Partial Variable Selection and Its' Applications in Biostatistics

Jingwen Gu[1], Ao Yuan's[1,2]* and Ming T Tan[1]

[1]*Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, USA*

[2]*Department of Epidemiology and Biostatistics Section, Rehabilitation Medicine, National Institutes of Health, USA*

Submission: December 23, 2017; **Published:** April18, 2018

*Corresponding author: Ao Yuan, Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington DC 20057, USA; Email: ay312@georgetown.edu

Abstract

We propose and study a method for partial covariates selection, which only select the covariates with values fall in their effective ranges. The coefficients estimates based on the resulting data is more interpretable based on the effective covariates. This is in contrast to the existing method of variable selection, in which some variables are selected/deleted in whole. To test the validity of the partial variable selection, we extended the Wilks theorem to handle this case. Simulation studies are conducted to evaluate the performance of the proposed method, and it is applied to a real data analysis as illustration.

Keywords : Covariate; Effective range; Partial variable selection; Linear model; Likelihood ratio test

Abbreviations : NIH: The National Institutes of Health; UPDRS: Unified Parkinson's Disease Rating Scale

Introduction

Variables selection is a common practice in biostatistics and there is vast literature on this topic. Commonly used methods include the likelihood ratio test [1], AIC [2], BIC [3], the minimum description length [4,5], etc. The principal components models linear combinations of the original covariates, reduces large number of covariates to a handful of major principal components, but the result is not easy to interpret in terms of the original covariates. The stepwise regression starts from the full model, and deletes the covariate one by one according to some statistical significance measure. May et al. [6] addressed variable selection in artificial neural network models, Mehmood et al. [7] gave a review for variable selection with partial least squares model. Wang et al. [8] addressed variable selection in generalized additive partial linear models. Liu et al. [9] addressed variable selection in semiparametric additive partial linear models.

The Lasso [10,11] and its variation [12,13] are used to select some few significant variables in presence of large number of covariates. However, existing methods only select the whole variable(s) to enter into/delete from the model, which may not the most desirable in some bio-medical practice. For example, in the heart disease study [14,15], there are more than ten risk factors identified by medical researchers in their long time investigations, with the existing variable selection methods, some of the risk factors will be deleted wholly from the

investigation, this is not desirable, since risk factors will be really risky only when they fall into some risk ranges. Thus delete the whole variable(s) in this case seems not reasonable in this case, while a more reasonable way is to find the risk ranges of these variables, and delete the un-risky ranges. In some other studies, some of the covariates values may just random errors which do not contribute to the influence of the responses, and remove these covariates values will make the model interpretation more accurate. In this sense we select the variables when they fall within some range. To our knowledge, method for partial variable selection hasn't been seen in the literature, and our goal here is to explore such a method. In the existing method of deleting whole variable(s), the validity of such selection can be justified using the Wilks result, under the null hypothesis of no effect of the deleted variable(s), the resulting two times log-likelihood ratio will be asymptotically chi-squared distributed. We extended the Wilks theorem to the case for partial variable deletion, and use it to justify the partial deletion procedure. Simulation studies are conducted to evaluate the performance of the proposed method, and it is applied to analyze a real data set as illustration.

The proposed method

The observed data is $(y_i, x_i)(i = 1,...,n)$, where y_i is the response and $x_i \in R^d$ is the covariates of the i th subject. Denote

$y_n = (y_1,...,y_n)'$ and $X_n = (x_1',...,x_n')'$. Consider the linear model

$$y_n = X_n\beta + \eth_n, \quad (1)$$

Where $\beta = (\beta_1,...,\beta_d)'$ is the vector of regression parameter, $\eth_n = (\eth_1,...,\eth_n)'$ is the vector of random errors. Without loss of generality we consider the case the $\eth_i's$ are iid, i.e. $Var(\eth) = \sigma^2 I_n$, where I_n is the n dimensional identity matrix. When the $\eth_i's$ are not iid, often it is assumed $Var(\eth) = \Omega$ for some known positive-definite Ω, then make the transformation $\tilde{y}_n = \Omega^{-1/2}y_n$, $\tilde{X}_n = \Omega^{-1/2}X_n$ and $\eth = \grave{U}^{-1/2}\eth$ then we get the model $\tilde{y}_n = \tilde{X}_n\beta + \eth$ and the $\eth_i's$ are iid with $Var(\eth) = I_n$. When Ω s unknown, it can be estimated by various ways. So below we only need to discuss the case the $\eth_i's$ are iid.

