Forum Statistika dan Komputasi, April 2008, p: 7-10 Vol 13 No.1
ISSN : 0853-8115

PENGGUNAAN JARINGAN SYARAF TIRUAN UNTUK PENDUGAAN MODEL LINEAR TERAMPAT DENGAN KOEFISIEN KERAGAMAN KONSTAN

Anik Djuraidah

Departemen Statistika, FMIPA IPB

Abstrak

Secara umum model linear terampat (GLIM) dapat dipetakan secara ekivalen pada Jaringan Syaraf Tiruan (JST) dengan satu lapisan atau disebut juga dengan perceptron. Fungsi aktifasi pada JST sama dengan invers dari fungsi hubung. Pada GLIM dengan komponen acak mempunyai sebaran gamma ekivalen dengan JST tanpa lapisan tersembunyi dengan fungsi galat adalah gamma dan fungsi tujuan adalah fungsi kemungkinan maksimum atau devians. Sedangkan fungsi aktifasi untuk model gamma adalah identitas, resiprokal, atau eksponensial. Makalah ini mengkaji pendugaan model pada data yang mempunyai sebaran gamma dengan metode JST dan seberapa besar perbedaan hasil pendugaannya dibandingkan dengan GLIM. Hasil kajian menunjukkan bahwa JST menghasilkan nilai dugaan yang sama dengan GLIM.

Kata kunci : jaringan syaraf tiruan, koefisien keragaman konstan, model gamma

PENDAHULUAN

Dalam statistika, jaringan saraf tiruan (artificial neural network selanjutnya disingkat JST) merupakan suatu kelas yang luas dari model regresi nonlinear, model reduksi data, dan sistem dinamik nonlinear. JST mampu memproses berbagai macam data dan membuat prediksi yang kadang kala sangat akurat.

Pada JST terdapat dua proses yaitu training dan testing. Proses training pada JST sama dengan metode pendugaan dalam statistika, akan tetapi algoritmanya lebih lambat dibandingkan dengan metode statistika. Sehingga untuk analisis data lebih disukai menggunakan paket program statistika dari pada paket program JST. Sedangkan proses testing pada JST identik dengan validasi model pada statistika (Cheng dan Titterington, 1994; Sarle, 1996). Beberapa model JST identik dengan metode statistika antara lain model linear terampat, regresi polinomial, regresi nonparametrik, komponen utama, analisis diskriminan dan analisis gerombol (Sarle, 1994; Jordan dan Bishop, 1996).

Model linear terampat (generalized linear model) merupakan pengembangan model linear klasik di mana nilaitengah populasi tergantung pada prediktor linear melalui suatu fungsi hubung (link function) dan sebaran peluang peubah respons dari keluarga eksponensial. Secara umum model linear terampat terdiri dari tiga komponen (McCullagh dan Nelder, 1989, Warner dan Misra, 1996) yaitu :

Komponen sistematik atau komponen linear

$$\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$$

Fungsi hubung h yang menjelaskan hubungan antara nilai harapan yi dengan prediktor linear

$$h(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$$

Komponen acak dari peubah respon yi yang mempunyai sebaran dari keluarga eksponensial dengan rataan μ dan ragam σ2

Pada model yang mempunyai koefisien keragaman konstan diasumsikan

$$\text{Ragam}(Y_i) = \sigma^2 E(Y_i) = \sigma^2 \mu_i$$

dengan σ adalah koefisien keragaman Y. Bila nilai σ kecil, dapat dilakukan transformasi logaritma untuk menstabilkan ragam. Atau bila komponen sistematik dari model bersifat multiplikatif dapat diubah menjadi aditif dengan melakukan transformasi logaritma pada data asli dan menggunakan metode kuadrat terkecil untuk pendugan parameter.

