Sure Explained Variability and Independence Screening

Zhengjun Zhang

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Co-authors: Min Chen, Yimin Lian, Zhao Chen

Outline

Sparsity tuning I

Variable selection

- the least absolute shrinkage and selection operator (LASSO) in Tibshirani (1996), group LASSO in Yuan and Lin (2006), and the adaptive LASSO in Zou (2006);
- the smoothly clipped absolute deviation (SCAD) in Fan (1997), Fan and Li (2001);
- the elastic net in Zou and Hasties (2005);
- the Dantzig selector in Candes and Tao (2007);
- ....

Sparsity tuning II

Independence screening

- Sure independence screening (SIS) based on Pearson's correlation: Fan and Lv (2008), Liu *et al.* (2013);
- Model based screening: Fan and Song (2010), Fan *et al.* (2011), Zhu *et al.* (2011), Song *et al.* (2012), Zhang *et al.* (2016);
- Model free screening: DC-SIS by Li *et al.* (2012), QaSIS by He *et al.* (2013), Q-SIS by Wu and Yin (2015);
- Our approach: based on explained variability and model free.

Outline

A new look of a twice-told tale:

$$\mathrm{var}(Y) = \mathrm{var}(\mathrm{E}(Y|X)) + \mathrm{E}(\mathrm{var}(Y|X)). \tag{2.1}$$

- $\mathrm{Var}\big(E(Y|X)\big)$ measures the spread of the conditional mean (center) of Y given X
- $\mathrm{Var}\big(E(Y|X)\big)/\mathrm{Var}(Y)$ can certainly be interpreted as the explained variance of Y by X.

$$\mathrm{SEV}(Y|X) = \frac{\mathrm{var}(\mathrm{E}(Y|X))}{\mathrm{var}(Y)} = 1 - \frac{\mathrm{E}(\mathrm{var}(Y|X))}{\mathrm{var}(Y)} = 1 - \frac{\mathrm{E}[\{Y - \mathrm{E}(Y|X)\}^2]}{\mathrm{var}(Y)}$$

Correlation ratios: Kendall and Stuart (1979), Doksum and Samarov (1995), Wang (2001), Zheng, Shi, and Zhang (2012)

Properties of SEV

Suppose both $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then

- (P.1) If $Y = g(X)$, *almost surely* (*a.s.*), then $\text{SEV}(Y|X) = 1$.
- (P.2) If $Y = g(X) + \varepsilon$, then $\text{SEV}(Y|X) = \text{var}(g(X))/(\text{var}(g(X)) + \text{var}(\varepsilon))$.
- (P.3) If $Y = ag(X) + b + \varepsilon$, where $g(x) = x$, $a \neq 0$, then $\text{SEV}(Y|X) = \rho_{XY}^2$. Here ρ_{XY} is Pearson's correlation coefficient.
- (P.4) If $\rho_{XY} \neq 0, \text{SEV}(Y|X) \neq 0$.
- (P.5) If $\rho_{XY} = -1$ or $1, \text{SEV}(Y|X) = 1$.
- (P.6) $\text{SEV}(Y|X) \geq \rho_{XY}^2$.
- (P.7) If $\text{SEV}(Y|X) = 0, \rho_{XY} = 0$.

Example 1

Suppose X is a standard normal random variable. Let $Y = X^2$. Then SEV($Y|X$)=1, while $\rho_{XY} = 0$, and the distance correlation between Y and X is approximately 0.54 by a numerical calculation.

Outline

Notations

- $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$, $\mathbf{X}_k = (X_{1k}, ...X_{nk})^T$
- n is the sample size, the dimension p is much larger than the sample size n, i.e. $n \ll p$.
- $\sum_{i=1}^{n} X_{ik} = 0$ and $\frac{1}{n-1} \sum_{i=1}^{n} X_{ik}^2 = 1$ for $k = 1, \ldots, p$.

