Advances in Applied Mathematics 应用数学进展, 2023, 12(3), 1045-1053
Published Online March 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.123106

Hans汉斯

Lasso正则化参数选择的修正AIC方法研究

赵哲宇，王少春，尚美晨，王延新*

宁波工程学院理学院，浙江 宁波

收稿日期：2023年2月15日；录用日期：2023年3月9日；发布日期：2023年3月16日

摘 要

Lasso正则化方法是常用的变量选择方法。但Lasso正则化方法的优劣取决于能否选取出最优的正则化参数。本文在AIC准则的基础上，提出适用于Lasso正则化参数选择的修正的AIC (MAIC)准则。数据模拟及实例分析表明，Lasso方法在MAIC准则下能够以更高的概率选择正确的模型，MAIC准则明显优于其它参数选择方法。

关键词

Lasso，变量选择，正则化方法，MAIC准则，高维数据

A Modified AIC Method for the Selection of Lasso Regularization Parameters

Zheyu Zhao, Shaochun Wang, Meicheng Shang, Yanxin Wang*

College of Science, Ningbo University of Technology, Ningbo Zhejiang

Received: Feb. 15th, 2023; accepted: Mar. 9th, 2023; published: Mar. 16th, 2023

Abstract

Lasso regularization method is a commonly used variable selection method. However, the merits of Lasso regularization method depend on whether the optimal regularization parameters can be selected. Based on the AIC criterion, a modified AIC (MAIC) criterion is proposed for the selection of Lasso regularization parameter selection. Through data simulation and practical application, Lasso method can select the correct model with higher probability under MAIC criterion, and MAIC criterion is obviously superior to other parameter selection methods.

*通讯作者。

文章引用: 赵哲宇, 王少春, 尚美晨, 王延新. Lasso 正则化参数选择的修正 AIC 方法研究[J]. 应用数学进展, 2023, 12(3): 1045-1053. DOI: 10.12677/aam.2023.123106

Keywords

Lasso, Variable Selection, Regularization Parameter, MAIC, High-Dimensional Data

1. 引言

高维数据分析在科学、人文和工程的各个领域变得越来越重要。面对这些高维数据，需要采取合适的方法对其进行降维或变量选择。经典的变量选择准则，如 Akaike 提出的最小信息准则(AIC) [1]、贝叶斯信息准则(BIC) [2]以及交叉验证(CV)或广义交叉验证(GCV) [3]等。近年来，变量选择的正则化方法已逐渐流行起来，这些方法包括 Lasso [4]，SCAD [5]，MCP [6]，Elastic net [7]及自适应 Lasso [8]等。

稀疏正则化方法优劣的关键是选择正则化参数，CV 被广泛用于正则化参数的选择，然而在高维情况下 CV 会导致模型不稳定，并且为 Lasso 估计[9]选择了过多变量。GCV 是另一种广泛使用的方法，但众所周知，GCV 与 AIC 相似，具有较好的渐近有效性，但不满足变量选择的一致性。Wang、Li 和 Tsai [10]指出，常用的 GCV 不能选择令人满意的正则化参数，在 SCAD 估计方法中存在不可忽视的过拟合现象。AIC 准则本质上是相对熵损失的最核心部分的渐近无偏估计，非常适合在预测模型中权衡所估计模型的复杂度和此模型拟合数据的优良性，但其在分析大样本时有时并不明显，并且在很复杂的模型识别中有不适用的情况。因此，Akaike [11]进一步对准则做了修改，利用贝叶斯原理得到后验分布，再使似然函数最小的办法得出后来的 BIC 准则，该选择器能够一致的识别正确的模型。Wang 和 Leng [12]也为自适应 Lasso 做过类似的工作。

