (a) The coefficients paths at each step of AEnet; (b) The Misclassification Error of AEnet

图 4. AEnet 方法交叉验证结果。(a) AEnet 方法的系数路径图; (b) AEnet 方法的误分类误差结果图

在图 1~图 4 中, 横坐标均为参数 λ 的对数值, 最上方的横坐标是变量个数[12], 表示相应位置的 λ 所对应的四种方法做变量选择使用的自变量个数。各图的(a)中的纵坐标为自变量回归系数的取值, (b)中的纵坐标为误分类误差(Misclassification Error)。随着参数 λ 值的增加, 四种变量选择方法

使用的自变量数目都在减少, 同时误分类误差在增大。图(b)中的两条竖线分别指的是 λ_{\min} 和 λ_{lse} 的值, 当 λ 取 λ_{\min} 一值时, 表示此时得到的模型误分类误差最小, 上方的横坐标可以读出此时被选入模型的变量个数。当 λ 取值为 λ_{lse} 时, 得到的就是一个误分类误差在一个标准差范围内但自变量个数最少的模型。因为被研究数据集中变量的数量并不多, 因此, 我们使用模型误分类误差最小的那个模型, 即 λ 取值为 λ_{\min} 时的模型作为最优模型, 下面分别给出四种方法对应的最优模型参数值:

```
> sprintf("YCJD:%f", accuracy(Datest[,20],Prel))
[1] "YCJD:0.944444"
> Lasso1$lambda.min
[1] 0.003875657
> coef(Lasso1,s="lambda.min")
20 x 1 sparse Matrix of class "dgCMatrix"
      1
(Intercept) -5.11371844
V1          -1.70737741
V2           0.20531100
V3          -0.36848307
V4          -1.21233004
V5          -0.39272746
V6           0.85852450
V7          -0.05834362
V8           .
V9          -0.02453052
V10         -2.18614981
V11         -4.40842885
V12         -1.78033285
V13          0.10459436
V14          0.00000000
V15           .
V16          0.05336984
V17         -0.11348688
V18         -0.11843865
V19         -2.92788771
```

(a)

```
> sprintf("YCJD:%f", accuracy(Datest[,20],ALPrel))
[1] "YCJD:0.888889"
> ALasso1$lambda.min
[1] 0.0004544507
> coef(ALasso,s="lambda.min")
20 x 1 sparse Matrix of class "dgCMatrix"
      1
(Intercept) -9.3506052
V1          -4.6973727
V2           0.9930832
V3          -1.3140683
V4          -3.6906896
V5          -1.1112815
V6           1.2417050
V7          -1.4547883
V8           .
V9          -0.2647259
V10         -6.9643987
V11         -5.5467250
V12         -1.2457476
V13           .
V14           0.0000000
V15           0.0000000
V16           .
V17          -0.1296827
V18          -0.3505799
V19         -4.4482491
```

(b)

```
> sprintf("YCJD:%f", accuracy(Datest[,20],EPre))
[1] "YCJD:0.833333"
> ENet1$lambda.min
[1] 0.002434026
> coef(ENet,s="lambda.min")
20 x 1 sparse Matrix of class "dgCMatrix"
      1
(Intercept) -1.16724711
V1          -0.15987586
V2           0.11796007
V3          -0.14272457
V4          -0.20123190
V5          -0.16306154
V6           0.14225289
V7          -0.01978239
V8          -0.08929059
V9          -0.24733722
V10         -0.22651396
V11         -0.30469378
V12         -0.16151322
V13         -0.01941499
V14         -0.08747687
V15         -0.07989726
V16           0.01148028
V17         -0.17644465
V18         -0.20744026
V19         -0.20840261
```

(c)

```
> sprintf("YCJD:%f", accuracy(Datest[,20],AEPrel))
[1] "YCJD:0.777778"
> AENet1$lambda.min
[1] 0.02060879
> coef(AENet,s="lambda.min")
20 x 1 sparse Matrix of class "dgCMatrix"
      1
(Intercept) -1.05699940
V1          -0.07938814
V2           0.04065741
V3          -0.06451283
V4          -0.08850314
V5          -0.07040409
V6           0.05041270
V7           0.00000000
V8          -0.04525372
V9          -0.10832536
V10         -0.09482419
V11         -0.32936888
V12         -0.14665566
V13         -0.02583268
V14         -0.06117597
V15         -0.04450907
V16           0.00000000
V17         -0.17032816
V18         -0.19895710
V19         -0.19612811
```

(d)

Figure 5. The parameters and coefficients of the optimal model of Lasso, ALasso, Enet and AENet. (a) The optimal model results of Lasso; (b) The optimal model results of ALasso; (c) The optimal model results of ENet; (d) The optimal model results of AENet