SEVIS procedure

- The index set of active predictors:

$$\mathscr{M} = \{1 \le k \le p : \text{ the predictor } X_k \text{ has contributed to the response Y} \tag{3.1}$$

- Denoted by $\omega_k = \text{SEV}(Y|X_k), k = 1, \ldots, p$. We define a new index set by

$$\mathscr{M}^* = \{1 \le k \le p : \ \omega_k > 0\}. \tag{3.2}$$

- Denoting $\widehat{\omega}_k$ as the estimator of ω_k, we define a truncated active index set by

$$\widehat{\mathscr{M}^*} = \{1 \le k \le p : \ \widehat{\omega}_k \ge cn^{-\tau}\}. \tag{3.3}$$

- A practically workable active index set:

$$\widehat{\mathscr{M}^*_d} = \{1 \le k \le p : \ \widehat{\omega}_k \text{ is among the first } d \text{ largest estimates in } \widehat{\mathscr{M}^*}\}. \tag{3.4}$$

SEVIS algorithm

- 1. Normalize $X_k, k = 1, \ldots, p$, such that $\sum_{i=1}^n X_{ik} = 0$ and $(n-1)^{-1} \sum_{i=1}^n X_{ik}^2 = 1$;
- 2. Calculate $\widehat{\omega}_k = \text{SEV}(Y|X_k)$ between Y and candidate predictors $X_k, \; k = 1, \ldots, p$;
- 3. Rank $\widehat{\omega}_k$ in a decreasing order, that is $\widehat{\omega}_{k_1} > \widehat{\omega}_{k_2} > \cdots > \widehat{\omega}_{k_p}$;
- 4. Choose the first $d[n/\log(n)]$, $X_{k_1}, \ldots, X_{k_{d[n/\log(n)]}}$, as the active predictors. d is a given integer.

Assumptions

- **(A1)** $f(x,y)$ has a bounded support on $(x,y) \in [s_x, S^x] \times [s_y, S^y]$. $f^X(x)$ is strictly positive, and is uniformly continuous on $x \in [s_x, S^x]$. The second derivative $(f^X(x))''$ and $(\phi^{Y|X}(x))''$ exist and are uniformly bounded on $x \in [s_x, S^x]$.
- **(A2)** The symmetric kernel function $K(z)$ has a bounded support, and satisfies

  $$\sup_{-\infty<z<\infty} |K(z)| < \infty, \quad \lim_{z\to\infty} |zK(z)| = 0, \quad \int z^2 |K(z)|\, dz < \infty.$$

  The bandwidth h satisfies $nh^3 \to \infty$ and $nh^4 \to 0$, as $n \to \infty$.
- **(A3)** $\min_{k\in\mathscr{M}^*} \omega_k \geq 2cn^{-\tau}$, where constant $c > 0$ and $0 \leq \tau < \frac{1}{6}$.
- **(A4)** For given constants $c > 0$, $0 \leq \tau < \frac{1}{6}$, and $\log(p) = o(n^{1/3-2\tau})$, $\liminf_{p\to\infty}(\min_{k\in\mathscr{M}^*} \omega_k - \max_{k\notin\mathscr{M}^*} \omega_k) > 2cn^{-\tau}$.

Sure Screening Property

Under Assumptions (A1),(A2), it follows that

$$\mathrm{P}\left(\max_{1\le k\le p}|\widehat{\omega}_k-\omega_k|\ge cn^{-\tau}\right)\le O(p\exp\{-c_1n^{1-2\tau}h^2\}). \tag{3.5}$$

Furthermore, if Assumption (A3) holds, we have that

$$\mathrm{P}\left(\mathscr{M}^*\subseteq\widehat{\mathscr{M}^*}\right)\ge 1-O(s_n\exp\{-c_1n^{1-2\tau}h^2\}), \tag{3.6}$$

where s_n is the cardinality of $\mathscr{M}^*$.

Ranking Consistency property

Under Assumptions (A1), (A2), and (A4), we have that

$$\liminf_{n\to\infty}\{\min_{k\in\mathcal{M}^*}\widehat{\omega}_k - \max_{k\notin\mathcal{M}^*}\widehat{\omega}_k\} \geq 0, \quad \text{a.s.} \tag{3.7}$$