上述关于正则化参数选择的研究主要针对固定维数。近年来，在高维模型中引入了各种调整参数的选择方法。Wang，Li 和 Leng [13]提出了一种修正 BIC (MBIC)准则，适用于协变量的维数 p 小于样本量 n，并且维数 p 为发散维数的情况。在 $p > n$ 的高维情形下，Chen 和 Chen [14]提出了广义的 BIC (EBIC)。Wang 和 Zhu [15]将 EBIC 的结果推广到超高维情况，提出了一个高维 BIC (HBIC)准则。Fan 和 Tang [16]在广义线性模型的罚似然估计中提出了广义信息准则(GIC)。

最近，Hui，Warton 和 Foster 针对自适应 Lasso 提出了 ERIC 准则[17]。受 Lepski 非参数回归带宽选择方法的启发，一种新的方案AV_∞被提出，用于选择高维线性回归中 Lasso 正则化参数[18]。Li 和 Lederer [19]将该方法用于逻辑回归中。更多正则化参数选择方法可见文献 Wu 和 Wang [20]。

本文在 AIC 准则基础上，提出 MAIC 准则选择 Lasso 正则化参数。该方法克服了 AIC 准则过拟合的缺陷，既适用于低维情形，也适用于高维情形。通过数据模拟表明，MAIC 准则能够以更高的概率选择正确的模型。

2. 罚估计方法

正则化方法是一类特殊的收缩方法，能够同时实现变量选择和系数估计，是一种主流的变量选择方法。该方法是基于惩罚的思想，在最小二乘和最大似然估计的基础上，对目标函数施加一个惩罚项，从而使新的目标函数的最优子集相对于原目标函数的最优子集有所收缩，从而达到稀疏的目的。以线性回归模型为例，正则化方法基本形式如下：

$$\min_\beta \left\{ \|y - X\beta\|_2^2 + \sum_{j=1}^p J_\lambda\left(\beta_j\right) \right\} \tag{1}$$

其中 J_λ 是依赖于参数 $\lambda > 0$ 的惩罚项。

当惩罚项为对系数的 L1 范数，即为 Lasso 估计，如下所示：

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \tag{2}$$

Lasso 估计使用了绝对值形式的惩罚项，由于其在零处的不可导性质，使其在模型的参数估计过程中较小系数会因为惩罚项的存在从而直接压缩到 0，从而实现变量选择的目的。

3. Lasso 正则化参数选择方法

在实际应用中，正则化模型的优劣与正则化参数取值密切相关，不同正则化参数会导致不同的惩罚力度，进而影响最终的模型。本节主要研究 Lasso 估计的各种正则化参数选择方法。

3.1. 交叉验证

交叉验证(CV)是常用的正则化参数选择手段。在模型选择时把已有的数据分为三个部分，第一部分用以训练，第二部分用以检验，第三部分用以测试。如果检验集的数据量不足，使用交叉验证也可以作为一种检验模型的方法。交叉验证的基本思想是把样本分割成训练集以及检验集后先采用训练集建立模型，用刚建立的模型对检验集进行预测，并求检验集的预测误差(PE)，多次重复后对所产生的多个预测误差(PE)平均处理并进行记录。交叉检验的误差越小，说明模型拟合效果越好。

K 折交叉验证(K-fold CV)表达式如下，

$$\mathrm{CV}_\lambda(K) = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\left(y_i - X_i\beta\right)^2\right] \tag{3}$$

但 CV 方法容易产生过拟合现象，即在参数选择时，λ 容易过小，则非零 β 数量就会过多，造成模型的过拟合，从而不满足变量选择的一致性要求。

3.2. AIC 准则

信息准则有一个优势是计算信息准则的计算量比交叉检验要少。在常用的信息准则中比较有名的是赤池信息准则(AIC)。通常情况下，AIC 定义为：

$$\mathrm{AIC}(\lambda) = \log\left(\frac{\|y - X\beta\|^2}{n}\right) + \frac{2}{n}\widehat{df_\lambda} \tag{4}$$

