PENDUGAAN PARAMETER REGRESI LOGISTIK DENGAN JARINGAN SYARAF TIRUAN

Anik Djuraidah
Jurusan Statistika
FMIPA Institut Teknologi Sepuluh Nopember

Abstract

Logistic regression can be mapped equivalent to artificial neural network (ANN) without hidden layer with logistic as activation function, hence logistic regression is subset of ANN. The result study on binary and poly-chotomous response data show that parameter estimation values of ANN and logistic regression are similar. In comparison with ANN, logistic regression has standard procedure for estimation and testing parameter.

***Keyword** : logistic regression, generalized linear model, artificial neural network, activation function, hidden layer*

1. PENDAHULUAN

Jaringan saraf tiruan (*artificial neural network* selanjutnya disingkat JST) di dalam statistika merupakan suatu kelas yang luas dari model regresi nonlinear, model reduksi data, dan sistem dinamik nonlinear. Seperti pada metode statistika, JST mampu memproses berbagai macam data dan membuat prediksi yang kadang kala sangat akurat. 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. 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* selanjutnya disingkat GLIM) merupakan perluasan dari model linear klasik. Pada GLIM nilaitengah populasi tergantung pada prediktor linear melalui suatu fungsi hubung *(link function*) dan sebaran peluang peubah respons termasuk keluarga eksponensial (*exponential family*). Secara umum GLIM terdiri dari tiga komponen (McCullagh dan Nelder, 1989, Warner dan Misra, 1996) yaitu :

1. Komponen sistematik atau komponen linear

$$\eta_{\mathrm{i}} = \mathbf{x}_{\mathrm{i}}'\boldsymbol{\beta}$$

2. Fungsi hubung g yang menjelaskan hubungan antara nilai harapan y_{i} dengan prediktor linear

$$g(\mu_{\mathrm{i}}) = \mathbf{x}_{\mathrm{i}}'\boldsymbol{\beta}$$

3. Komponen acak dari peubah respon y_{i} mempunyai sebaran keluarga eksponensial dengan rataan μ dan ragam σ^2

Pada regresi logistik, hubungan antara peubah respon y yang bernilai biner dengan prediktor $\mathbf{x} = (x_1, \;\cdots, \; x_{\mathrm{I}})$ dimodelkan sebagai berikut :

$$\mathrm{P}(y = 1|\mathbf{x}) = \Lambda(\beta_{\mathrm{o}} + \sum_{\mathrm{i}=1}^{\mathrm{I}} \beta_{\mathrm{i}}\, x_{\mathrm{i}}) \tag{1}$$

dengan $\Lambda(u) = \dfrac{1}{(1+e^{-u})}$ yang menyatakan fungsi logistik. Parameter $\boldsymbol{\beta}$ diduga dari data dan diinterpretasikan sebagai log-odds rasio. Regresi logistik adalah regresi nonlinear yang merupakan kasus khusus GLIM dengan

$$\mathrm{E}(y|\mathbf{x}) = \pi(\mathbf{x}, \boldsymbol{\beta}),$$

$$\mathrm{Var}(y|\mathbf{x}) = \pi(\mathbf{x}, \boldsymbol{\beta})\;(1 - \pi(\mathbf{x}, \boldsymbol{\beta}))$$

dan fungsi hubung logistik

$$\pi(\mathbf{x}, \boldsymbol{\beta}) \;=\; \Lambda(\beta_0 + \sum_{\mathrm{i}=1}^{\mathrm{p}} \beta_{\mathrm{i}}\, \mathbf{x}_{\mathrm{i}}) = \;\Lambda(\boldsymbol{\beta}'\,\mathbf{x})$$

dengan $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)$ dan menambahkan $x_0 \equiv 1$ pada vektor **x**.

Parameter $\boldsymbol{\beta}$ diduga dengan menggunakan metode kemungkinan maksimum (*maximum likelihood* selanjutnya disingkat ML) yaitu dengan maksimisasi fungsi log-kemungkinan

$$L(\boldsymbol{\beta}) = \sum_{n=1}^{N} y^{(n)} \log \pi(\mathbf{x}^{(n)}, \boldsymbol{\beta}) + (1 - y^{(n)}) \log(1 - \pi(\mathbf{x}^{(n)}, \boldsymbol{\beta}))$$

dengan $\left(\mathbf{x}^{(n)}, \quad y^{(n)}\right)$ menyatakan data pengamatan individu ke-n, n = 1, ..., N (McCullagh dan Nelder, 1989).

