PEMILIHAN VARIABEL DAN REDUKSI DIMENSI DALAM REGRESI NONPARAMETRIK BERDIMENSI BESAR

Eva Yanti Siregar
Dosen Program Studi Pendidikan Matematika STKIP Tapanuli Selatan
Jl. Sutan Muhammad Arif Kel. BatangAyumi Jae Padangsidimpuan – 22716
Email: evayanti13@yahoo.com

ABSTRAK

Prosedur l_1 pada model regresi Gauss non-parametrik. Dalam banyak contoh konkrit,dimensi d pada variabel X tergantung pada jumlah pengamatan. Dalam tulisan ini, dibangun dua prosedur. Yang pertama, memilih probabilitas tinggi pada koordinat ini. Kemudian, dengan menggunakan metode pemilihan subset, menjalankan polinomial Estimator untuk memperkirakan fungsi regresi $n^{2\beta/(2\beta+d)}$, di mana d^* merupakan dimensi ”real” dari masalah jumlah variabel yang tergantung pada *f*, telah mengganti bentuk dimensi *d*. Untuk mencapai hasil ini, digunakan metode l_1-penalization dalam setup nonparametrik.

Kata kunci: Reduksi dimensi, Dimensi besar, LASSO.

PENDAHULUAN

Analisa regresi adalah analisis statistik yang mempelajari bagaimana membangun sebuah model fungsional dari data untuk dapat menjelaskan ataupun meramalkan suatu fenomena alami atas dasar fenomena yang lain. Analisa regresi merupakan salah satu teknik statistik yang digunakan secara luas dalam ilmu pengetahuan terapan. Regresi di samping digunakan untuk mengetahui bentuk hubungan antar peubah regresi, juga dapat dipergunakan untuk peramalan.

Model regresi linier merupakan model regresi dalam fungsi regresi yang berbertuk linier. Persamaan $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$ merupakan model regresi linier dengan parameter regresi yang diestimasi berdasarkan data pengamatan. Dengan menggunakan *n* pengamatan untuk suatu model linier sederhana.

Metode yang biasanya digunakan untuk estimasi parameter regresi adalah metode kuadrat terkecil. Metode kuadrat terkecil dapat memberikan hasil yang optimal jika sesatannya diasumsikan berdistribusi normal $\varepsilon \sim N(0,\sigma^2)$. Dengan pemenuhan terhadap asumsi kenormalan dapat digunakan regresi parametrik untuk mengetahui bentuk hubungan antar peubah regresi pada contoh data yang diamati.

Dalam asumsi-asumsi sering terjadi dan terkadang peubah acak yang diamati tidak dapa dianggap menyebar normal. Dari segi statistika persoalan tersebut harus dapat diselesaikan dengan menggunakan teknik statistika. Dalam statistika parametrik, teknik-teknik yang digunakan berhubungan dengan pendugaan parameter serta pengujian hipotesis yang berhubungan dengan parameternya. Asumsi-asumsi yang digunakan pada umumnya menspesifikasikan bentuk sebarannya. Salah satu analisis alternatif lain yang dapat digunakan adalah dengan regresi nonparametrik karena dalam regresi nonparametrik tidak diperlukan pemenuhan asumsi kenormalan. Dalam penelitian ini masalah yang dipertimbangkan adalah masalah dimensi besar.

METODE PENELITIAN

Metode penelitian yang dilakukan adalah bersifat literatur kepustakaan dan dilakukan dengan mengumpulkan informasi dari referensi beberapa buku dan jurnal, memahami penelitian-penelitian yang telah pernah dilakukan oleh peneliti lain yang berhubungan dengan penelitian yang dilakukan.

Teknik Analisa Data

1. Menjelaskan tentang Regresi Parametrik dan Regresi Nonparametrik.
2. Menjelaskan tentang Reduksi Dimensi dalam Regresi Nonparametrik.
3. Menjelaskan tentang Regresi Nonparametrik Berdimensi Besar.
4. Mengidentifikasi kesalahan-kesalan yang ditemukan pada Regresi Gauss Non-parametrik.
5. Memperhitungkan estimator dan konvergensi pada kesalahan yang muncul.
6. Menguraikan pendekatan Polynomial Taylor dalam harga mutlak yang dipakai dalam estimasi dan kekonvergensian.