Outline

Settings and Example 1

- Existing methods to be compared: SIRS by Zhu et al. (2011), DC-SIS by Li et al. (2012) and Q-SIS with $\alpha = 0.75$ by Wu and Yin (2015), i.e. Model free
- $d_1 = [n/\log(n)]$, $d_2 = 2[n/\log(n)]$, and $d_3 = 3[n/\log(n)]$

$$
\begin{aligned}
&\textbf{(1.a)} \quad Y = c_1X_1 + c_1X_2 + c_1X_{12} + c_1X_{22} + \varepsilon,\\
&\textbf{(1.b)} \quad Y = c_1X_1 + c_2X_2 + c_3\,\mathbf{I}(X_{12} < 0) + c_4X_{22} + \varepsilon,\\
&\textbf{(1.c)} \quad Y = c_1X_1X_2 + c_3\,\mathbf{I}(X_{12} < 0) + c_4X_{22} + \varepsilon,\\
&\textbf{(1.d)} \quad Y = c_1X_1 + c_2X_2 + c_3\,\mathbf{I}(X_{12} < 0) + c_5\exp(|X_{22}|)\varepsilon.
\end{aligned}
$$

The predictors $X = (X_1, ..., X_p)$ are generated from a normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$, where $\sigma_{ij} = \sigma^{|i-j|}$. The random error ε follows a standard normal distribution. We set the vector $(c_1, c_2, c_3, c_4, c_5) = (5, 2, 7, 5, 2)$. The sample size $n = 200$, the dimension $p = 2000, 5000$, and $\sigma = 0.5, 0.8$.

Table 1 : The Proportions of $\mathscr{P}_s$ and $\mathscr{P}_a$, $p = 2000$ $\sigma = 0.5$

		SIRS	DC-SIS	Q-SIS	SEVIS				
		$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$				$\mathscr{P}_a$
Model	Size	*ALL*	*ALL*	*ALL*	X_1	X_2	X_{12}	X_{22}	*ALL*
(1.a)	d_1	**1.00**	**1.00**	**0.98**	1.00	1.00	1.00	1.00	**1.00**
	d_2	**1.00**	**1.00**	**0.99**	1.00	1.00	1.00	1.00	**1.00**
	d_3	**1.00**	**1.00**	**1.00**	1.00	1.00	1.00	1.00	**1.00**
(1.b)	d_1	**0.96**	**0.99**	**0.94**	1.00	1.00	0.98	1.00	**0.98**
	d_2	**0.98**	**1.00**	**0.96**	1.00	1.00	0.99	1.00	**0.99**
	d_3	**0.99**	**1.00**	**0.97**	1.00	1.00	0.99	1.00	**0.99**
(1.c)	d_1	**0.01**	**0.76**	**0.25**	0.98	0.99	0.99	1.00	**0.97**
	d_2	**0.01**	**0.90**	**0.47**	0.99	0.99	1.00	1.00	**0.98**
	d_3	**0.03**	**0.94**	**0.60**	0.99	0.99	1.00	1.00	**0.99**
(1.d)	d_1	**0.03**	**0.14**	**0.03**	1.00	1.00	0.82	0.56	**0.41**
	d_2	**0.06**	**0.30**	**0.06**	1.00	1.00	0.92	0.64	**0.57**
	d_3	**0.08**	**0.43**	**0.08**	1.00	1.00	0.94	0.68	**0.62**

Table 2 : The Proportions of $\mathscr{P}_s$ and $\mathscr{P}_a$, $p = 2000 \quad \sigma = 0.8$

		SIRS	DC-SIS	Q-SIS	SEVIS				
		$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$				$\mathscr{P}_a$
Model	Size	*ALL*	*ALL*	*ALL*	X_1	X_2	X_{12}	X_{22}	*ALL*
(1.a)	d_1	**1.00**	**1.00**	**0.97**	1.00	1.00	1.00	1.00	**1.00**
	d_2	**1.00**	**1.00**	**1.00**	1.00	1.00	1.00	1.00	**1.00**
	d_3	**1.00**	**1.00**	**1.00**	1.00	1.00	1.00	1.00	**1.00**
(1.b)	d_1	**0.45**	**0.74**	**0.46**	1.00	1.00	0.64	1.00	**0.64**
	d_2	**0.60**	**0.85**	**0.61**	1.00	1.00	0.76	1.00	**0.76**
	d_3	**0.67**	**0.89**	**0.67**	1.00	1.00	0.83	1.00	**0.83**
(1.c)	d_1	**0.02**	**0.99**	**0.65**	1.00	1.00	0.88	1.00	**0.88**
	d_2	**0.04**	**1.00**	**0.80**	1.00	1.00	0.95	1.00	**0.95**
	d_3	**0.08**	**1.00**	**0.88**	1.00	1.00	0.97	1.00	**0.97**
(1.d)	d_1	**0.05**	**0.13**	**0.02**	1.00	0.99	0.62	0.53	**0.28**
	d_2	**0.10**	**0.27**	**0.07**	1.00	1.00	0.75	0.61	**0.42**
	d_3	**0.15**	**0.40**	**0.12**	1.00	1.00	0.82	0.68	**0.53**