图 5. 四种方法的最优模型参数和系数结果。(a) Lasso 方法计算结果; (b) ALasso 方法计算结果; (c) Enet 方法计算结果; (d) AENet 方法计算结果

1) 结合图 1(b)和图 5(a), 利用 Lasso 方法做变量选择, 当 lambda 的值为 0.00388 时, 选入模型的变量为 16 个, 此时的模型预测精度(YCJD)为 0.94444。

2) 结合图 2(b)和图 5(b), 利用 ALasso 方法做变量选择, 当 lambda 的值为 0.00045 时, 选入模型的变量为 14 个, 此时的模型预测精度(YCJD)为 0.88889。

3) 结合图 3(b)和图 5(c), 利用 Enet 方法做变量选择, 当 lambda 的值为 0.00243 时, 选入模型的变量为 19 个, 此时的模型预测精度(YCJD)为 0.83333。

4) 结合图 4(b)和图 5(d), 利用 AEnet 方法做变量选择, 当 lambda 的值为 0.02061 时, 选入模型的变量为 17 个, 此时的模型预测精度(YCJD)为 0.77778。

四种方法的最优模型中各变量的系数估计值如表 1:

Table 1. The coefficient estimates of the variables

表 1. 四种方法对应最优模型各变量的系数估计值

变量	Intercept	V1	V2	V3	V4	V5	V6
Lasso	-5.11372	-1.70738	0.20531	-0.36848	-1.21233	-0.39273	0.85852
ALasso	-9.35061	-4.69737	0.99308	-1.31407	-3.69069	-1.11128	1.24171
Enet	-1.16725	-0.15988	0.11796	-0.14272	-0.20123	-0.16306	0.14225
AEnet	-1.05700	-0.07939	0.04066	-0.06451	-0.08850	-0.07040	0.05041
变量	V7	V8	V9	V10	V11	V12	V13
Lasso	-0.05834	0.00000	-0.02453	-2.18615	-4.40843	-1.78033	0.10459
ALasso	-1.45479	0.00000	-0.26473	-6.96440	-5.54673	-1.24575	0.00000
Enet	-0.01978	-0.08929	-0.24734	-0.22651	-0.30469	-0.16151	-0.01941
AEnet	0.00000	-0.04525	-0.10833	-0.09482	-0.32937	-0.14666	-0.02583
变量	V14	V15	V16	V17	V18	V19	
Lasso	0.00000	0.00000	0.05337	-0.11349	-0.11844	-2.92789	
ALasso	0.00000	0.00000	0.00000	-0.12968	-0.35058	-4.44825	
Enet	-0.08748	-0.07990	0.01148	-0.17644	-0.20744	-0.20840	
AEnet	-0.06118	-0.04451	0.00000	-0.17033	-0.19896	-0.19613	

应注意的是, 需要检测四种方法得到的最优模型中, 哪一种更具有可解释性。我们对 72 个数据随机选取 54 个作为训练集, 18 个数据作为测试集, 对所得模型的优良性进行检测。在这里, 由于使用的是随机选取, 每一次不同的选取, 得到的结果自然不尽相同, 因此, 为使计算结果更具有说服力, 我们将四种变量选择方法都重复进行 20 次计算。对 20 次计算得到的模型预测精度求平均值、预测分类正确案例数求平均值和选入模型的变量个数求平均值, 得到表 2:

Table 2. The summary based on 20 replications

表 2. 20 次重复计算结果

变量选择方法	模型预测精度(平均值)	预测分类正确案例	选入模型的变量个数(平均值)
Lasso	0.88917	16	10
ALasso	0.85629	15	13
Enet	0.84715	15	17
AEnet	0.83270	15	18

根据表 2 中的结果, Lasso 方法的模型预测精度平均值最高, 为 0.88917; AEnet 方法的模型预测精度平均值最低, 为 0.83270。Lasso 方法使用测试集进行预测分类, 测试集中的 18 个案例被正确分类的案例数平均值为 16, 其他三种方法被正确分类的案例数平均值为 15; 对选入模型的变量个数进行比较, Lasso 方法选入模型的变量个数平均值最小, 为 10 个, ALasso、Enet、AEnet 三种方法选入模型的变量个数平均值依次递增。根据 ALasso 方法的定义可知, 它克服了 Lasso 方法对系数过度压缩的特点, 我们可以从表中看出这一点, ALasso 方法选入模型的变量个数平均为 13, 平均比 Lasso 方法多 3 个。Enet 方法与 AEnet 方法较前两种方法来说, 能够将具有强相关的变量一同选入模型, 从表中也可以看出, Enet 与 AEnet 方法选入模型的变量个数均比 Lasso 和 ALasso 方法要多。通过对以上结果的分析可知, 使用 Lasso、ALasso、Enet、AEnet 四种变量选择方法对上述宫颈癌数据进行变量选择, Lasso 方法表现最优, 其得到的最优模型以最少的变量个数, 达到最高的模型预测精度。ALasso、Enet、AEnet 三种方法的表现依次下降。因此, 我们可以利用 Lasso 方法创建的 Logistic 回归模型, 根据行为决定因素对宫颈疾病的安全或危险性进行分类, 作为宫颈癌的早期发现或一级预防。