Perbedaan Regresi Parametrik dan Regresi Nonparametrik

Ada beberapa perbedaan khusus dalam penggunaan prosedur parametrikdan prosedur nonparametrik antara lain:

1. Penggunaan prosedur parametrik didasarkan pada asumsi-asumsi tertentu,misalnya mengasumsikan bahwa sampel yang diambil dari populasi yang berdistribusi normal. Prosedur nonparametrik tidak didasarkan pada asumsi-asumsi yang mengikuti suatu distribusi tertentu dan dapat digunakan apabila asumsi yang diperlukan pada penggunaan prosedur parametrik menjaditidak valid.
2. Dalam kasus parametrik untuk mengetahui bentuk hubungan antar peubah respon pada contoh data yang diamati dapat digunakan Metode KuadratTerkecil dan Metode Maksimum Likelihood. Dalam regresi nonparametricuntuk memperkirakan parameter digunakan metode Theil dengan koefisienkemiringan garis regresi sebagai median kemiringan dari seluruh pasangangaris dari titik-titik dengan nilai-nilai X yang berbeda atau independen.
3. Pengujian hipotesis untuk model parametrik menggunakan statistik uji $t$yang merupakan sebuah hasil asumsi secara normal yang didasarkan darimetode kuadrat terkecil. Pengujian hipotesis pada regresi nonparametricmenggunakan metode Theil yang disusun berdasarkan statistik t Kendall.
4. Interval kepercayaan pada regresi parametrik adalah pembentukan interval kepercayaan untuk parameter yang didasarkan pada metode kuadratterkecil dan asumsi yang digunakan masih sama dengan asumsi yang digunakan pada pengujian hipotesis. Interval kepercayaan pada regresi non-parametrik adalah pembentukan interval kepercayaan hanya untuk koefisienkemiringan.

Kurva regresi digunakan untuk menjelaskan hubungan antara peubah penjelas dengan peubah terikat. Pendekatan yang paling sering digunakan adalah pendekatan parametrik. Asumsi yang mendasari pendekatan ini adalah kurva regresi yang diwakili oleh suatu model parametrik (Hardle, 1990). Dalam regresi parametrik, diasumsikan bahwa bentuk kurva regresi diketahui berdasarkan teori, informasi sebelumnya, atau sumber lain yang dapat memberi pengetahuan secara rinci.

Apabila model dari pendekatan parametrik diasumsikan benar, maka pendugaan parametrik akan sangat efisien. Tetapi jika tidak, menyebabkan interpretasi data yang menyesatkan. Selain itu, model parametrik mempunyai keterbatasan untuk menduga pola data yang tidak diharapkan. Jika asumsi bentuk kurva parametrik ini tidak terpenuhi, maka kurva regresi dapat diduga menggunakan model regresi dari pendekatan nonparametrik.

Pendekatan nonparametrik merupakan metode pendugaan model yang dilakukan berdasarkan pendekatan yang tidak terikat asumsi bentuk kurva regresi tertentu. Kurva regresi berdasarkan pendekatan nonparametrik ini, diwakili oleh model yang disebut model regresi nonparametrik. Karena sebelumnya tidak ada asumsi mengenai bentuk kurva regresi, model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linier atau nonlinier. Semua fungsi dapat digunakan untuk pendugaan dalam model regresi. Komputasi atau perhitungan dalam menduga model, merupakan kendala utama dalam regresi nonparametrik. Seiring dengan perkembangan media komputer yang sangat pesat dewasa ini, regresi nonparametrik turut berkembang pula. Ada beberapa teknik pendugaan nilai peubah respons dalam regresi nonparametrik yakni penduga kernel, regresi spline, regresi lokal, dll.

Regresi Nonparametrik Berdimensi Besar

Tujuan analisa regresi adalah untuk mempelajari bagaimana respon sebuah peubah variabel *Y* terhadap perubahan yang terjadi pada peubah lain yaitu variabel *X*. Hubungan antara variabel *X* dan variabel *Y* dapat dituliskan sebagai berikut (Bertin dan Lecue, 2008):

$$y = f(X_i) + \varepsilon_i, \quad i = 1, 2, ..., n$$

Dengan *Y* adalah peubah terikat, fungsi *f(x)* adalah Fungsi regresi nonparametrik, ε adalah Error faktor gangguan yang tidak dapat dijelaskan oleh model.