Table 3 : The Proportions of $\mathscr{P}_s$ and $\mathscr{P}_a$, $p = 5000$ $\sigma = 0.5$

		SIRS	DC-SIS	Q-SIS	SEVIS				
		$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$				$\mathscr{P}_a$
Model	Size	*ALL*	*ALL*	*ALL*	X_1	X_2	X_{12}	X_{22}	*ALL*
(1.a)	d_1	**1.00**	**1.00**	**0.93**	1.00	1.00	1.00	1.00	**1.00**
	d_2	**1.00**	**1.00**	**0.96**	1.00	1.00	1.00	1.00	**1.00**
	d_3	**1.00**	**1.00**	**0.98**	1.00	1.00	1.00	1.00	**1.00**
(1.b)	d_1	**0.98**	**0.99**	**0.92**	1.00	1.00	0.97	1.00	**0.97**
	d_2	**0.99**	**1.00**	**0.95**	1.00	1.00	0.99	1.00	**0.99**
	d_3	**1.00**	**1.00**	**0.96**	1.00	1.00	0.99	1.00	**0.99**
(1.c)	d_1	**0.00**	**0.50**	**0.08**	0.98	0.98	0.96	1.00	**0.92**
	d_2	**0.01**	**0.68**	**0.19**	0.98	0.99	0.99	1.00	**0.97**
	d_3	**0.01**	**0.77**	**0.28**	0.99	0.99	0.99	1.00	**0.98**
(1.d)	d_1	**0.01**	**0.08**	**0.01**	0.98	0.95	0.71	0.48	**0.29**
	d_2	**0.02**	**0.18**	**0.02**	1.00	0.98	0.83	0.56	**0.43**
	d_3	**0.04**	**0.24**	**0.03**	1.00	0.98	0.88	0.60	**0.50**

Table 4 : The Proportions of $\mathscr{P}_s$ and $\mathscr{P}_a$, $p = 5000$ $\sigma = 0.8$

		SIRS	DC-SIS	Q-SIS	SEVIS				
		$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$				$\mathscr{P}_a$
Model	Size	*ALL*	*ALL*	*ALL*	X_1	X_2	X_{12}	X_{22}	*ALL*
(1.a)	d_1	**1.00**	**1.00**	**0.96**	1.00	1.00	1.00	1.00	**1.00**
	d_2	**1.00**	**1.00**	**0.98**	1.00	1.00	1.00	1.00	**1.00**
	d_3	**1.00**	**1.00**	**0.99**	1.00	1.00	1.00	1.00	**1.00**
(1.b)	d_1	**0.39**	**0.66**	**0.37**	1.00	1.00	0.57	1.00	**0.57**
	d_2	**0.53**	**0.76**	**0.47**	1.00	1.00	0.69	1.00	**0.69**
	d_3	**0.59**	**0.82**	**0.54**	1.00	1.00	0.76	1.00	**0.76**
(1.c)	d_1	**0.01**	**0.98**	**0.50**	1.00	1.00	0.73	1.00	**0.73**
	d_2	**0.02**	**0.99**	**0.67**	1.00	1.00	0.83	1.00	**0.83**
	d_3	**0.03**	**0.99**	**0.76**	1.00	1.00	0.90	1.00	**0.90**
(1.d)	d_1	**0.03**	**0.08**	**0.00**	0.98	0.97	0.46	0.44	**0.14**
	d_2	**0.06**	**0.15**	**0.02**	0.99	0.99	0.62	0.55	**0.30**
	d_3	**0.08**	**0.19**	**0.04**	1.00	1.00	0.69	0.60	**0.37**

Example 2

$$\textbf{(2.a)} \quad Y = 5X_1 + 3X_2^2 + X_3^3 + \varepsilon,$$
$$\textbf{(2.b)} \quad Y = 3\sin(X_1) + 7\cos(X_2) + \tan(X_3) + \varepsilon.$$

$X = (X_1, ..., X_p)$ from a normal distribution with mean zero and covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$, where $\sigma_{ii} = 1, i = 1, ..., p$ and $\sigma_{ij} = \sigma, i \neq j$. Model (2.b) is a summation of trigonometric functions of predictors. Considering the property of tangent function, we generate $X = (X_1, ..., X_p)$ from a uniform distribution $[-1, 1]$ and keep the covariance matrix unchanged. The random error ε follows a standard normal.