We first give a brief review of the existing method of variable selection. Assume $\in = y - x'\beta$ has some known density $f(\cdot)$ (such as normal), with possibly some unknown parameter(s). For simple of discussion we assume there is no unknown parameters. Then the log-likelihood is

$$\ell_n(\beta) = \sum_{i=1}^{n}\log f(y_i - x_i'\beta).$$

Let $\hat{\beta}$ be the MLE of β (when $f(\cdot)$ is the standard normal density, $\hat{\beta}$ is just the least squares estimate). If we delete $k(\le d)$ columns of X_n and the corresponding components of β, denote the remaining covariate matrix as X_n^- and the resulting β as β^- and the corresponding MLE as $\hat{\beta}^-$. Then under the hypothesis H_0: the deleted columns of X_n has no effects, or equivalently the deleted components of β are all zeros, then asymptotically

$$2\left[\ell_n(\hat{\beta}) - \ell_n(\hat{\beta}^-)\right] \xrightarrow{D} \chi_k^2,$$

Where χ_k^2 is the chi-squared distribution with k degrees of freedom. Let $\chi_d^2(1-\alpha)$ be the $(1-\alpha)^{th}$ upper quantile of the χ_k^2 distribution, if $2[\ell_n(\hat{\beta}) - \ell_n(\hat{\beta}^-)] \ge \chi_k^2(1-\alpha)$, then H_0 is rejected at significance level α, and its' not good to delete these columns of X_n; otherwise we accept H_0 and delete these columns of X_n. There are some other methods to select columns of X_n, such as AIC, BIC and their variants, as in the model selection field. In These methods, the optimal deletion of columns of X_n corresponds to the best model selection, which maximize the AIC or BIC. These methods are not as solid as the above one, as may sometimes depending on eye inspection to choose the model which maximize the AIC or BIC.

All the above methods require the models under consideration be nested within each other, i.e., one is a sub-model of the other. Another more general model selection criterion is the minimum description length (MDL) criterion, a measure of complexity, developed by Kolmogorov [4], Wallace and Boulton, etc. The Kolmogorov complexity has close relationship with the entropy, it is the output of a Markov information source, normalized by the length of the output. It converges almost surely (as the length of the output goes to infinity) to the entropy of the

source. Let $\mathcal{G} = \{g(\cdot,\cdot)\}$ be a finite set of candidate models under consideration, and $\grave{E} = \{\theta_j : j = 1,...,h.\}$ be the set of parameters of interest. θ_i may or may not be nested within some other θ_j, or θ_i and θ_j both in $\grave{E}$ may have the same dimension but with different parametrization. Next consider a fixed density $f(\cdot|\theta_j)$, with parameter θ_j running through a subset $\grave{A}_j \subset R^k$ to emphasize the index of the parameter, we denote the MLE of θ_j under model $f(\cdot|\cdot)$ by $\hat{\theta}_j$ $\hat{\theta}_j$ (instead of by $\hat{\theta}_n$ to emphasize the dependence on the sample size), $I(\theta_j)$ the Fisher information for θ_j under $f(\cdot|\cdot)$ $|I(\theta_j)|$ its determinant, and k_j the dimension of θ_j. Then the MDL criterion [16] chooses θ_j to minimize

$$-\sum_{i=1}^{n}\log f(Y_i|\hat{\theta}_j) + \frac{k_j}{2}\log\frac{n}{2\pi} + \log\int_{\grave{A}_j}\sqrt{|I(\theta_j)|}d\theta_j, \quad (j=1,...,h).$$

(3)

This method does not require the models be nested, but still require select/delete some whole columns, and does not apply to our case.