其中 $\widehat{df_\lambda}$ 为广义自由度。

Tibshirani [4]介绍了怎样去计算 Lasso 自由度(DF)的近似值。Efron [21]阐述了 Lasso 的求解路径，并且用最小角回归 LARS 去求解 Lasso。而 Efron 认为自由度(DF)用非零系数项的个数进行近似能得到比较好的近似值：

$$\widehat{df_\lambda} = \left|\widehat{\mathcal{A}_\lambda}\right| \tag{5}$$

其中 $\hat{\mathcal{A}} = \left\{ j : \widehat{\beta_{\lambda j}} \neq 0 \right\}$ 代表非零系数项的集合，$\left|\widehat{\mathcal{A}_\lambda}\right|$ 是集合非零元素的个数。Zou [22]证明了如果有 $y \sim N\left(\mu, \sigma^2 I\right)$ 和 $\mathrm{rank}(X) = p$，那么对 Lasso 正则化问题自由度的估计为无偏和一致的。若无特殊注明，本文默认自由度使用上述公式进行计算。

3.3. BIC 准则

基于 BIC 准则选择正则化参数已被证明在满足模型选择的一致性[22]。BIC 准则定义为：

$$\mathrm{BIC}(\lambda)=\log\left(\frac{\left\|y-X\widehat{\beta_\lambda}\right\|^2}{n}\right)+\frac{\log(n)}{n}\widehat{df_\lambda} \tag{6}$$

$\widehat{df_\lambda}$ 的定义如式(5)所示。

但上述的 BIC 主要针对的是固定维度。Wang *et al.* [13]提出了 MBIC 准则，将 BIC 准则推广到参数发散 $p>n$ 的情况。定义如下

$$\mathrm{MBIC}(\lambda)=\log\left(\frac{\left\|y-X\widehat{\beta_\lambda}\right\|^2}{n}\right)+\widehat{df_\lambda}\frac{\log(n)}{n}C_n \tag{7}$$

其中 $C_n>0$ 是一些正常数，一般设置为 $C_n=\log\{\log(p)\}$。

然而当 $p>n$，特别是维数 p 相对于样本量 n 呈指数快速增长时，经典的 BIC 准则已经不适合。Chen 和 Chen [14]提出了一种广义的 BIC(EBIC)准则，

$$\mathrm{EBIC}(\lambda)=n\log\left(\frac{\left\|y-X\widehat{\beta_\lambda}\right\|^2}{n}\right)+\{\log(n)+2\gamma\log(p)\}\widehat{df_\lambda} \tag{8}$$

其中 $\gamma\in[0,1]$。现在已经被证明如果 $p=O(n^\kappa)$，并且 $\gamma>1-(2\kappa)^{-1}$，则 EBIC 满足变量选择的一致性。然而，尚不清楚在 $\log(p)=O(n^\kappa),0<\kappa<1$ 这样的超高维情况下是否 EBIC 仍然满足变量选择的一致性。因此 Wang 和 Zhu [15]提出了高维 BIC(HBIC)作为

$$\mathrm{HBIC}(\lambda)=n\log\left(\frac{\left\|y-X\widehat{\beta_\lambda}\right\|^2}{n}\right)+2\gamma\log(p)\widehat{df_\lambda} \tag{9}$$

其中 $\gamma\geq 1$。在一定的正则化条件下，HBIC 以概率趋于 1 选择真实模型。此外，还有 Wu 和 Wang 还提出了各种正则化参数选择方法[20]，这里不再赘述。

3.4. MAIC

AIC 准则为选择最优模型提供了有效手段。但是 AIC 准则容易选择过多的变量，因此本文提出修正 AIC 准则，即 MAIC。定义如下：

$$\mathrm{AIC}(\lambda)=\log\left(\frac{\left\|y-X\beta\right\|^2}{n}\right)+\frac{2}{n}\widehat{df_\lambda}+\log\left(\widehat{df_\lambda}\right) \tag{10}$$

其中自由度 $\widehat{df_\lambda}$ 的定义如式(5)所示。本文将通过数据模拟和实例分析证明 MAIC 准则能够以更高的概率选择真实的模型，不仅适用于低维情形，也适用于高维情形。同时不需要像 BIC 型准则一样，在不同的情形下需要选择不同的准则类型。