Fungsi *f* adalah fungsi matematik yang disebut sebagai fungsi regresi dan ε_i adalah error yang mengijinkan terjadinya deviasi dari hubungan yang murni deterministik. Pada aplikasi dikumpulakan data $(X_i, Y_i), ..., (X_n, Y_n)$ yang berisi informasi tentang fungsi *f*. Dari data-data ini dapat diduga ataupun mengestimasi fungsi *f* tersebut. Jika pengetahuan tentang fungsi *f* ini minim, maka estimasi terhadap fungsi f ini dapat didekati secara nonparametrik. Agar pendekatan nonparametrik ini menghasilkan estimasi terhadap fungsi *f* yang masuk akal, maka hal yang harus diperhatikan adalah asumsi bahwa fungsi *f* memiliki derajat kemulusan. Biasanya kontinuitas dari fungsi *f* merupakan syarat yang cukup untuk menjamin sebuah estimator akan konvergen pada fungsi *f* yang sesungguhnya bila jumlah data bertambah tanpa batas.

Dalam penggunaan umum, dimensi berarti parameter atau pengukuran yang dibutuhkan untuk mendefinisikan sifat-sifat suatu objek yaitu panjang, lebar, dan tinggi atau ukuran dan bentuk. Dalam matematika, tidak ada satu pun definisi yang mencukupi untuk menyatakan konsep dalam segala situasi yang digunakan. Konsekuensinya, matematikawan membagi sejumlah definisi dimensi ke dalam tipe-tipe yang berbeda. Semuanya didasarkan pada konsep dimensi Euclides berruang *n* sehingga menjadi E^n. Maka diperoleh:

1. Titik E^0 adalah dimensi 0,

2. Garis E^1 adalah dimensi 1,
3. Bidang E^2 adalah dimensi 2,
4. E^n adalah dimensi *n*.

Estimator Kernel

Suatu fungsi *K(•)* disebut fungsi kernel jika fungsi *K* fungsi kontinu berharga riil, simetris, terbatas dan $\int_{-\infty}^{\infty} K(y)dy = 1$. Jika *K* suatu kernel dengan sifat

1. $\int_{-\infty}^{\infty} x^j K(x)dx = 0,$ untuk $j = 1, 2, ..., r-1$
2. $\int_{-\infty}^{\infty} x^r K(x)dx \neq 0,$ atau ∞ maka *K* disebut kernel order *r*.

Secara umum estimator regresi kernel dari *g* adalah estimator kuadrat terkecil, dengan fungsi bobot $W_{ni(x)}$ tergantung pada kernel *K*. Jika densitas variabel *X* tidak diketahui, Hardle (1990) memberikan bobot $W_{ni}(x) = \frac{K_h(x - X_i)}{\hat{f}_h(x)}$ dengan $\hat{f}_h(x) = n^{-1}\sum K_h(x - X_i)$ dan $K_h(u) = h^{-1}K\left(\frac{u}{h}\right)$ sehingga estimator kernel dari regresi *g* adalah:

$$\hat{g}h(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{j-1}^{n} K_h(x - X_j)}.$$

Selanjutnya, jika densitas variabel *X* diketahui, Greblicki (1974) cit. Hardle (1990) memberikan bobot $W_{ni}(x) = K_h(x - X_i)/f(x)$, sehingga estimator kernel dari regresi *g* adalah:

$$\hat{g}h(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{f(x)}.$$

Kemudian dalam model rancangan tetap dari ruang yang sama dengan $\{X_i\}, i = 1, 2, ..., n$ tetap pada [0,1], Priestley dan Chao (1972) cit. Hardle (1990) memberikan bobot $W_{ni}(x) = n(X_i - X_{i-1})K_h(x - X_i), X_0 = 0$ dan $\hat{f}(x) = (n(X_i - X_{i-1}))^{-1}$ untuk $x \in (X_{i-1} - X_i)$, sehingga estimator kernel regresi *g* adalah:

$$\hat{g}h(x) = (nh)^{-1}\sum_{i=1}^{n} K_h(x - X_i).$$

Lemma Pada model rancangan tetap variabel *X* satu dimensi didefinisikan $C_K = \int_0^1 K^2(u)du$ dan $d_K = \int_0^1 u^2 K(u)du$. Diambil $W_{ni} = n(X_i - X_{i-1}) K_h(x - X_i)$ dengan asumsi