Pada penelitian ini akan dikaji tentang arsitektur JST untuk regresi logistik dan hasil pendugaan parameternya pada beberapa data. Di samping itu akan dibandingkan seberapa besar perbedaan nilai dugaan parameter model antara JST dengan GLIM.

2. JARINGAN SYARAF TIRUAN UNTUK REGRESI LOGISTIK

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).

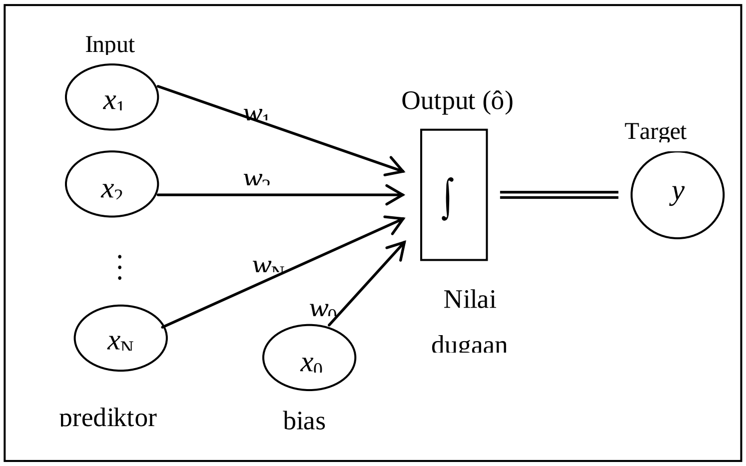
Pada JST terdapat dua proses yaitu proses *training* dan *testing*. Proses *training* adalah proses pembelajaran JST yang mengatur input-input yang digunakan dan pemetaannya pada output hingga diperoleh model JST, sehingga pada proses ini terjadi pengaturan bobot dan bias. Dalam statistika proses *training* tidak lain adalah pendugaan parameter. Sedangkan proses *testing* adalah proses pengujian ketelitian model yang diperoleh dari proses *training*. Proses *testing* pada JST identik dengan validasi model pada statistika (Cheng dan Titterington, 1994 ; Sarle, 1996).

Proses *training* pada JST terbagi dalam tiga tahap utama yaitu : *feed-forward*, *backpropagation*, 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 sisaan. Pada tahap *update* dilakukan dilakukan pembaruan nilai bobot sampai diperoleh nilai fungsi sisaan minimal.

Bentuk *feed-forward* JST yang paling sederhana disebut *perceptron* logistik tertera pada Gambar 1. Perceptron mempunyai p+1 unit input dengan x_0 adalah input konstan bernilai 1 dan satu unit output. Nilai input x_i diboboti dengan pembobot w_i dan jumlah input-input terboboti ini kemudian ditransformasi dengan fungsi logistik. Nilai output $\hat{\mathrm{O}}$ dapat dinyatakan sebagai sebuah fungsi f dari nilai input $\mathbf{x}$ dan bobot w_i yaitu :

$$\hat{\mathrm{O}} \;=\; f(\mathbf{x},\mathbf{w}) = \;\Lambda(w_0 + \sum_{\mathrm{i}=1}^{\mathrm{p}} w_\mathrm{i} x_\mathrm{i}) \qquad (2)$$

dengan $\mathbf{x} = (x_1, \;\cdots,\; x_\mathrm{p})$ adalah vektor input, $\mathbf{w} = (w_1, \;\cdots,\; w_\mathrm{p})$ adalah vektor pembobot, dan $\Lambda(\mathbf{x})$ adalah fungsi aktifasi seperti pada persamaan (1).