				SIRS	DC-SIS	Q-SIS	SEVIS			
				$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$			$\mathscr{P}_a$ (2.a)
n	p	σ	d	*ALL*	*ALL*	*ALL*	X_1	X_2	X_3	*ALL*
200	1000	0	d_1	**0.07**	**1.00**	**0.70**	1.00	1.00	1.00	**1.00**
200	1000	0	d_2	**0.13**	**1.00**	**0.83**	1.00	1.00	1.00	**1.00**
200	1000	0	d_3	**0.18**	**1.00**	**0.87**	1.00	1.00	1.00	**1.00**
100	1000	0	d_1	**0.04**	**0.63**	**0.15**	1.00	0.96	0.90	**0.86**
100	1000	0	d_2	**0.08**	**0.80**	**0.28**	1.00	0.98	0.93	**0.90**
100	1000	0	d_3	**0.11**	**0.85**	**0.39**	1.00	0.99	0.95	**0.93**
200	5000	0	d_1	**0.02**	**0.98**	**0.38**	1.00	1.00	0.99	**0.99**
200	5000	0	d_2	**0.03**	**0.99**	**0.54**	1.00	1.00	1.00	**1.00**
200	5000	0	d_3	**0.04**	**1.00**	**0.65**	1.00	1.00	1.00	**1.00**
100	5000	0	d_1	**0.01**	**0.32**	**0.02**	0.97	0.89	0.77	**0.65**
100	5000	0	d_2	**0.02**	**0.48**	**0.04**	0.99	0.93	0.83	**0.75**
100	5000	0	d_3	**0.02**	**0.59**	**0.08**	1.00	0.95	0.86	**0.81**
200	1000	0.5	d_1	**0.01**	**0.32**	**0.22**	1.00	0.91	0.98	**0.89**
200	1000	0.5	d_2	**0.01**	**0.47**	**0.34**	1.00	0.95	0.98	**0.93**
200	1000	0.5	d_3	**0.01**	**0.55**	**0.40**	1.00	0.96	0.99	**0.95**
100	5000	0.5	d_1	**0.00**	**0.11**	**0.03**	0.96	0.65	0.78	**0.44**
100	5000	0.5	d_2	**0.01**	**0.19**	**0.07**	0.97	0.72	0.86	**0.57**
100	5000	0.5	d_3	**0.02**	**0.26**	**0.12**	0.97	0.76	0.89	**0.64**
200	5000	0.5	d_1	**0.00**	**0.13**	**0.05**	0.99	0.79	0.94	**0.73**
200	5000	0.5	d_2	**0.00**	**0.21**	**0.09**	0.99	0.85	0.97	**0.81**
200	5000	0.5	d_3	**0.01**	**0.24**	**0.13**	1.00	0.89	0.98	**0.86**
100	5000	0.5	d_1	**0.00**	**0.01**	**0.01**	0.90	0.44	0.58	**0.19**
100	5000	0.5	d_2	**0.00**	**0.04**	**0.01**	0.92	0.52	0.65	**0.26**
100	5000	0.5	d_3	**0.00**	**0.05**	**0.03**	0.93	0.57	0.70	**0.32**