Now come to our question, which is non-standard and we are not aware of a formal method to address this problem. However, we think the following question is of practical meaning. Consider deleting some of the components within fixed $k(k \le d)$ columns of X_n the deleted proportions for these columns are $\gamma_1,...,\gamma_k (0 < \gamma_j < 1)$. Denote X_n^- for the remaining covariate matrix, which is X_n with some entries replaced by 0's, corresponding to the deleted elements. Before the partial deletion, the model is $y_n = X_n\beta + \eth_n$.

After the partial deletion of covariates, the model becomes $y_n = X_n^-\beta^- + \eth_n$.

Note that here β and β^- have the same dimension, as no covariate is completely deleted. β is the effects of the original covariates, β^- is the effects of the covariates after some possible partial deletion. It is the effects of the effective covariates. Thus, though β and β^- have the same structure, they have different interpretation. The problem can be formulated as testing the hypothesis:

$$H_0 : \beta = \beta^- \quad vs \quad H_1 : \beta \ne \beta^-.$$

If H_0 is accepted, the partial deletion is valid.

Note that different from the standard null hypothesis that some components of the parameters be zeros, the above null hypothesis is not a nested hypothesis, or β^- is not a subset of β, so the existing Wilks' theorem for likelihood ratio statistic does not directly apply to our problem. Denote $\ell_n^-(\beta)$ be the corresponding log-likelihood based on data (y_n, X_n^-) and the corresponding MLE as $\hat{\beta}^-$. Since after the partial deletion, β^- is the MLE of under a constrained log-likelihood, while $\hat{\beta}$ is the MLE under the full likelihood, we have $\ell_n^-(\hat{\beta}^-) \le \ell_n(\hat{\beta})$. Parallel to the log-likelihood ratio statistic for (whole) variable deletion, let, for our case, $\ddot{E}_n = 2[\ell_n(\hat{\beta}) - \ell_n^-(\hat{\beta}^-)]$.

Let $(j_1,...,j_k)$ be the columns with partial deletions, $C_{j_r} = \{i : x_{j,i} \text{ is deleted } 1 \le i \le n\}$ be the index set for the deleted covariates in the j_r^{th} column be the cardinality of C_{j_r} thus

$ma_r = |C_{j_r}|/n (r = 1,...,k)$. We first give the following Proposition, in the simple case in which the index sets C_{j_r} are mutually exclusive. Then in Corollary 1 we give the result in more general case in which the index sets C_{j_r} are not need to be mutually exclusive. For given X_n, there are many different ways of partial column deletions, we may use Theorem 1 to test each of these deletions. Given a significance level α, a deletion is valid at level α if $\ddot{E}_n < \chi^2(1-\alpha)$, where $\chi^2(1-\alpha)$ is the $(1-\alpha)^{th}$ upper quantile of the $\Sigma_{j=1}^k \gamma_j \chi_j^2$ distribution, which can be computed by simulation for given $(\gamma_1,...,\gamma_k)$.

The following Theorem is a generalization of the Wilks [1] Theorem. Deleting some whole columns in X_n, corresponds to $\gamma_j = 1$ $(j = 1,...,k)$ in the theorem, and then we get the existing Wilks' Theorem.