4. 结语

我们知道, Lasso、ALasso、Enet、AEnet 四种变量选择方法各有优势, Lasso、ALasso 两种方法在处理维数较低、没有强相关变量的数据时, 效果是相当好的, 而 Enet、AEnet 两种方法更优于处理高维数据和具有强相关变量的数据。因此, 在使用 Lasso、ALasso、Enet、AEnet 四种方法做变量选择时, 我们还应当结合实际的数据来进行具体的分析。实际上, 由于宫颈癌数据只有 72 个案例, 19 个变量, 1 个分类变量, 数据量较少, 这对数据处理结果具有一定程度的影响, 所以, 利用 Lasso、ALasso、Enet、AEnet 四种方法对上述宫颈癌数据进行变量选择得到的结果并不是非常理想, 比如表现最好的 Lasso 方法, 它的模型预测精度为 0.88917, 测试集中 18 个案例被预测正确的约 16 个, 还有 2 个被错误预测。

利用 Lasso 方法创建的 Logistic 回归模型, 可以帮助女性对自己的行为因素做预判提供一定的参考, 通过知情个人健康行为对宫颈癌进行一级预防。将宫颈癌的行为决定因素作为宫颈癌的预测因子, 通过建立 Logistic 模型帮助进行早期检测和筛查, 具有成本效益。当然, 更加精确的情况还应当结合其他医疗手段等方式进行查证, 同时, 我们也希望能够收集到更多的数据, 以便建立更具有可解释性的模型, 从而根据行为决定因素得到宫颈癌预防的更多防范措施。

基金项目

2021 年度广西高校中青年教师科研基础能力提升项目: Elastic Net 方法在广义线性模型中的应用研究与实例分析(编号: 2021KY1513); 2020 年度广西金融职业技术学院中青年教师科研基础能力提升项目: Elastic Net 方法在广义线性模型中的应用研究与实例分析(编号: GXJZ202010); 2020 年度广西高校中青年教师科研基础能力提升项目: 基于模糊数学的综合评价法在高职院校学生顶岗效果评价的应用(编号: 2020KY50012)。

参考文献

- [1] Bray, F. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer Journal for Clinicians*, **68**, 394-424. <https://doi.org/10.3322/caac.21492>
- [2] 刘慧强. 我国宫颈癌流行病学特征和高危因素分析[J]. 中国妇幼保健, 2016, 12(6): 1258-1260.
- [3] 乔友林, 赵宇倩. 宫颈癌的流行病学现状和预防[J]. 中华妇幼临床医学杂志(电子版), 2015, 11(2): 141-147.
- [4] 包鹤龄, 刘温宁, 王黎君, 等. 中国 2006-2012 年子宫颈癌死亡情况与变化趋势分析[J]. 中华流行病学杂志, 2017, 38(1): 58-64.

-
- [5] van der Aalst, C.M., van Klaveren, R.J. and de Koning, H.J. (2010) Does Participation to Screening Unintentionally Influence Lifestyle Behaviour and Thus Life Style-Related Morbidity? *Best Practice & Research: Clinical Gastroenterology*, **24**, 465-478. <https://doi.org/10.1016/j.bpg.2010.06.001>
- [6] Sobar, R.M. and Wijaya, A. (2016) Behavior Determinant Based Cervical Cancer Early Detection with Machine Learning Algorithm. *Advanced Science Letters*, **22**, 3120-3023. <https://doi.org/10.1166/asl.2016.7980>
- [7] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. <https://doi.org/10.1198/016214506000000735>
- [8] Zou, H. and Hastie, T. (2005) A Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, **67**, 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [9] 李春红, 黄登香, 戴洪帅. Adaptive Elastic net 方法在 Logistic 回归模型中的应用[J]. 工程数学学报, 2015, 32(5): 759-771.
- [10] Larkey, L. (2006) Las Mujeres. Saludables: Reaching Latinas for Breast, Cervical and Colorectal Cancer Prevention and Screening. *Community Health*, **31**, 79-77. <https://doi.org/10.1007/s10900-005-8190-2>
- [11] Webb, T.L. and Sheeran, P. (2006) Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence. *Psychological Bulletin*, **132**, 249-268. <https://doi.org/10.1037/0033-2909.132.2.249>
- [12] 薛震, 孙玉林. R 语言——统计分析与机器学习[M]. 北京: 中国水利水电出版社, 2020.