				SIRS	DC-SIS	Q-SIS	SEVIS			
				$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$			$\mathscr{P}_a$ (2.b)
n	p	σ	d	*ALL*	*ALL*	*ALL*	X_1	X_2	X_3	*ALL*
200	1000	0	d_1	**0.03**	**0.97**	**0.52**	1.00	1.00	1.00	**1.00**
200	1000	0	d_2	**0.07**	**0.99**	**0.73**	1.00	1.00	1.00	**1.00**
200	1000	0	d_3	**0.10**	**1.00**	**0.84**	1.00	1.00	1.00	**1.00**
100	1000	0	d_1	**0.01**	**0.27**	**0.03**	1.00	0.94	0.82	**0.77**
100	1000	0	d_2	**0.02**	**0.50**	**0.10**	1.00	0.97	0.89	**0.86**
100	1000	0	d_3	**0.04**	**0.64**	**0.20**	1.00	0.98	0.92	**0.90**
200	5000	0	d_1	**0.00**	**0.67**	**0.08**	1.00	1.00	0.95	**0.95**
200	5000	0	d_2	**0.00**	**0.82**	**0.20**	1.00	1.00	0.98	**0.98**
200	5000	0	d_3	**0.01**	**0.88**	**0.35**	1.00	1.00	0.99	**0.99**
100	5000	0	d_1	**0.00**	**0.05**	**0.00**	1.00	0.84	0.63	**0.51**
100	5000	0	d_2	**0.00**	**0.12**	**0.00**	1.00	0.91	0.72	**0.64**
100	5000	0	d_3	**0.00**	**0.19**	**0.02**	1.00	0.93	0.77	**0.71**
200	1000	0.5	d_1	**0.00**	**0.27**	**0.00**	0.98	0.93	0.98	**0.92**
200	1000	0.5	d_2	**0.02**	**0.39**	**0.00**	0.98	0.95	0.99	**0.95**
200	1000	0.5	d_3	**0.05**	**0.52**	**0.00**	0.98	0.96	0.99	**0.96**
100	1000	0.5	d_1	**0.00**	**0.06**	**0.00**	0.97	0.53	0.77	**0.41**
100	1000	0.5	d_2	**0.02**	**0.11**	**0.00**	0.98	0.64	0.85	**0.53**
100	1000	0.5	d_3	**0.03**	**0.16**	**0.00**	0.98	0.72	0.90	**0.64**
200	5000	0.5	d_1	**0.00**	**0.09**	**0.00**	0.98	0.81	0.94	**0.77**
200	5000	0.5	d_2	**0.00**	**0.14**	**0.00**	0.98	0.87	0.96	**0.85**
200	5000	0.5	d_3	**0.00**	**0.17**	**0.00**	0.98	0.90	0.97	**0.88**
100	5000	0.5	d_1	**0.00**	**0.01**	**0.00**	0.97	0.36	0.59	**0.20**
100	5000	0.5	d_2	**0.00**	**0.02**	**0.00**	0.98	0.44	0.68	**0.29**
100	5000	0.5	d_3	**0.00**	**0.03**	**0.00**	0.98	0.50	0.72	**0.35**

Example 3

The underlying model is adapted from Ravikumar *et al.* (2008). We generate the data from the additive model

$$\textbf{(3)} \quad Y = f_1(X_1) + f_2(X_2) + f_3(X_3) + f_4(X_4) + \varepsilon,$$

where

$$f_1(x) = -2sin(2x),\ f_2(x) = x^2 - 1/3,\ f_3(x) = x - 1/2,\ f_4(x) = e^{-x} + e^{-1} - 1.$$

$X_1, \ldots, X_p$ are drawn from an independent and identically distributed uniform distribution on $[-2, 2]$. The random error ε follows the standard normal distribution. Let $n = 150$ and $p = 2000$ which is ten times of that in *et al.* (2008). The censored observations $Y_i^* = \min(Y_i, C_i), 1 \leq i \leq n$, where C_i follows a uniform distribution on $[0, 8.5]$ to control the censoring rate to be approximately 25%. The replications are 500.

Table 5 : The Proportions of $\mathscr{P}_s$ and $\mathscr{P}_a$ in Example 3

	SIRS	DC-SIS	Q-SIS	SEVIS				
	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_a$	$\mathscr{P}_s$				$\mathscr{P}_a$
d	*ALL*	*ALL*	*ALL*	X_1	X_2	X_3	X_4	*ALL*
d_1	**0.00**	**0.16**	**0.01**	0.93	0.90	0.97	1.00	**0.81**
d_2	**0.00**	**0.33**	**0.02**	0.97	0.94	0.98	1.00	**0.89**
d_3	**0.00**	**0.46**	**0.05**	0.98	0.96	0.99	1.00	**0.92**