1. *K* mempunyai support [-1,1] dengan *K(-1) = K(1) = 0*
2. $g \in c^2$
3. $X_i = i/n$, $i = 1, 2, ..., n$
4. $\text{var}(\varepsilon_i) = \sigma^2$, $i = 1, 2, ..., n$
5. $n \to \infty, h \to 0$ dan $nh \to \infty$

Dalam estimator kernel, tingkat kemulusan $\hat{g}h$ ditentukan oleh fungsi kernel *K* dan *h* yang disebut parameter pemulus, tetapi pengaruh kernel *K* tidak sedominan parameter pemulus *h*. Nilai *h* kecil memberikan grafik yang kurang mulus sedangkan nilai *h* besar memberikan grafik yang sangat mulus. Oleh karena itu, perlu dipilih nilai *h* optimal untuk mendapatkan grafik optimal.

Seleksi Prosedur Estimasi

Estimasi adalah keseluruhan proses yang menggunakan sebuah estimator untuk menghasilkan sebuah estimate dari suatu parameter. Untuk menyeleksi Prosedur Estimasi yang dilakukan pertama-tama adalah menentukan himpunan indeks $J = \{i_1, ..., i_{d^*}\}$. Kemudian menyusun sebuah penaksir dari nilai *f(x)* yang konvergen untuk nilai $n^{-2\beta/(2\beta+d^*)}$ pada $f \in \sum(\beta, x)$ untuk $\beta > 1$. Dalam menilai tujuan pertama, gunakan l_1 polinomial estimator.

Estimasi Titik untuk Kurva Regresi

Estimasi kurva regresi umumnya dilakukan dengan pendekatan parametrik yang mulai diperkenalkan oleh Laplace sejak abad ke XVIII dan juga Boscovich pada tahun 1757. Dalam regresi parametrik diasumsikan bahwa bentuk kurva regresi fungsi *f* diketahui. Pembuatan asumsi tersebut berdasarkan pada teori, pengalaman masa lalu atau tersedianya sumber-sumber lain yang dapat memberi pengetahuan atau informasi yang terperinci.

Estimasi dapat juga dilakukan berdasarkan pendekatan yang tidak terikat dengan asumsi bentuk kurva regresi tertentu, yang memberikan fleksibilitas yang lebih besar dari kurva regresi. Metode pendekatan seperti ini dinamakan pendekatan nonparametrik yang mulai dikenal sejak abad ke XIX. Ada beberapa teknik untuk mengestimasi dalam regresi nonparametrik, antara lain histogram, estimator Kernel, Spline, dan lain-lain.

Masalah yang sering muncul dalam regresi adalah tidak semua variabel penjelas dapat didekati dengan pendekatan parametrik, karena tidak adanya informasi tentang bentuk hubungan variabel penjelas tersebut dengan variabel responnya, sehingga harus digunakan pendekatan nonparametrik. Dengan menggabungkan dua pendekatan tersebut dalam suatu pendekatan regresi akan didapatkan suatu model semiparametrik. Estimasi model semiparametrik ekuivalen dengan mengestimasi parameter-parameter pada komponen parametrik dan estimasi kurva pada komponen nonparametrik.

Bentuk kurva regresi fungsi *f* diasumsikan oleh smooth, dalam arti bahwa fungsi *f* termuat di dalam ruang Sobolev $W_2^p[a,b]$, dengan

$$W_2^p[a,b] = \left\{g; \int \left(f^{(p)}(x)\right)^2 dx < \infty\right\}$$

Untuk suatu p bilangan bulat positif, dan e_i sesatan random yang diasumsikan berdistribusi normal dengan rata-rata nol dan variansi σ^2. Untuk mendapatkan estimasi kurva regresi fungsi f menggunakan optimasi

$$\underset{f \in W_2^p\{a,b\}}{Min} \sum_{i=1}^{n} (y_i - f(x))^2$$

dengan suatu syarat,

$$g(f) = \int_a^b \left(f^{(p)}(x)\right)^2 dx \leq \rho, \quad \rho \geq 0$$
.