4. 数据分析

4.1. 模拟研究

模拟数据由线性回归模型生成：

$$Y=X\beta+\varepsilon,\varepsilon\sim n(0,\sigma^2)$$

其中，p 维向量 X 由多变量正态分布生成 $N(0,\Sigma)$ 生成的，并且预测因子之间的相关性为 $Cov(X_i,X_j)=\rho^{|i-j|}$，$i,j=1,\cdots,p$，且 $\rho=0.5$。所有模拟实验重复进行 100 次。

构建两个数值研究来说明 MAIC 方法的有效性。利用 Lasso 正则化方法进行变量选择，MAIC 准则选择正则化参数。为了说明 MAIC 准则的有效性，同时采用 AIC 准则和 BIC 型准则选择正则化参数。正则化模型由 ADMM 算法计算[23]。

为了评价方法的有效性和变量选择的一致性，考虑如下 7 个特征：1) MSE 表示 100 次重复实验中模型均方误差 ME 的中位数，即 $\mathrm{MSE}=\left\|\hat{\beta}-\beta^*\right\|_2$。2) SD 表示为 100 次重复实验中模型误差 ME 的标准差的中位数。3) C 表示 100 次重复实验中非零系数被正确估计为非零个数的均值。4) IC 表示 100 次重复实验中零系数被错误估计为非零个数的均值。5) Underfit 表示欠拟合，即在 100 次模拟实验中将非零系数错误估计为零的比例。6) Correctfit 表示正确拟合，即在 100 次模拟实验中将非零系数正确估计为非零且零系数正确被估计为零的比例；7) Overfit 表示过拟合，即在 100 次模拟实验中选择了所有重要变量并且包含了零系数被估计为非零的比例。

模拟 1：在这个例子中，变量个数 p 取 8，考虑 $p<n$ 的情况，另 $\beta=(3,1.5,0,0,2,0,0,0)^{\mathrm{T}}$，$\sigma^2=1$，$n=50$，100，200，模拟结果如表 1 所示。

Table 1. Simulation results one
表 1. 模拟 1 结果

n	准则	MSE	SD	C	IC	Underfit	Correctfit	Overfit
50	AIC	0.1410	0.1526	3.0000	2.7300	0	0.0600	0.9400
	MAIC	0.1912	0.1922	3.0000	**0.5700**	0	**0.6100**	**0.3900**
	BIC	**0.1387**	**0.1604**	3.0000	1.6600	0	0.1900	0.8100
100	AIC	**0.0753**	**0.0627**	3.0000	2.7700	0	0.0800	0.9200
	MAIC	0.1058	0.0986	3.0000	**0.1400**	0	**0.8600**	**0.1400**
	BIC	0.0749	0.0652	3.0000	1.0400	0	0.3800	0.6200
200	AIC	0.0352	0.0315	3.0000	2.2000	0	0.1900	0.8100
	MAIC	0.0437	0.0666	3.0000	**0.0100**	0	**0.9900**	**0.0100**
	BIC	**0.0319**	**0.0296**	3.0000	0.7400	0	0.4800	0.5200

从表 1，可以看出，首先所有方法的 C 值均为 3，说明所有模型都准确估计了非零系数。其次，在模型误差中，所有变量选择的方法模型误差相当，MAIC 准则的误差略微偏高。再次， Lasso 方法在 MAIC 准则具有更小的 IC 值，能够以更高概率选择真实模型，选择正确模型的比例明显高于 AIC 和 BIC 准则。特别是随着样本量 n 的增加，选择正确模型的比例趋于 1。

模拟 2：考虑 $p>n$ 的高维情况，令 $\sigma^2=2$，$\beta=(3,1.5,2,2,2,\cdots,0)^{\mathrm{T}}$。取 n = 200 且 p = 200、400、1000、2000 的四种维数，模拟结果如表 2 所示。