Outline

Data

- 593 patients and their 12042 gene expression levels
- Divide all patients into 17 groups by tissue source site (TSS)
- 8 groups: TCGA-04, TCGA-09, TCGA-13, TCGA-23, TCGA-24, TCGA-25, TCGA-29, TCGA-61
- At least 30 patients ($n \geq 30$), total 487 samples
- Response variable 1: TP53, a well known tumor suppressor gene and mutates in 303 among 316 ovarian carcinoma patients
- Response variable 2: survival time of patients

Evaluate the performance of those top ranked genes

- consider the following nonparametric additive model,

$$Y = f_1(X_{first}) + f_2(X_{second}) + f_3(X_{third}) + \varepsilon, \tag{5.1}$$

- compare the adjusted R^2 and deviance explained by each method
- use SEV$(Y|\widehat{Y})$ to measure nonlinearly explained variability, where $\widehat{Y} = \widehat{f_1}(X_{first}) + \widehat{f_2}(X_{second}) + \widehat{f_3}(X_{third})$

Table 6 : The results of data fitting for $Y = $ TP53

Group	n	SIRS adj R^2	SIRS deviance explained	SIRS SEV$(Y\|\hat{Y})$	DC-SIS adj R^2	DC-SIS deviance explained	DC-SIS SEV$(Y\|\hat{Y})$
TCGA-04	43	0.405	0.467	0.410	0.578	0.641	0.562
TCGA-09	30	0.451	0.529	0.435	0.629	0.697	0.577
TCGA-13	113	0.294	0.342	0.311	0.319	0.373	0.335
TCGA-23	38	0.450	0.495	0.419	0.534	0.635	0.556
TCGA-24	100	0.296	0.361	0.326	**0.311**	**0.373**	**0.337**
TCGA-25	45	0.310	0.359	0.315	0.704	0.773	0.652
TCGA-29	52	0.556	0.583	0.525	0.539	0.566	0.484
TCGA-61	66	**0.494**	**0.558**	**0.489**	0.440	0.505	0.455

Table 7 : The results of data fitting for $Y = \text{TP53}$

Group	n	Q-SIS			SEVIS		
		adj R^2	deviance explained	SEV$(Y\|\widehat{Y})$	adj R^2	deviance explained	SEV$(Y\|\widehat{Y})$
TCGA-04	43	0.303	0.447	0.404	**0.861**	**0.912**	**0.788**
TCGA-09	30	0.200	0.400	0.341	**0.677**	**0.774**	**0.654**
TCGA-13	113	0.188	0.249	0.243	**0.348**	**0.421**	**0.382**
TCGA-23	38	0.538	0.681	0.591	**0.625**	**0.717**	**0.684**
TCGA-24	100	0.181	0.250	0.237	0.230	0.294	0.278
TCGA-25	45	0.562	0.716	0.613	**0.906**	**0.956**	**0.815**
TCGA-29	52	0.530	0.585	0.505	**0.620**	**0.643**	**0.545**
TCGA-61	66	0.151	0.246	0.272	0.267	0.365	0.380

Table 8 : The results of data fitting for Y being days to last followup

Group	n	SIRS adj R^2	SIRS deviance explained	SIRS SEV$(Y\|\hat{Y})$	DC-SIS adj R^2	DC-SIS deviance explained	DC-SIS SEV$(Y\|\hat{Y})$
TCGA-04	38	0.822	0.894	0.777	**0.884**	**0.940**	**0.799**
TCGA-09	30	0.527	0.577	0.507	0.720	0.778	0.646
TCGA-13	109	0.274	0.294	0.265	0.287	0.361	0.334
TCGA-23	38	0.660	0.725	0.639	0.662	0.720	0.657
TCGA-24	98	0.321	0.364	0.371	0.258	0.293	0.273
TCGA-25	45	0.456	0.493	0.456	**0.573**	**0.660**	**0.589**
TCGA-29	49	0.401	0.439	0.389	**0.503**	**0.554**	0.483
TCGA-61	64	0.224	0.261	0.241	0.302	0.336	0.311