Theorem 1

Under H_0, suppose $C_{j_r} \cap C_{j_s} = \phi$, the empty set, for all $1 \le r \ne s \le k$, then we have

$$\ddot{E}_n \xrightarrow{D} \sum_{j=1}^k \gamma_j \chi_j^2,$$

Where $\chi_1^2,...,\chi_k^2$ are iid chi-squared random variable with 1-degree of freedom. The case the C_{j_r} are not mutually exclusive is a bit more complicated. We first re-write the sets C_{j_r} such that

$$\cup_{r=1}^k C_{j_r} = \cup_{r=1}^k \cup_{j_1,...,j_r} D_{j_1,...,j_r},$$

where the $D_{j_1,...,j_r}$'s are mutually exclusive, $D_{j_1},...,D_{j_k}$ are index sets for one column of X_n only; the D_{j_1,j_2}'s are index sets common for columns j_1 and j_2 only; the D_{j_1,j_2,j_3}'s are index sets common for columns j_1, j_2 and j_3 only,.... Generally some of the $D_{j_1,...,j_r}$'s are empty sets. Let $\gamma_{j_1,...,j_r} = |D_{j_1,...,j_r}|$ be the cardinality of $D_{j_1,...,j_r}$ and $\gamma_{j_1,...,j_r} = |D_{j_1,...,j_r}|/n$ $(r = 1,...,k)$. By examining the proof of Theorem 1, we get the following corollary which gives the result in the more general case.

Corollary 1: Under H_0, we have

$$\ddot{E}_n = 2\left[\ell_n(\hat{\beta}) - \ell_n^-(\hat{\beta})\right] \xrightarrow{D} \sum_{r=1}^k \sum_{j_1,...,j_r} \gamma_{j_1,...,j_r} \chi_{j_1,...,j_r}^2,$$

Where the $\chi_{j_1,...,j_r}^2$'s are all independent chi-squared random variables with r-degrees of freedom $(r = 1,...,k)$.

Below we give two examples to illustrate the usage of Proposition.

Example 1: $n = 1000$, $d = 5$ $k = 3$. Columns (1,2,4) has some partial deletions with $C_1 = \{201, 202,...,299, 300\}$, $C_2 = \{351, 352,..., 549, 550\}$, $C_3 = \{601, 602,..., 849, 850\}$ the have no C_j's overlap; $\gamma_1 = 1/10$ $\gamma_2 = 1/5$, $\gamma_3 = 1/4$. o by the Proposition, under H_0 we have

$$2\left[\ell_n(\hat{\beta}) - \ell_n^-(\hat{\beta}^-)\right] \xrightarrow{D} \frac{1}{10}\chi_1^2 + \frac{1}{5}\chi_2^2 + \frac{1}{4}\chi_3^2$$

Where all the chi-squared random variables are independent, each has 1 degree of freedom.

Example 2: $n = 1000$, $d = 5$, $k = 3$. Columns (1,2,4) has some partial deletions with $C_1 = \{101, 102,...,299, 300; 651, 652,...,749, 750\}$, $C_2 = \{201, 202,..., 349, 350\}$, $C_3 = \{251, 252,..., 299, 300; 701, 702,..., 799, 800\}$. In this case the C_j's have overlaps, the Proposition can not be used directly, so we use the Corollary. Then $D_1 = \{101, 102,..., 199, 200\}$, $D_2 = \{301, 302,..., 349, 350\}$, $D_3 = \{701, 702,..., 799, 800\}$, $D_{1,2} = \{201, 202,..., 249, 250\}$, $D_{1,3} = \{701, 702,..., 749, 750\}$ $D_{2,3} = \phi$, $D_{1,2,3} = \{251, 252,..., 299, 300\}$, $\gamma_1 = 1/5$ $\gamma_2 = 1/20$, $\gamma_3 = 1/10$, $\gamma_{1,2} = 1/20$, $\gamma_{1,3} = 1/20$, $\gamma_{2,3} = 0$, $\gamma_{1,2,3} = 1/20$. So by the Corollary, under H_0 we have

$$2\left[\ell_n(\hat{\beta}) - \ell_n^-(\hat{\beta}^-)\right] \xrightarrow{D} \frac{1}{5}\chi_1^2 + \frac{1}{20}\chi_2^2 + \frac{1}{10}\chi_3^2 + \frac{1}{20}\chi_{1,2}^2 + \frac{1}{20}\chi_{1,3}^2 + \frac{1}{20}\chi_{1,2,3}^2$$