Menurut Eubank (1988), estimasi ini ekuivalen dengan penalized least square (PLS) yaitu penyelesaian optimasi seperti berikut:

$$\underset{f \in W_2^p\{a,b\}}{Min} \left\{ n^{-1} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_a^b (f^{(p)}(x_i))^2 dx \right\}$$

Ruas pertama pada persamaan di atas merupakan fungsi yang mengukur kecocokan data *(goodness of fit)*, sedangkan ruas merupakan ukuran kekasaran kurva *(roughness penalty)* dengan λ sebagai parameter yang mengontrol *goodness of fit* dan *roughness penalty*.

Umumnya estimasi fungsi $\hat{f}$ diperoleh dari meminimumkan *Penalized Likelihood* (PL). Untuk menyelesaikan optimasi *Penalized Likelihood* (PL), digunakan pendekatan *Reproducing Kernel Hilbert Space* (RKHS) atau *Gateaux*. Sedangkan untuk persoalan inferensi seperti estimasi interval untuk fungsi f yang menggunakan pendekatan Bayesian. Tetapi pendekatan ini memerlukan pengetahuan matematika yang relative tinggi dan sulit dipahami oleh banyak pengguna Statistika.

Namun untuk menduga kurva regresi yang diperoleh dari optimasi *Likelihood* dapat menjadi pilihan yang cukup baik karena secara matematik mudah dan sederhana. Sedangkan untuk mengkonstruksi selang kepercayaan pada kurva regresi, beberapa peneliti seperti wahba (1983) menggunakan pendekatan Bayesian dengan menggunakan *prior improper* sehingga secara matematis cukup sulit. Akan tetapi jika selang kepercayaan diperoleh dengan pendekatan *Privotal Quantity* tidak akan melibatkan distribusi *prior*, sehingga diperoleh model yang sederhana dan inferensi statistik yang relative mudah (Eubank, 1988).

PEMBAHASAN

Hasil utama yang diperoleh berdasarkan penjelasan-penjelasan yang telah dipaparkan pada bab-bab sebelumnya. Hasil utama dari penelitian ini dapat diperoleh dari model regresi nonparametrik berikut $y = f(X_i) + \varepsilon_i, \; i = 1, 2, ..., n$. Dan untuk mengestimasi fungsi regresi dilakukan dengan pendekatan nonparametrik. Salah satu metode dalam mengetimasi regresi nonparametrik adalah dengan metode kernel *(K)* di mana $K : \Re^d \to \Re$, bandwidth yaitu $\theta = (\theta_0, ..., \theta_d)^t \in \Re^{d+1}, h > 0$ dan parameter regularisasi di mana $\lambda > 0$. Dan untuk menghindari kerumitan teknis akan diasumsikan pada fungsi μ desain X dalam asumsi sebagai berikut:

Asumsi 1 Terdapat beberapa konstanta $\eta, \mu_m > 0, \mu_m \geq 1$ dan $L_\mu > 0$ sedemikian sehingga

1. $B_\infty(x,\eta) \subset \sup(\mu)$ dan $\mu_m \le \mu(y) \le \mu_M$ untuk hampir setiap $y \in B_\infty(x,\eta)$,
2. μ adalah L_μ-Lipschitzian sekitar *x*, yaitu untuk sebarang $t \in B_\infty(x,1), |\mu(x) - \mu(t)| \le L_\mu \|x-t\|_\infty$.

Hasil pertama berhubungan dengan sifat statistik dari prosedur seleksi. Untuk tahap ini, memerlukan asumsi keteraturan untuk fungsi regresi *f*. Asumsi ini dipenuhi untuk sebarang β-Hlderian dalam fungsi *x* dengan $\beta > 1$.

Asumsi 2 Terdapat konstan mutlak $L > 0$ yang memenuhi kondisi. Fungsi regresi *f* terdiferensial dan,

$$| f(t) - P_1(f)(t,x)| \le L\|t-x\|_1^\beta, \forall t \in B_\infty(x,1) \; |$$

dengan $P_1(f)(.,x)$P1(f)(., x) adalah polinomial Taylor dari derajat 1 dan dari fungsi *f* pada titik *x*.

Untuk mencapai suatu seleksi efisien pada koordinat yang menarik, harus dapat membedakan turunan parsial tidak nol dari fungsi *f* dari turunan parsial nol.