Bila peubah respon Y berdistribusi gamma dengan indeks konstan $\upsilon = 1/\sigma^2$ yang bebas terhadap nilaitengah maka fungsi kepekatannya adalah

$$f(y) = \begin{cases} \dfrac{1}{\Gamma\upsilon}\left(\dfrac{\upsilon y}{\mu}\right)^{\upsilon} \exp\left(-\dfrac{\upsilon y}{\mu}\right) d(\log y) & \text{untuk } y \geq 0, \\ & \upsilon > 0,\ \mu > 0 \\ 0 & \text{untuk lainnya} \end{cases}$$

Parameter υ menentukan bentuk sebaran. Nilaitengah dan ragam Y adalah

$$E(Y) = \mu \quad \text{dan} \quad \text{Ragam}(Y) = \mu^2 / \upsilon.$$

Parameterisasi sebaran gamma dalam bentuk keluarga eksponen menghasilkan parameter kanonik $\theta = -1/\mu$ dan fungsi kumulan $b(\theta) = -\log(-\theta)$. Sehingga rataannya adalah $b'(\theta) = \mu$ dan fungsi ragam $b'(\theta) = \mu^2$. Fungsi hubung kanonik menghasilkan statistik cukup yang merupakan fungsi linear dari data dinyatakan sebagai $\eta = \mu^{-1}$. Selain fungsi hubung kanonik juga bisa digunakan fungsi hubung logaritma atau identitas (McCullagh dan Nelder, 1989).

Untuk υ yang konstan dan pengamatan yang bebas stokastik dan identik, fungsi log kemungkinan maksimum dapat ditulis sebagai

$$\ell(y, \mu) = \sum_{i} \upsilon\left(y_i / \mu_i - \log \mu_i \right) \quad \ldots . \ (1)$$

Sehingga deviannya adalah

$$D(y, \hat{\mu}) = -2\sum_{i}\left[\log\left(y_i / \hat{\mu}_i \right) - \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i}\right] \quad \ldots \ (2)$$

Sedangkan penduga kemungkinan maksimum untuk parameter dispersi

$$\upsilon^{-1} \cong \frac{\overline{D}(6 + \overline{D})}{(6 + 2\overline{D})}$$

Pada tulisan ini akan dikaji pendugaan model dengan koefisien keragaman konstan dengan metode jaringan syaraf tiruan (JST) dan seberapa besar perbedaan hasil pendugaannya dibandingkan dengan model linear terampat (GLIM).

JARINGAN SYARAF TIRUAN UNTIK MODEL DENGAN GALAT MENYEBAR GAMMA

Sesuai dengan sistem kerjanya struktur JST terdiri dari tiga lapisan yang terdiri dari lapisan input (input layer), lapisan antara atau tersembunyi (hidden layer) dan lapisan output (output layer). Setiap lapisan terdiri dari beberapa neuron dan antar neuron-neuron ini akan dihubungkan dengan neuron lain pada lapisan terdekat. Pada setiap lapisan diberi pembobot yang akan mentransformasi nilai input menjadi nilai output. Sedangkan output adalah suatu penjumlahan dari hasil kali antara bobot dengan neuron-neuron pada lapisan input dan lapisan tersembunyi dengan suatu fungsi aktifasi tertentu (Sarle,1994; Warner dan Misra, 1996).

Secara umum model linear terampat dapat dipetakan secara ekivalen pada JST dengan satu lapisan atau disebut juga dengan perceptron (Sarle, 1994). Fungsi aktifasi pada JST sama dengan invers dari fungsi hubung (Λ = g-1) (Warner dan Misra, 1996). Untuk model dengan galat menyebar gamma ekivalensi fungsi hubung pada GLIM dengan fungsi aktifasi tertera pada Tabel 1.

Proses training pada JST terbagi dalam tiga tahap utama yaitu : feed-forward, back-propagation, dan update nilai bobot. Pada tahap feed-forward dilakukan dari proses input sampai diperoleh output, sedangkan tahap back-propagation dilakukan perbandingan nilai output dari feed-forward dengan nilai target, kemudian dilanjutkan ke depan ke lapisan input sehingga diperoleh galat. Pada tahap update dilakukan dilakukan pembaruan nilai bobot sampai diperoleh nilai fungsi galat minimal.