where all the chi-squared random variables are independent, with χ_1^2, χ_2^2 and χ_3^2 are each of 1 degree of freedom, $\chi_{1,2}^2$ and $\chi_{1,3}^2$ are each of 2-degrees of freedom, and $\chi_{1,2,3}^2$ is of 3-degrees of freedom. Next, we discuss the consistency of estimation of $\hat{\beta}^-$ under the null hypothesis H_0 Let $x^- = x_r^-$ with probability $\gamma_r (r = 0, 1,..., k)$, where x_r^- is an i.i.d. copy of the $x_{i,r}^-$'s whose components with index in C_{j_r}, in particular C_{j_0} is the index set for those covariates without partial deletion.

Theorem 2

Under conditions of Theorem 1,

$$\hat{\beta}^- \to \beta_0$$

$$\sqrt{n}\left(\hat{\beta}^- - \beta_0\right) \xrightarrow{D} N(0, \Omega), \to \beta_0$$

Where,

$$\Omega = E_{\beta_0}\left[\dot{\ell}(\beta_0)\dot{\ell}(\beta_0)\right] = E\left[(x^- - \mu^-)(x^- - \mu^-)'\right]\int\frac{\dot{f}^2(\dot{\theta})}{f(\dot{\theta})}d\dot{\theta}$$

To extend the results of Theorem 2 to the general case, we need the following more notations. Let $x_{(j_1,...,j_k)}$ be an i.i.d. copy of data in the set $D_{j_1,...,j_k}$. Let $x^- = x_{j_1,...,j_r}^-$ with probability $\gamma_{j_1,...,j_r} (r = 0, 1,..., k)$ where $x_{j_1,...,j_r}^-$ is an i.i.d. copy of the $x_{i,j_1,...,j_r}^-$'s, whose components with index in $C_{j_1,...,j_r}$,

Corollary 2: Under conditions of Corollary 1, results of Theorem 2 hold with J given above.

Computationally $E\left[(x^- - \mu^-)(x^- - \mu^-)'\right]$ is well approximated by

$$E\left[(x^- - \mu^-)(x^- - \mu^-)'\right] \approx \sum_{r=0}^k \frac{|D_{j_1,...,j_r}|}{n}\frac{1}{|D_{j_1,...,j_r}|}\sum_{(i,j)\in D_{j_1,...,j_r}}(x_{i,j} - \hat{\mu}_{j_1,...,j_r}^-)(x_{i,j} - \hat{\mu}_{j_1,...,j_r}^-)',$$

Where the notation $\dot{\Theta}_{(i,j)\in D_{j_1,...,j_r}}$ means summation over those x_{ij}^-'s with deletion index in $D_{j_1,...,j_r}$, and $(\hat{\mu}_{j_1,...,j_r}^-) = \frac{1}{|D_{j_1,...,j_r}|}\Sigma_{(i,j)\in D_{j_1,...,j_r}} x_{i,j}^-$.

Simulation study and application

Simulation study: We illustrate the proposed method with two examples, Example 3 and Example 4 below. The former rejects the null hypothesis H_0 while the latter accepts. In each case we simulate $n = 1000$ i.i.d. data with response y_i and with covariates $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5})$. We first generate the covariates, sample

the x_i's from the 5-dimensional normal distribution with mean vector $\mu = (3.1, 1.8, -0.5, 0.7, 1.5)'$ and a given covariance matrix $\tilde{A}$. Then we generate the response data, which, given the covariates. The are y_i's generated as $y_i = x_i' \beta_0 + \dot{o}_i$, $(i = 1, ..., n)$, $\beta_0 = (0.42, 0.11, 0.65, 0.83, 0.72)'$ the $\dot{o}_i$'s are i.i.d. $N(0,1)$. Hypothesis test is conducted to examine if the partial deletion is valid or not. Significant level is set as $\alpha = 0.05$. The experiment repeated 1000 times, prop represents the proportion $\ddot{E}_n > \chi^2(1-\alpha)$.