Gambar 1. Arsitetektur *Perceptron* Logistik

Pada proses *training* dilakukan pendugaan terhadap bobot $\mathbf{w}$ berdasarkan pengamatan $\left(\mathbf{x}^{(1)}, y^{(1)}\right) \; \cdots \; \left(\mathbf{x}^{(N)}, y^{(N)}\right)$ dengan meminimumkan fungsi tujuan (*objective function*). Bentuk fungsi tujuan tergantung bentuk fungsi sisaan (*error/cost function*). Fungsi sisaan yang paling umum adalah jumlah kuadrat sisaan, tetapi bisa juga fungsi kemungkinan (Warner dan Misra, 1996). 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

$$\mathrm{p}(\mathbf{w} \mid \mathcal{D}) = \frac{\mathrm{p}(\mathcal{D} \mid \mathbf{w})\, p(\mathbf{w})}{\mathrm{p}(\mathcal{D})} \qquad (3)$$

Memaksimumkan persamaan (3) sama dengan memaksimumkan suku pembilangnya karena penyebut tidak tergantung $\mathbf{w}$. Sehingga maksimisasi persamaan (3) ekivalen dengan minimisasi dari negatif logaritma suku pembilangnya. Fungsi ini dikenal dengan fungsi sisaan kemungkinan, yaitu

$$\mathrm{J}(\mathbf{w}) = -\ln \mathrm{p}(\mathcal{D} \mid \mathbf{w}) - \ln \mathrm{p}(\mathbf{w}) \qquad (4)$$

Jika pengamatan pada data D menyebar bebas stokastik dan identik maka

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

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

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

Bila diasumsikan peluang awal dari parameter seragam untuk semua parameter maka suku kedua pada persamaan (4) dapat dihilangkan. Hal ini berati minimisasai fungsi sisaan sama dengan maksimisasi fungsi log kemungkinan yaitu memilih $\mathbf{w}_{ML}$ yang memaksimumkan $\mathrm{p}(\mathbf{w} \mid \mathcal{D})$.

Setelah menentukan model probabilistik, fungsi sisaan, masalah selanjutnya adalah menemukan prosedur yang efisien untuk perhitungan gradien dari fungsi sisaan. Misalkan $z_i^{(n)}$ adalah nilai output dari lapisan ke-i pada pengamatan ke-n, untuk model (2) karena hanya satu lapisan maka

$$z^{(n)} = w_0 + \sum_{i=1}^{N} \mathbf{w}_i \mathbf{x}_i^{(n)}$$

dan $\hat{\mathrm{O}}^{(n)}$ adalah output pengamatan ke-n, maka pada model (2) adalah

$$\hat{\mathrm{O}}^{(n)} = \Lambda(z^{(n)})$$

Bentuk kanonik gradien $\mathrm{J}(\mathbf{w})$ terhadap z adalah

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

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

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

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

3. *PERCEPTRON* LOGISTIK PADA K UNIT OUTPUT

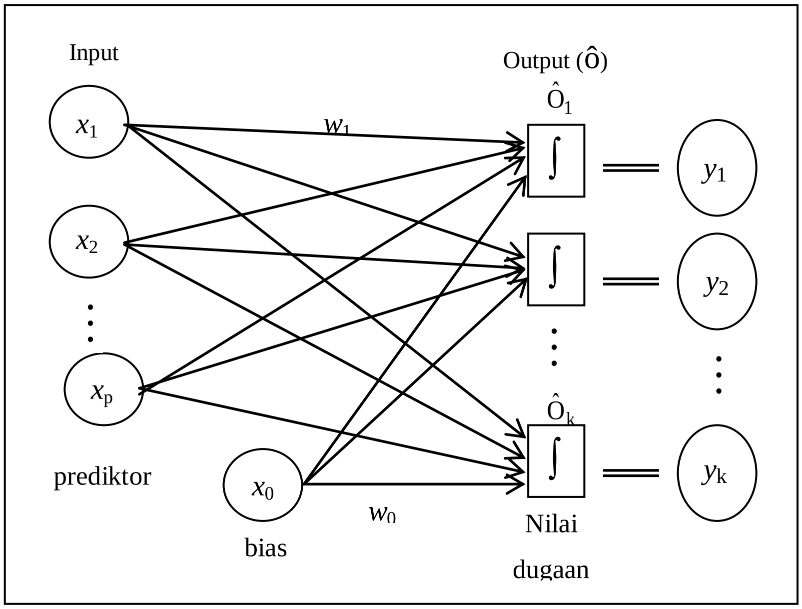
Perceptron logistik dengan satu unit output dapat diperluas untuk K>1 unit output. JST untuk K unit output tertera pada Gambar 2 yang dicirikan dengan bobot w_{ik} (i = 1, ... , I; k = 1, ... , K) terhadap nilai output $\hat{O}_k$ yaitu

$$\hat{O}_k \quad = \quad f(\mathbf{x}, \mathbf{w}_k) = \quad \Lambda(w_{0k} + \sum_{i=1}^{I} w_{ik} x_{ik})$$