Tabel 1. Ekivalensi antara Fungsi Hubung dengan Fungsi Aktifasi

Fungsi Hubung	Fungsi Aktifasi
Resiprokal	Resiprokal
Logaritma	Eksponensial
Identitas	Identitas

Bentuk feed-forward JST untuk model linear terampat dengan galat menyebar gamma disebut perceptron gamma tertera pada Gambar 1. Perceptron mempunyai p+1 unit input dengan x0 adalah input konstan dan satu unit output. Nilai input xi diboboti dengan pembobot wi dan jumlah input-input terboboti ini kemudian ditransformasi dengan fungsi aktifasi. Nilai output $\hat{O}$ dapat dinyatakan sebagai sebuah fungsi f dari nilai input x dan bobot wi yaitu :

$$\hat{O} \;=\; f(\mathbf{x}, \mathbf{w}) = \Lambda(w_0 + \sum_{i=1}^{p} w_i x_i) \quad \ldots . \ (3)$$

dengan x = (x1, . . . , xp) adalah vektor input, w = (w1, . . . , wp) adalah vektor pembobot, dan Λ (x) = fungsi aktifasi.

Pada proses training dilakukan pendugaan terhadap bobot $\hat{\mathbf{w}}$ berdasarkan pengamatan $(\mathbf{x}^{(1)}, y^{(1)}) \ \cdots \ (\mathbf{x}^{(N)}, y^{(N)})$ dengan meminimum-kan fungsi tujuan (objective function). Bentuk fungsi tujuan tergantung bentuk fungsi galat (error/cost function). Fungsi tujuan yang paling umum adalah jumlah kuadrat galat, tetapi bisa juga fungsi kemungkinan (Warner dan Misra, 1996). Sedangkan menurut SAS Online Help v 8.2 fungsi tujuan ada tiga, yaitu kemungkinan maksimum, devians, dan pendugaan-M

Menurut Jordan dan Bishop (1996), proses training bertujuan mendapatkan penduga MAP (maximum a priori) bagi parameter-parameter dengan memaksimumkan peluang dari parameter-parameter pada data D sehingga

$$p(\mathbf{w} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathbf{w})\, p(\mathbf{w})}{p(\mathcal{D})} \quad \ldots\ldots.(4)$$

Maksimisasi persamaan (4) sama dengan memaksimumkan suku pembilangnya karena penyebut tidak tergantung w. Sehingga maksimisasi persamaan (4) ekivalen dengan minimisasi dari negatif logaritma suku pembilangnya. Fungsi ini dikenal dengan fungsi galat kemungkinan, yaitu

$$J(\mathbf{w}) = -\ln\ p(\mathcal{D} \mid \mathbf{w}) - \ln\ p(\mathbf{w}) \quad \ldots\ldots..(5)$$

Jika pengamatan pada data D menyebar bebas stokastik dan identik maka

$$p(\mathbf{w} \mid \mathcal{D}) = \prod_{n=1}^{N} p(y_n \mid x_n, \mathbf{w})$$

dengan yn adalah nilai target ke-n, sehingga fungsi log kemungkinan

$$\mathcal{L} = \sum_{n=1}^{N} \ln p(y_n \mid x_n, \mathbf{w}).$$

Bila diasumsikan peluang awal dari parameter seragam untuk semua parameter maka suku kedua pada persamaan (5) dapat dihilangkan. Hal ini berati minimisasai fungsi galat sama dengan maksimisasi fungsi log kemungkinan yaitu memilih wML yang memaksimumkan $p(\mathbf{w} \mid \mathcal{D})$.