Example 3: In this example, we are interested to know whether covariates with $|x_{ij}| < \frac{1}{10}$ can be deleted. Five data set with different β_0 values are simulated. With $\gamma = (\gamma_1, ..., \gamma_k)$ the results are shown

in Table 1. We see that the proportion of rejecting H_0 $(\text{Pr}op)$ are all smaller than 0.05 in the five set of β_0. This suggests that covariates with $|x_{ij}| < \frac{1}{10}$ should not be deleted at 0.05 significance level. Example 4. In this example, the original X as in Example 3, but now we replace the entries in first 100 rows and first three columns by $\dot{O}$, where $\dot{o} N\left(0, \frac{1}{9}\right)$. We are interested to see in this case whether these noises can be deleted, i.e. H_0 can be rejected or not. The results are shown in the following. We see that the proportion of rejecting $\|$ $(\text{Pr}op)$ are all greater than 0.95 for the five sets of β_0. It suggests that the data provided strong evidence to conclude that the deleted value are noises and they are not necessary to the data set at 0.05 significance level.

Table 1: The simulation result of $\gamma \ddot{E}_n, \chi^2(-\alpha)$ and its $Prop$ according to β_0.

No.	$\hat{a}_0$	γ	$\ddot{E}_n$	$\chi^2(1-\alpha)$	$Prop$
1	(0.42,0.11,0.65,0.83,0.72)	(0.008,0.022,0.043,0.037,0.030)	14492.91	4.5767	0.006
2	(0.12,0.85,0.44,0.73,0.62)	(0.004,0.020,0.041,0.040,0.020)	13010.97	4.5748	0.016
3	(0.59,0.27,0.73,0.35,0.66)	(0.008,0.032,0.031,0.048,0.025)	13505.9	4.5786	0.000
4	(0.21,0.45,0.78,0.56,0.63)	(0.007,0.022,0.039,0.053,0.033)	12487.58	4.5281	0.005
5	(0.77,0.51,0.48,0.89,0.32)	(0.01,0.022,0.042,0.045,0.026)	15437.66	4.5317	0.000

Application to real data problem

We analyze a data set from the Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonnism, which is obtained from The National Institutes of Health (NIH)[17]. It is a multi-center, placebo-controlled clinical trial that aimed to determine a treatment for early Parkinson's disease patient to prolong their time requiring levodopa therapy. The number of patients enrolled was 800. The selected object were untreated patients with Parkinson's disease (stage I or II) for less than five years and met other eligible criteria. They were randomly assigned according to a two-by-two factorial design to one of four treatment groups:

Placebo

Active tocopherol

Active deprenyl

Active deprenyl and tocopherol.

The observation continued for 14±6 months and reevaluated every 3 months. At each visit, Unified Parkinson's Disease Rating Scale (UPDRS) including its motor, mental and activities of daily living components were evaluated. Statistical analysis result was based on 800 subjects. The result revealed that no beneficial effect of tocopherol. Deprenyl effect was found significantly prolong the time requiring levodopa therapy which reduced the

risk of disability by 50 percent according to the measurement of UPDRS. Our goal is to examine whether some of the covariates can be partially deleted. The response variables to examined are PDRS, TREMOR, S/E ADL by Rater, PIGD, Days from enrollment and Days from enrollment to Need for LEVODOPA. The covariates are Age, Motor and ADL for all these responses [18]. The deleted covariates are the ones with values below the lth data quantile, with $\gamma = 0.01, 0.02, 0.03$ and 0.05. We examine the responses one by one. The results are shown in Tables 2-5 below.