Lemma 3 Jika $\bar{\theta} \in \Re^{d+1}$ dan $\bar{\theta}^2 \in \Re^{d+1}$ dua solusi pada *(S)* sehingga $A\bar{\theta} = A\bar{\theta}^2$
Bukti: Menunjukkan bahwa $S(\bar{\theta})$ himpunan $j \in \{0,...,d\} : \bar{\theta}_j \ne 0$. Untuk sebarang $v \in \Re^{d+1}$, diperoleh

$$\varphi(\bar{\theta}+v) - \varphi(\bar{\theta}) = 2\lambda \sum_{j\in S(\bar{\theta})} \left|\bar{\theta}_j + v_j\right| - \left|\bar{\theta}_j\right| -$$

$$v_j \; sigh(\bar{\theta}_j) + 2\lambda \sum_{j\in S(\bar{\theta})} \left|v_j\right| - \eta_j v_j + \|Av\|_2^2$$

dengan $\eta_j = \lambda^{-1}(A_j)^t(Z - A\bar{\theta})$. Untuk sebarang $j \notin S(\bar{\theta})$, diperoleh $\left|\eta_j\right| \le 1$ sehingga $\left|v_j\right| - \eta_j v_j \ge 0$. Oleh karena itu,

$$\phi(\bar{\theta}+v) - \phi(\bar{\theta}) \ge \|Av\|_2^2$$

Ambil $v \in \Re^{d+1}$ sedemikian hingga $\bar{\theta}^2 = \bar{\theta} + v$. Vektor $\bar{\theta}^2$ dan $\bar{\theta}$ keduanya solusi dari *(S)*, sehingga meminimalkan φ dan sehingga $\varphi(\bar{\theta}^2) = \varphi(\bar{\theta})$. Oleh karena itu, diperoleh $\|Av\|_2^2 = 0$.Selanjutnya, membuktikan hasil yang berkaitan dengan model identifiability serta keunikan LASSO. Diperkenalkan cara

$$\Omega_{01} := \forall\theta \in \Re^{d+1} : \frac{1}{2}\sqrt{\frac{\mu_m}{2}}\|\theta\|_2 \le \|A\theta\|_2 \le$$

$$2\sqrt{\frac{3_{\mu M}}{2}}\|\theta\|_2$$

KESIMPULAN

Adapun kesimpulan yang diperoleh berdasarkan hasil pembahasan adalah sebagai berikut:

Untuk mengestimasi fungsi regresi yang sulit dapat dilakukan dengan pendekatan nonparametrik. Salah satu metode dalam mengestimasi regresi nonparametrik adalah dengan metode kernel *(K)*, bandwidth atau smoothing parameter dan parameter regularisasi. Suatu ukuran kebaikan estimator dari fungsi regresi *f* dapat dilihat dari tingkat kesalahannya. Semakin kecil tingkat kesalahannya semakin baik estimasinya.

Kelayakan penggunaan metode regresi parametrik dapat diuji dengan membandingkan terhadap metode regresinonparametrik. Adapun model regresi nonparametrik adalah $y = f(X_i) + \varepsilon_i, \quad i = 1, 2, ..., n$. Estimasi terhadap fungsi regresi β-regular fungsi *f* tidak lebih cepat dari laju $n^{-2\beta/(2\beta+d)}$. Dalam beberapa situasi, fungsi *f* tergantung hanya pada beberapa dari koordinat *X*. Dalam banyak contoh konkrit, dimensi *d* pada variabel *X* tergantung pada jumlah pengamatan. Dalam tulisan ini, dibangun dua prosedur. Yang pertama, memilih probabilitas tinggi pada koordinat ini. Kemudian, dengan menggunakan metode pemilihan subset, menjalankan polinomial Estimator untuk memperkirakan fungsi regresi $n^{-2\beta/(2\beta+d)}$, dengan d^* merupakan dimensi ”real” dari masalah jumlah variabel yang tergantung pada fungsi *f*, telah mengganti bentuk dimensi *d*. Dan untuk mencapai hasil ini, digunakan metode l_1-penalization dalam setup nonparametrik.

DAFTAR PUSTAKA

Bertin, Karine. and Lecue, Guillaume. (2008). Selection of Variables and Dimension Reduction in High-Dimensional Nonparametric Regression.

Eubank, R. L. (1988). Spline Smoothing and Nonparametrik Regression, Marcel Dekter, New York.

Hardle, W. (1990). Applied Nonparametric Regression. Cambridge, University Press. New York.

Wahba, G. (1983). Bayesian Confidence Interval for the Cross validated Smoothing parameter in the Generalized Spline Smoothing Problems, The Annals of Statistics.