Dalam statistika model ini ekivalen dengan model regresi logistk multinomial atau ordinal dengan peubah respon y bernilai {1, 2, ... , k). Model untuk respon y = k didefinisikan sebagai

$$P(y = k | \mathbf{x}) = f(\mathbf{x}, \mathbf{w}_k) / \sum_{k=1}^{K} f(\mathbf{x}, \mathbf{w}_k)$$

Bobot-bobot $\mathbf{w}_k$ dapat diinterpretasikan sebagai koefisien regresi logistik (Schumacher *et al*, 1996).



Gambar 2 Arsitektur Perceptron Logistik dengan K Unit Output

4. DATA DAN METODE EVALUASI

Untuk mengetahui hasil pendugaan model dengan JST dan GLIM digunakan data di bawah ini :

1. DATA-1 adalah data dengan respon biner dari Hosmer dan Lemeshow (1989) dengan peubah respon berat badan bayi (1 = kurang 2500 gram, 0 = lebih 2500 gram) dengan peubah penjelasnya adalah :

Peubah (Label)	***Kategori***
Ras (RACE)	1 = Putih, 2 = Hitam , 3 = Lainnya
Perokok (SMOKE)	1 = Ya, 0 = Tidak
Hipertensi (HT)	1 = Ya, 0 = Tidak
Riwayat Prematur (PTL)	1 = Pernah, 0 = Tidak
Uterine Irritability (UI)	1 = Ya, 0 = Tidak
Kunjungan ke dokter (FTV)	1 = Tidak pernah, 1 = pernah
Umur Ibu (AGE)	Kontinu
Berat Badan Ibu terakhir (LWT)	Kontinu

2. DATA-2 adalah data dengan respon ordinal yang mempunyai 3 kategori diambil dari McCullagh dan Nelder (1989) pada Tabel 5.2 dengan peubah respon tingkat serangan penyakit paru (1 = normal/ tidak ada, 2 = sedikit, 3 = banyak) dan peubah penjelasnya adalah logaritma lama bekerja di pertambangan (lnTIME).
3. DATA-3 adalah data dengan respon ordinal yang mempunyai 9 kategori diambil dari McCullagh dan Nelder (1989) pada Tabel 5.1 dengan peubah respon rasa (1 = sangat tidak disukai sampai 9 = sangat disukai) dan peubah penjelasnya jenis bahan aditif pada keju (A, B, C. D).

Pada ketiga data ini, pendugaan parameter modelnya dilakukan dengan metode JST dan GLIM dengan bantuan perangkat lunak SAS.

5. HASIL DAN PEMBAHASAN

Nilai dugaan parameter DATA-1, DATA-2, dan DATA-3 dengan JST dan GLIM masing-masing disarikan pada Tabel 2, Tabel 3, dan Tabel 4. Pada Tabel 2 dan Tabel 3 tampak nilai dugaan parameter antara metode JST dengan GLIM hampir sama.

Tabel 2. Nilai Dugaan Parameter DATA-1

Parameter	JST	GLIM
RACE 1 vs 3	0.7765	0.7765
RACE 2 vs 3	-0.4259	-0.4259
SMOKE 0 vs 1	0.8074	0.8074
HT 0 vs 1	1.8203	1.8204
PTL 0 vs 1	1.2462	1.2463
UI 0 vs 1	0.7052	0.7052
FTV 0 vs 1	-0.1283	-0.1284
AGE	0.0368	0.0368
LWT	0.0147	0.0147
BIAS	-5.8935	-5.8940

Tabel 3. Nilai Dugaan Parameter DATA-2

Parameter	JST	Parameter	GLIM
BIAS 1	9.6755	INTERCEPT 1	9.6761
BIAS 2	10.5811	INTERCEPT 2	10.5817
lnTIME	-2.5966	lnTIME	-2.5968