从表 2，可以得出如下结论。首先，在高维情况下 AIC 显然是一个失败的选择器，不能识别真正的模型。其次，MAIC、MBIC、EBIC 和 HBIC 都能够一致的选择真实模型，但是 MAIC 选择正确模型的比例更高。第三，各种方法的模型误差相当，相比模拟 1 的低维情形，MAIC 准则的误差偏高的态势并不明显。当然，这也表明维度越高 MAIC 表现越好，MAIC 更适用于高维情况下的正则化参数的选择。

Table 2. Simulation results two
表 2. 模拟 2 结果

p	准则	MSE	SD	C	IC	Underfit	Correctfit	Overfit
200	AIC	3.9769	1.4952	5.0000	107.0900	0	0.0100	0.9900
	MAIC	0.3629	0.2162	5.0000	**0.0100**	0	**0.9900**	**0.0100**
	BIC	**0.3129**	**0.1683**	5.0000	1.0200	0	0.4000	0.6000
	MBIC	0.3241	0.1822	5.0000	0.3200	0	0.7300	0.2700
400	AIC	3.8139	0.5886	5.0000	158.5800	0	0	1.0000
	MAIC	0.4191	0.2253	5.0000	**0**	0	**1.0000**	**0**
	BIC	**0.3570**	**0.1937**	5.0000	1.1500	0	0.4300	0.5700
	MBIC	0.3895	0.2055	5.0000	0.3600	0	0.7400	0.2600
	EBIC	0.4053	0.2074	5.0000	0.2000	0	0.8400	0.1600
1000	AIC	2.4308	0.3419	5.0000	180.3500	0	0	1.0000
	MAIC	0.4545	0.1939	5.0000	**0.0100**	0	**0.9900**	**0.0100**
	EBIC	**0.4368**	**0.1869**	5.0000	0.1200	0	0.9100	0.0900
	HBIC	0.4462	0.1886	5.0000	0.0200	0	0.9800	0.0200
2000	AIC	1.9577	0.3030	5.0000	189.5500	0	0	1.0000
	MAIC	0.4856	0.2153	5.0000	**0.0100**	0	**0.9900**	**0.0100**
	HBIC	**0.4591**	**0.1949**	5.0000	0.0100	0	0.9000	0.1000

4.2. 实际数据分析

本文利用由 Deeksha Russell 和 Duan Wang 从世界卫生组织下属的全球卫生观察站(GHO)数据库及联合国网站整理而成的数据集 Life Expectancy Data，该数据集包含 22 个变量，2938 个样本，描述关于 193 个国家从 2000 年到 2015 预期寿命及其影响因素的全部数据或部分数据，这意味着有 20 个预测变量，经过预处理，即删除缺失值等后共获取 1649 个样本，其中包含 133 个国家从 2000 年到 2015 年的全部数据或部分数据。数据集各个变量描述如下：

X_1：年份；X_2：国家状态(发展中国家：1，发达国家：2)；X_3：每 1000 人中的成人(15 至 60 岁)死亡数；X_4：每 1000 人中的未成年人死亡数；X_5：15 岁以上人均酒精消费量(纯酒精升数)；X_6：卫生支出占人均国内生产总值的百分比；X_7：1 岁儿童乙型肝炎(HepB)免疫接种覆盖率；X_8：每 1000 人中报告的麻疹病例数；X_9：全部人口的平均体重指数；X_{10}：每 1000 人中五岁以下死亡人数；X_{11}：1 岁儿童脊髓灰质炎(Pol3)免疫接种覆盖率；X_{12}：政府卫生总支出占政府总支出的百分比；X_{13}：1 岁儿童白喉、破伤风和百日咳疫苗接种覆盖率；X_{14}：每 1000 名活产婴儿中的艾滋病毒/艾滋病死亡人数(0~4 岁)；X_{15}：人均国内生产总值(美元)；X_{16}：人口数；X_{17}：10 至 19 岁儿童和青少年的偏瘦率；X_{18}：5 至 9 岁儿童的偏瘦率；X_{19}：资源收入构成方面的人类发展指数(0~1)；X_{20}：受教育年限；Y：预期寿命。