In Table 3, response TREMOR is examined. For covariable Age, the likelihood ratio $\ddot{E}_n$ is larger than the cutoff point $\chi^2(1-\alpha)$ at 0.03 and 0.05 levels, it suggests that for Age, partial deletions with these proportions are not valid. For covariable Motor, $\ddot{E}_n$ is smaller than the cutoff point $\chi^2(1-\alpha)$ at the 0.05 and 0.1 levels, this covariable can be partially deleted at these proportions. For covariable of ADL, with deletion proportions 0.01-0.1, the likelihood ratio $\ddot{E}_n$ is smaller than $\chi^2(1-\alpha)$ which suggest that the lower percentage of 1%-10% can be deleted. In Table 4, PIGD is the response variable. For covariable age, $\ddot{E}_n$ is larger than the cutoff point $Q(1-\alpha)$ at 0.01, 0.02, 0.03 and 0.05 level, suggests that it cannot be partially deleted with these proportions [19]. For covariable Motor, da_n is smaller than cutoff point $\chi^2(1-\alpha)$ at the deletion proportions of 0.02 and 0.03, suggests that the lower percentage of 2% and 3% can be deleted from the covariable Motor. For the variable ADL, $\ddot{E}_n$ is larger than the

cutoff point $\varrho(1-\alpha)$ at the delete proportion of 0.01, 0.02, 0.03 and 0.05, hence partial deletion is not valid.

In Tables 5, the response is PDRS. The likelihood ratios $\ddot{E}_n$ of Age, Motor and ADL all are larger than $\chi^2(1-\alpha)$ at the deletion proportions of 0.01, 0.02, 0.03 and 0.05. Thus the null hypothesis are rejected at all these proportions. Note that the coefficient for Age is insignificant, and hence the corresponding $\ddot{E}_n$ values with deleted proportions are senseless **Appendix.**

Table 2: The simulation result of $\gamma \ddot{E}_n \chi^2(1-\alpha)$ and its *Prop* according to β_0.

No.	$\hat{a}_0$	γ	$\ddot{E}_n$	$\chi^2(1-\alpha)$	*Prop*
1	(0.42,0.11,0.65,0.83,0.72)	(0.1,0.1,0.1)	1.0146	4.6034	0.998
2	(0.12,0.85,0.44,0.73,0.62)	(0.1,0.1,0.1)	0.3576	4.6414	0.977
3	(0.59,0.27,0.73,0.35,0.66)	(0.1,0.1,0.1)	3.2480	4.6756	0.965
4	(0.21,0.45,0.78,0.56,0.63)	(0.1,0.1,0.1)	3.3003	4.6306	0.972
5	(0.77,0.51,0.48,0.89,0.32)	(0.1,0.1,0.1)	3.3531	4.6326	0.955

Table 3: Response TREMOR: $\ddot{E}_n$ values and estimated regression coefficients.

Age	Motor	ADL	
Estimated coefficient	0.0240456	0.1801616	0.00451205
Delete Proportion	$\ddot{E}_n \chi^2(1-\alpha)$	$\ddot{E}_n \chi^2(1-\alpha)$	$\ddot{E}_n \chi^2(1-\alpha)$
0.01	11.5171 0.0593	0.35929 0.8787	0.00425 0.1897
0.03	20.0485 0.1245	6.2598 0.6924	0.00425 0.1861
0.05	14.0114 0.2937	8.7075 0.9034	0.00425 0.1496
0.1			0.0238 0.3841

Table 4: Response TREMOR: $\ddot{E}_n$ values and estimated regression coefficients.

Age	Motor	ADL	
Estimated coefficient	-0.0049032	0.02467423	0.2084862
Delete Proportion	$\ddot{E}_n \chi^2(1-\alpha)$	$\ddot{E}_n \chi^2(1-\alpha)$	$\ddot{E}_n \chi^2(1-\alpha)$
0.02	0.4956 0.0849	0.0031 0.0972	3.5607 0.2210
0.03	1.3908 0.1306	0.0166 0.1513	3.5607 0.2256
0.05	0.4816 0.2596	1.2607 0.2188	3.6607 0.1925