Setelah menentukan model probabilistik, fungsi tujuan, masalah selanjutnya adalah menemukan prosedur yang efisien untuk perhitungan gradien dari fungsi galat. Misalkan $z_i^{(n)}$ adalah nilai output dari lapisan ke-i pada pengamatan ke-n, untuk model (3) karena hanya satu lapisan maka

$$z^{(n)} = w_0 + \sum_{i=1}^{N} w_i x_i^{(n)}.$$

Sedangkan $\hat{o}^{(n)}$ adalah output pengamatan ke-n pada model (3) adalah

$$\hat{o}^{(n)} = \Lambda(z^{(n)}).$$

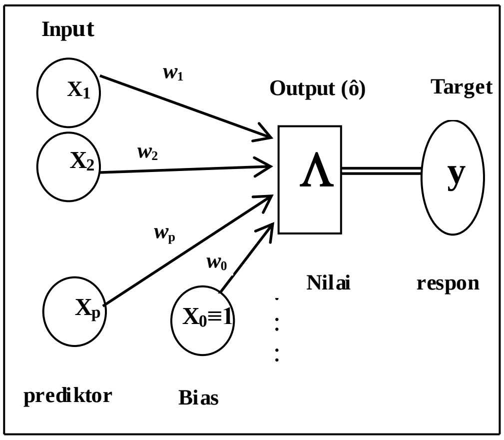
Bentuk kanonik gradien J(w) terhadap z adalah

$$\frac{\partial J}{\partial \mathbf{w}} = \sum_{i=1}^{N} (y^{(i)} - \hat{o}^{(i)}) \frac{\partial z^{(i)}}{\partial \mathbf{w}}$$

Permasalahan pada training adalah minimisasi fungsi tujuan yang tergantung pada vektor parameter adaptif w. Aspek penting dari masalah ini adalah vektor gradien $\nabla_{\mathbf{w}} J$ dapat dievaluasi secara efisien. Minimisasi berdasarkan nilai gradien adalah masalah standar dalam optimasi nonlinear, dimana banyak teknik yang sangat kuasa. Secara umum algoritma ini diawali dengan menentukan nilai awal sembarang untuk parameter w , kemudian secara iteratif nilai w diperbarui dengan

$$\mathbf{w}^{(r+1)} = \mathbf{w}^{(r)} + \Delta\mathbf{w}^{(r)}$$

dengan r menunjukkan nomor iterasi. Setiap algoritma yang berbeda, berbeda pula dalam memperbarui $\Delta\mathbf{w}^{(r)}$. Metode yang terkenal antara lain conjugate gradient, quasi-newton, dan untuk kasus jumlah kuadrat galat, algoritma Levenberg-Marquardt sangat efektif.



Gambar 1. Arsitetektur Perceptron Gamma

Pada model linear terampat dengan komponen acak mempunyai sebaran gamma ekivalen dengan JST tanpa lapisan tersembunyi dengan fungsi galat negatif log kemungkinan pada persamaan (1) atau devians pada persamaan (2).

METODE EVALUASI

Data yang digunakan adalah mengenai waktu pembekuan darah dari McCullagh, and Nelder (1989 halaman 300). Waktu pembekuan darah dalam detik (y) untuk plasma normal diukur pada 9 macam persentase konsentrasi prothrombin-free plasma (u) dan 2 jenis thromboplastin (lot).

Untuk mencapai kelinearan dilakukan transformasi logaritma pada konsentrasi prothrombin-free plasma (ln(u)) Data dianalisa dengan menggunakan GLIM dan JST dengan fungsi hubung (aktifasi) resiprokal dan galat berdistribusi gamma.