假设各国人预期寿命与这些因素呈线性关系，即有如下线性模型：

$$y_i = \sum_{j=1}^{20} x_{ij}\beta_j + \varepsilon_i, i = 1, 2, \cdots, n$$

其中 y_i 表示第 i 个样本的预期寿命，x_{ij} 是它的第 j 个变量，ε_i 是均值为 0，方差为 σ^2 的 i.i.d 的随机误差项。

首先将数据划分为训练集(90%)及测试集(10%)，在训练集上选择模型，在测试集上预测并计算预测误差(PE)，PE 定义如下：

$$\mathrm{PE} = \frac{1}{n}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2$$

基于不同准则，利用最小二乘估计(OLS)、Lasso 估计分析该数据。变量选择结果如表 3 所示。从表 3 可以看出，无罚的最小二乘估计(OLS)选择了所有的变量(由于小数精度原因，其中有 3 个变量系数绝对值为 0.0000)；AIC、BIC 及 CV 均选择了 16 个变量，其中 AIC 及 BIC 估计结果相同；MAIC 则选择了最少的 12 个变量，在选择了相对稀疏的模型的同时，误差并没有明显增大，R^2 值接近最小二乘。不同方法选择的模型在测试集上预测误差以及 R^2 值如表 4 所示，因此该用方法作变量选择效果更好。

Table 3. Parameter estimation under different methods
表 3. 不同方法下的参数估计

变量	OLS	AIC	MAIC	BIC	CV
X_1	0.0261	0.0262	0.0273	0.0262	0.0262
X_2	0.9991	0.9229	0	0.9229	0.8848
X_3	−0.0163	−0.0164	−0.0149	−0.0164	−0.0165
X_4	0.0928	0.0932	0.0984	0.0932	0.0934
X_5	−0.0879	−0.0809	0	−0.0809	−0.0775
X_6	0.0003	0	0	0	0
X_7	−0.0075	−0.0074	0	−0.0074	−0.0074
X_8	−0.0000	0	0	0	0
X_9	0.0322	0.0326	0.0397	0.0326	0.0328
X_{10}	−0.0697	−0.0700	−0.0734	−0.0700	−0.0701
X_{11}	0.0094	0.0094	0	0.0094	0.0094
X_{12}	0.1124	0.1108	0.0808	0.1108	0.1099
X_{13}	0.0135	0.0137	0.0192	0.0137	0.0139
X_{14}	−0.4427	−0.4432	−0.4493	−0.4432	−0.4434
X_{15}	0.0000	0	0	0	0
X_{16}	−0.0000	0	0	0	0
X_{17}	−0.0202	−0.0206	−0.0313	−0.0206	−0.0208
X_{18}	−0.0479	−0.0480	−0.0494	−0.0480	−0.0481
X_{19}	10.1277	9.6578	1.5495	9.6578	9.4241
X_{20}	0.8427	0.8580	1.1308	0.8580	0.8657

Table 4. Prediction error of models selected under different methods on test set
表 4. 不同方法下选择模型在测试集上的预测误差

	OLS	AIC	MAIC	BIC	CV
PE	13.9143	13.9171	14.3470	13.9171	13.8665
R^2	0.8508	0.8472	0.8360	0.8472	0.8474

5. 结论

本文提出了 MAIC 准则选择 Lasso 正则化参数，它不仅适用于低维情况，而且也适用于高维情况。模拟结果表明，MAIC 方法满足变量选择的一致性。此外，与 BIC 型选择器相比，MAIC 方法在高维情况下表现得更好。当然，本文还没有讨论 MAIC 方法的理论性质，如变量选择的一致性，这将是未来的研究课题。此外，本文提出的 MAIC 准则可以推广到非凸罚方法(如 SCAD 和 MCP 等)的正则化参数选择。

基金项目

国家级大学生创新创业训练计划项目(202211058015)；宁波市自然科学基金项目(2021J143, 2021J144)。

参考文献

[1] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, 267-281.

