BAB 2

TINJAUAN PUSTAKA

Model Regresi Gauss Nonparametrik ditulis sebagai berikut (Bertin dan Lecue (2008)):

$$Y = f(X_i) + e_i, \quad i = 1, 2, ..., n,$$

dengan variabel input $X_1, ..., X_n$ merupakan variabel acak n bersebaran bebas identik (i.i.d) dengan nilai R^d, dengan $e_1, ..., e_n$ sampai n merupakan variabel acak Gauss dengan variansi σ^2 bebas dari X_i dan f fungsi regresi. Terpenting pada penilaian pointwise dari fungsi f pada titik tertentu $x = (x_1, ..., x_d) \in R^d$. Dibutuhkan beberapa konsep proses penilaian $\widehat{f_n}$ memiliki pointwise terkecil yang digabungkan dengan kuadrat resiko.

$$E\left(\widehat{f_n}(x) - f(x)\right) \tag{2.1}$$

hanya menggunakan sekumpulan data $D_n = (Y_i, X_i)_{1 \leq i \leq n}$.

Asumsikan bahwa fungsi regresi memiliki beberapa sifat beraturan sekitar x adalah suatu asumsi klasik untuk permasalahan ini. Pada tulisan ini asumsikan fungsi f sebagai $\beta - Hlderian$ sekitar x. Diingat kembali bahwa fungsi $f : R^d \mapsto R$ adalah $\beta - Hlderian$ pada titik x dengan $\beta > 0$, dinotasikan oleh $f \in \sum(\beta, x)$ ketika dua titik berikut memenuhi:

1. Fungsi f adalah $l - kali$ terdiferensial pada x (dengan $l = [\beta]$ adalah bilangan bulat terbesar yang tepat lebih kecil dari β),
2. Terdapat $L > 0$ sedemikian hingga untuk sebarang $t = (t_1, ..., t_n) \in B_\infty(x, 1)$,

$$|f(t) - P_i(f)(t, x)| \leq L \|t, x\|_1^\beta,$$

dengan $P_l(f)(., x)$ adalah Polinomial Taylor pada orde l menghubungkan dengan fungsi f pada titik x, $\|\cdot\|_1$ adalah l_1 norm dan $B_\infty(x, 1)$ adalah satuan l_∞-bola pada pusat x dan jari-jari 1.

Dalam matematika, seri Taylor adalah representasi dari suatu fungsi sebagai jumlah tak terbatas, dihitung dari nilai turunannya pada satu titik. Seri Taylor secara resmi diperkenalkan oleh matematikawan Brook Inggris Taylor. Jika seri ini berpusat di nol, seri ini juga disebut seri maclaurin, dinamai ahli matematika Skotlandia Colin Maclaurin yang menggunaka banyak kasus dari deret taylor di abad ke-18. Seri Taylor dapat dianggap sebagai batas dari Polinomial Taylor.

ketika fungsi f hanya diasumsikan pada $\sum(\beta, x)$, tidak ada estimator yang dapat konvergen ke fungsi f (untuk kemungkinan yang diberikan pada persamaan (1.1)) lebih cepat dari,

$$n^{-2\beta/(2\beta+d)}. \tag{2.2}$$

Asumsi 2.1 *Terdapat bilangan bulat $d^* \leq d$, sebuah fungsi $g : R^{d^*} \to R$ dan sebuah subset $J = \{i_1, ..., i_{d^*}\} \subset \{1, ..., d\}$ kardinalitas d^* sehingga untuk setiap $(x_1, ..., x_d) \in R^d$ berlaku*

$$f(x_1, ..., x_d) = g(x_{i_1}, ..., x_{i_{d^*}}).$$

Berdasarkan Asumsi (2.1) dimensi "real" pada permasalahan tidak lagi disebut fungsi d tetapi fungsi d^*. Selanjutnya, diharapkan bahwa jika $f \in \sum(\beta, x)$ (yang mana dapat juga dikatakan bahwa g adalah β-*Hlderian* pada titik x), memungkinkan mengestimasi fungsi $f(x)$ seperti pada persamaan (1.2) di mana fungsi d digantikan oleh fungsi d^*, mengarahkan pembuktian kekonvergensian ketika $d^* << d$. Namun demikian, pembuktian dimulai dari data D_n, hal ini tidak menunjukkan bahwa pendeteksian himpunan koordinat J adalah tugas yang mudah. Untuk memilih himpunan ini, gunakan teknik l_1-penalization. Teknik ini banyak digunakan dalam masalah-masalah yang bersifat parametrik (cf. Bickel et al. (2008), Zhao dan Yu (2006), Meinshausen dan Yu (2008) dan referensi di dalamnya).

Teorema 2.2 *Berdasarkan Asumsi (2.1) sangat tepat untuk menyusun konsep, hanya dari nilai data D_n, sebuah subset $\hat{J} \subset \{1, ..., d\}$ sedemikian sehingga, probabilitas yang lebih besar dari $1 - c_0 \exp(c_0 d - c_1 n h^{d+2})$ (untuk sebuah parameter*

bebas $o < h < 1$*) Karine Bertin dan Guillaume Lecue (2008),*

$$\hat{J} = J.$$

Teorema 2.3 *Untuk sebarang* $f \in \sum(\beta, x)$ *dengan* $\beta > 1$ *yang memenuhi A-sumsi (2.1) memungkinkan untuk mengkonstruksi berdasarkan data* D_n*, prosedur estimasi* $\hat{f}_n$ *dapat dituliskan sebagai berikut*

$$P\left[\left|\hat{f}_n(x) - f(x)\right| \geq \delta\right] \leq c\,\exp\left(-c\,\delta^2\,n^{2\beta/(2\beta+d^*)}\right), \forall \delta > 0$$

dengan c tidak bergantung terhadap n (Karine Bertin dan Guillaume Lecue (2008)).

Masalah yang dipertimbangkan dalam tulisan ini disebut masalah dimensi besar. Banyak tulisan sebelumnya yang mempelajari macam-macam permasalah yang meringkas keadaan yang tidak mungkin (Lafferty dan Wasserman (2008)). Dalam Bickel dan Li (2007), Levina dan Bickel (2005), Belkin dan Niyogi (2003), Donoho dan Grimes (2003), diasumsikan bahwa bentuk variabel X termasuk dimensi kecil dengan dimensi $d^* < d$. Semua permasalahan didasarkan pada teknik heuristik. Lafferty dan Wasserman (2008), masalah yang sama sebagai satu pertimbangan disini teratasi.