HASIL DAN PEMBAHASAN

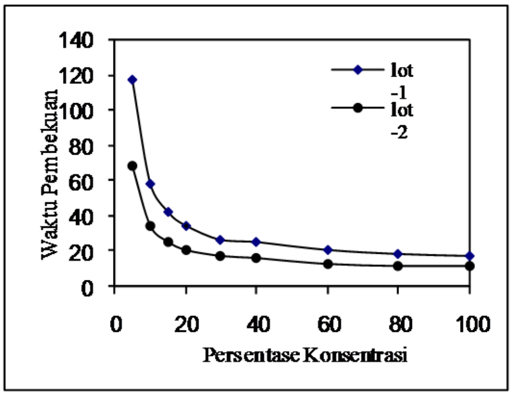
Plot data antara y dengan konsentrasi tertera pada Gambar 2. Dari plot tampak waktu pembekuan lot1 lebih tinggi dibandingkan lot2 dengan koefisien arah yang tidak sama. Model yang sesuai untuk data ini adalah $\mu^{-1} = \text{lot} + \text{lot} * \ln(\text{u})$ .Dengan fungsi galat menyebar gamma dan fungsi hubung resiprokal, nilai penduga parameter dari JST dan GLIM tertera pada Tabel 2. Kedua metode menghasilkan dugaan model yang sama, yaitu :

Lot-1:

$$\mu^{-1} = 0.03677 - 0.020231 + 0.15343 * \ln(\text{u})$$
$$= 0.016554 + 0.15343 * \ln(\text{u})$$

Lot-2:

$$\mu^{-1} = -0.03677 - 0.020231 + 0.023599 * \ln(\text{u})$$
$$= 0.23908 + 0.023599 * \ln(\text{u})$$



Gambar 2. Plot antara Waktu Pembekuan denganPersentase Konsentrasi

Tabel 2. Nilai Penduga Parameter dari JST dan GLIM

Parameter	JST	GLIM
Lot 0	0.003677	0.0074
Lot 1	-0.003677	0.0000
Lot 0 * ln(u)	0.015343	0.0015
Lot 1 * ln(u)	0.023599	0.0236
Bias	-	-
	0.020231	0.0239
Fungsi tujuan : Log likelihood	1.4776439057	-26.5976
Fungsi tujuan : Deviance	0.0016334151	0.0294

Berdasarkan hasil pada Tabel 2, pada JST nilai fungsi tujuan kemungkinan maksimum sebesar 1.4776439057, bila nilai ini dikalikan dengan banyaknya pengamatan yaitu 18, maka akan diperoleh 26.59757 sama dengan negatif log likelihood pada GLIM. Demikian juga pada fungsi tujuan devians, bila nilainya dikalikan dengan banyaknya pengamatan akan diperoleh 0.02940 sama dengan nilai devians dari GLIM.

KESIMPULAN

Model gamma adalah salah satu bentuk khusus dari jaringan syaraf tiruan dengan bentuk arsitekturnya disebut perceptron. Fungsi aktifasi untuk model gamma adalah resiprokal, identitas, atau eksponensial, dan fungsi tujuannya dapat kemungkinan maksimum atau devians. Ditinjau dari kerangka pemodelan statistik, GLIM memberikan tata cara yang baku daripada JST terutama pada tahapan pemilihan parameter, peubah penjelas, pengujian signifikansi parameter dan pemilihan model terbaik.

DAFTAR PUSTAKA

Cheng, B. & Titterington, D.M. 1994. Neural Network : A review from a Statistical Prespective. Statistical science 9 : 2-54.

Jordan MI, Bishop CM. 1996. Neural Network. AI Memo No. 152, C,B.C.L. Memo No 131. Massachusetts Institute of Technology.

McCullagh P, Nelder JA. 1989. Generalized Linear Models. Chapman and Hall. New York.

Sarle, WS. 1994. Neural Network and Statistical Models. Proceddings of the 19th Anual SAS User Group International Conference.

Sarle, WS. 1996. Neural Network and Statistical Jargon. [terhubung berkala] ftp://ftp.sas.com/pub/neural/jargon.

Warner B, Misra M. 1996. Understanding Neural Network as Statistical Tools. American Statistician 50 : 284-293