**Table 5:** Response TREMOR: $\ddot{E}_n$ values and estimated regression coefficients.

Age	Motor	ADL	
Estimated coefficient	1.563389	0.0476914	-1.139864
Delete Proportion	$\ddot{E}_n \chi^2(1-\alpha)$	$\ddot{E}_n \chi^2(1-\alpha)$	$\ddot{E}_n \chi^2(1-\alpha)$
0.01	1.0073 0.0497	80.7411 0.0944	6.3392 0.0453
0.02	1.0475 0.0669	142.3528 0.0841	57.5051 0.2216
0.03	0.8609 0.1486	321.5332 0.1906	57.5051 0.2111
0.05	0.8650 0.2379	397.6481 0.2227	57.5051 0.2199

Concluding remarks

We proposed a method for partial variable deletion, which is a generalization of the existing variable selection. The question is motivated from practical problems. It can used to find the effective ranges of the covariates, or to remove possible noises in the covariates, and thus the corresponding estimated effects are more interpretable. The procedure is a generalization of the Wilks likelihood ratio statistic, and is simple to use. Simulation studies are conducted to evaluate the performance of the method, and it is applied to analyze a real Parkinson disease data as illustration.

References

1. Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. Annals of Mathematical Statistics 9: 6062.

2. Akaike H (1974) A new look at the statistical identification model. IEEE Transaction on Automatic Control 19: 716-723.

3. Schwarz G (1978) Estimating the dimension of a model. Annals of Statistics 6: 461-464.

4. Kolmogorov A (1963) On tables of random numbers. Sankhya Ser A 25: 369-375.

5. Hansen M, Yu B (2001) Model selection and the principle of minimum description length. Journal of American Statistical Association 96: 746-774.

6. May RJ, Maier HR, Dandy GC, Fernando TG (2008) Non-linear variable selection for artificial neural networks using partial mutual information. Environmental Modelling and Software 23: 1312-1326.

7. Mehmood T, Liland KH, Snipen L, Saeb S (2012) A review of variable selection methods in partial least squares regression. Chemometrics and Intelligent Laboratory Systems 118: 62-69.

8. Wang L, Liu X, Linag H, Carroll R (2011) Estimation and variable selection for generalized additive partial linear models. Annals of Statistics 39: 1827-1851.

9. Liu X, Wang L, Liang H (2011) Estimation and variable selection for semiparametric additive partial linear models. Stat Sin 21(3): 1225-1248.

10. Tibshirani R (1996) Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B 58: 267-288.

11. Tibshirani R (1997) The lasso method for variable selection in the Cox model. Stat Med 16(4): 385-395.

12. Fan J, Li R (2001) Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96: 1348-1360.

13. Fan J, Li R (2002) Variable selection for Cox's proportional hazards model and frailty model. Annals of Statistics 96: 1348-1360.

14. Wang HX, Leineweber C, Kirkeeide R, Svane B, Theorell T, et al. (2007) Psychosocial stress and atherosclerosis: family and work stress accelerate progression of coronary disease in women. The Stockholm Female Coronary Angiography Study J Intern Med (3): 245-254.

15. Shara NM, Wang H, Valaitis E, Pehlivanova M, Carter EA, et al. (2011) Comparison of estimated glomerular filtration rates and albuminuria in predicting risk of coronary heart disease in a population with high prevalence of diabetes mellitus and renal disease. Am J Cardiol 107(3): 399-405.

16. Rissanen J (1996) Fisher information and stochastic complexity. IEEE Transactions on Information Theory 42: 40-47.

17. Bickel PJ, Klaassen CA, Ritov Y, Wellner JA (1993) Efficient and Adaptive Estimation for Semiparametric Models, Johns Hopkins University Press, Baltimore, Maryland, USA.

18. Shoulson I (1989) Deprenyl and tocopherol antioxidative therapy of parkinsonism (DATATOP). Parkinson Study Group. Acta Neurol Scand Suppl 126: 171-175.

19. http://www2.math.umd.edu/slud/s701/WilksThm.pdf
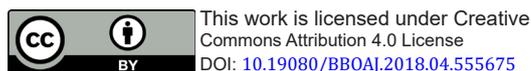
Your next submission with Juniper Publishers

will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php