Table 9 : The results of data fitting for Y being days to last followup

Group	n	Q-SIS			SEVIS		
		adj R^2	deviance explained	SEV$(Y\|\hat{Y})$	adj R^2	deviance explained	SEV$(Y\|\hat{Y})$
TCGA-04	38	0.655	0.792	0.666	0.710	0.794	0.740
TCGA-09	30	0.533	0.641	0.597	**0.725**	**0.800**	**0.728**
TCGA-13	109	0.257	0.301	0.278	**0.355**	**0.403**	**0.431**
TCGA-23	38	0.476	0.531	0.523	**0.691**	**0.752**	**0.690**
TCGA-24	98	0.289	0.361	0.382	**0.330**	**0.400**	**0.423**
TCGA-25	45	0.317	0.364	0.329	0.528	0.622	0.547
TCGA-29	49	0.320	0.414	0.364	0.398	0.495	**0.514**
TCGA-61	64	0.253	0.345	0.342	**0.332**	**0.416**	**0.429**

the mutual pairwise explained variabilities

- the explained variabilities of more than 72.5% of genes selected using SIRS, DC-SIS and Q-SIS by the genes selected using SEVIS are larger than the explained variabilities in the opposite directions with the response variable being TP53.
- The proportion increases to 81.4% with the response variable being changed to days to last followup.

Outline

- Suppose $f(x,y)$ is the joint density function of random variables (X,Y),
- and $f^X(x)$ is the marginal density function of predictor X.

$$\begin{aligned}\text{SEV}(Y|X) &= \frac{\mathrm{E}[\mathrm{E}(Y|X)]^2 - (\mathrm{E}(Y))^2}{\text{var}(Y)} \\ &= \frac{\int (r^{Y|X}(x))^2 f^X(x)\,dx - (E(Y))^2}{\text{var}(Y)},\end{aligned} \tag{6.1}$$

where

$$r^{Y|X}(x) := \mathrm{E}(Y|X=x) := \frac{\phi^{Y|X}(x)}{f^X(x)} := \frac{\int yf(x,y)dy}{f^X(x)}. \tag{6.2}$$

Nadaraya-Watson estimators

$$\phi_n^{Y|X}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) Y_i,$$
$$f_n^X(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \tag{6.3}$$

$$r_n^{Y|X}(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)} & \text{if } \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right) \neq 0, \\ 0 & \text{if } \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right) = 0. \end{cases} \tag{6.4}$$

The estimator

$$\widehat{\omega} = \widehat{\mathrm{SEV}}(Y|X) = S_Y^{-2}\left(\int (r_n^{Y|X}(x))^2 f_n^X(x)\,dx - (\bar{Y})^2\right), \tag{6.5}$$

Nadaraya-Watson estimators

$$\phi_n^{Y|X}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) Y_i, \qquad f_n^X(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \tag{6.3}$$

$$r_n^{Y|X}(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} K(\frac{x-X_i}{h}) Y_i}{\sum_{j=1}^{n} K(\frac{x-X_j}{h})} & \text{if } \sum_{j=1}^{n} K(\frac{x-X_j}{h}) \neq 0, \\ 0 & \text{if } \sum_{j=1}^{n} K(\frac{x-X_j}{h}) = 0. \end{cases} \tag{6.4}$$

The estimator

$$\widehat{\omega} = \widehat{\text{SEV}}(Y|X) = S_Y^{-2}\left(\int (r_n^{Y|X}(x))^2 f_n^X(x)\,dx - (\bar{Y})^2\right), \tag{6.5}$$

Asymptotics

Under Assumptions (A1) and (A2), we have

$$\sqrt{n}(\mathbf{A}^T\Sigma\mathbf{A})^{-\frac{1}{2}}(\widehat{\mathrm{SEV}}(Y|X) - \mathrm{SEV}(Y|X)) \Rightarrow N(0,1), \tag{6.6}$$

where

$$\Sigma = \mathrm{Cov}\Big(\sigma_Y^{-2}\int(2Y_i - \frac{\phi^{Y|X}(x)}{f^X(x)})\frac{\phi^{Y|X}(x)}{f^X(x)}\frac{1}{h}K(\frac{x-X_i}{h})dx,\ \frac{Y_i}{\sigma_Y},\ Y_i^2\Big),$$

and

$$\mathbf{A} = (1,\ -\frac{2\mu_Y}{\sigma_Y} + 2\mu_Y\sigma_Y^{-3}(\int\frac{(\phi^{Y|X}(x))^2}{f^X(x)}dx - \mu_Y^2),\ -\sigma_Y^{-4}(\int\frac{(\phi^{Y|X}(x))^2}{f^X(x)}dx -$$

μ_Y and σ_Y are the population mean and standard deviation of Y, respectively.