[2] Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461-464. https://doi.org/10.1214/aos/1176344136

[3] Craven, P. and Wahba, G. (1979) Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerische Mathematik*, **31**, 377-403. https://doi.org/10.1007/BF01404567

[4] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[5] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. https://doi.org/10.1198/016214501753382273

[6] Zhang, C.-H. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. https://doi.org/10.1214/09-AOS729

[7] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), **67**, 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[8] 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

[9] Lim, C. and Yu, B. (2016) Estimation Stability with Cross-Validation (ESCV). *Journal of Computational and Graphical Statistics*, **25**, 464-492. https://doi.org/10.1080/10618600.2015.1020159

[10] Wang, H., Li, R. and Tsai, C.L. (2007) Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method. *Biometrika*, **94**, 553-568. https://doi.org/10.1093/biomet/asm053

[11] Akaike, H. (1977) On Entropy Maximization Principle. *Application of Statistics*, **543**, 27-41.

[12] Wang, H. and Leng, C. (2007) Unified Lasso Estimation by Least Squares Approximation. *Journal of the American Statistical Association*, **102**, 1039-1048. https://doi.org/10.1198/016214507000000509

[13] Wang, H., Li, B. and Leng, C. (2009) Shrinkage Tuning Parameter Selection with a Diverging Number of Parameters. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), **71**, 671-683. https://doi.org/10.1111/j.1467-9868.2008.00693.x

[14] Chen, J. and Chen, Z. (2008) Extended Bayesian Information Criteria for Model Selection with Large Model Spaces. *Biometrika*, **95**, 759-771. https://doi.org/10.1093/biomet/asn034

[15] Wang, T. and Zhu, L. (2011) Consistent Tuning Parameter Selection in High Dimensional Sparse Linear Regression. *Journal of Multivariate Analysis*, **102**, 1141-1151. https://doi.org/10.1016/j.jmva.2011.03.007

[16] Fan, Y. and Tang, C.Y. (2013) Tuning Parameter Selection in High Dimensional Penalized Likelihood. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), **75**, 531-552. https://doi.org/10.1111/rssb.12001

[17] Hui, F.K., Warton, D.I. and Foster, S.D. (2015) Tuning Parameter Selection for the Adaptive Lasso Using ERIC. *Journal of the American Statistical Association*, **110**, 262-269. https://doi.org/10.1080/01621459.2014.951444

[18] Chichignoud, M., Johannes, L. and Wainwright, M. (2016) A Practical Scheme and Fast Algorithm to Tune the Lasso with Optimality Guarantees. *Journal of Machine Learning Research*, **17**, 1-20.

[19] Li, W. and Lederer, J. (2019) Tuning Parameter Calibration in High-Dimensional Logistic Regression with Theoretical

Guarantees. *Journal of Statistical Planning and Inference*, **202**, 80-98. https://doi.org/10.1016/j.jspi.2019.01.006

[20] Wu, Y. and Wang, L. (2020) A Survey of Tuning Parameter Selection for High-Dimensional Regression. *Annual Review of Statistics and Its Application*, **7**, 209-226. https://doi.org/10.1146/annurev-statistics-030718-105038

[21] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least Angle Regression. *The Annals of Statistics*, **32**, 407-451. https://doi.org/10.1214/009053604000000067

[22] Zou, H., Hastie, T. and Tibshirani, R. (2007) On the "Degrees of Freedom" of the Lasso. *The Annals of Statistics*, **35**, 2173-2192. https://doi.org/10.1214/009053607000000127

[23] Zhu, Y.Z. (2017) An Augmented ADMM Algorithm with Application to the Generalized Lasso Problem. *Journal of Computational and Graphical Statistics*, **26**, 195-204. https://doi.org/10.1080/10618600.2015.1114491