Berdasarkan Tabel 3, maka penduga respon dari metode JST untuk masing-masing kategori adalah :

$$P(y=1|\ \boldsymbol{x}) = \frac{1}{\exp-(9.675508-2.596641\ln \mathrm{TIME})}$$

$$P(y\leq 2|\ \boldsymbol{x}) = \frac{1}{\exp-(10.581136-2.596641\ln \mathrm{TIME})}$$

$$P(y=3|\ \boldsymbol{x}) = 1-P(y\leq 2|\ \boldsymbol{x})$$

Sedangkan penduga respon metode GLIM dapat diperoleh dengan cara mengganti koefisien yang sesuai pada Tabel 3 .

Tabel 4. Nilai Dugaan Parameter DATA-3

Parameter	JST	Parameter	GLIM
CHEESE A	-0.8623	CHEESE A	1.6128
CHEESE B	2.4896	CHEESE B	4.9646
CHEESE C	0.8477	CHEESE C	3.3227
CHEESE D	-2.4750	CHEESE D	0.0000
BIAS_TASTE 1	-4.6051	INTERCEPT 1	-7.0802
BIAS_TASTE 2	-3.5499	INTERCEPT 2	-6.0250
BIAS_TASTE 3	-2.4504	INTERCEPT 3	-4.9254
BIAS_TASTE 4	-1.3818	INTERCEPT 4	-3.8568
BIAS_TASTE 5	-0.0455	INTERCEPT 5	-2.5206
BIAS_TASTE 6	0.9065	INTERCEPT 6	-1.5685
BIAS_TASTE 7	2.4081	INTERCEPT 7	-0.0669
BIAS_TASTE 8	3.9680	INTERCEPT 8	1.4930

Berdasarkan Tabel 4, maka penduga respon dari metode JST untuk jenis keju A adalah

$$P(y=1 \mid \text{CHEESE A}) = \frac{1}{\exp-(-0.8623-4.6051)} = \frac{1}{\exp-(-5.4674)}$$

$$P(y\leq 2 \mid \text{CHEESE A}) = \frac{1}{\exp-(-0.8623-3.5499)} = \frac{1}{\exp-(-4.4122)}$$

Sedangkan GLIM memberikan penduga respon untuk keju A adalah :

$$P(y=1 \mid \text{CHEESE A}) = \frac{1}{\exp-(1.6128-7.0802)} = \frac{1}{\exp-(-5.4674)}$$

$$P(y\leq 2 \mid \text{CHEESE A}) = \frac{1}{\exp-(1.6128-6.0250)} = \frac{1}{\exp-(-4.4122)}$$

Demikian juga untuk respon kelas rasa dan jenis keju lainnya bisa dilakukan dengan cara mengganti koefisien yang sesuai. Nilai dugaan parameter dari kedua metode tersebut sama.

6. KESIMPULAN

Model regresi logistik adalah salah satu bentuk dari JST, yaitu dengan arsitektur tanpa lapisan tersembunyi dan fungsi aktivasi logistik. Dengan demikian model regresi logistik merupakan subset dari JST. Pendugaan parameter regresi logistik dengan menggunakan metode JST dan GLIM menghasilkan nilai dugaan yang hampir sama. Ditinjau dari kerangka pemodelan statistik, GLIM memberikan tata cara yang baku dari pada 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.

Hosmer, D.W. & Lemeshow, S. (1989). *Applied Logistic Regression.* New York : John Wiley and Sons.

Jordan, M.I. & Bishop, C.M. (1996). Neural Network. *AI Memo No. 152, C,B.C.L. Memo No 131. Massachusetts Institute of Technology.*

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models.* New York: Chapman and Hall.

Sarle, W.S. (1994). Neural Network and Statistical Models. *Proceedings of the 19th Anual SAS User Group International Conference*.

Sarle, W.S. (1996). *Neural Network and Statistical Jargon*. Diambil tanggal 16 Juli 2007 dari ftp://ftp.sas.com/pub/neural/jargon.

Schumacher, M., Roβner R., & Vach W. (1996). Neural Network and logistic regression : Part I. *J. Computational Statistics and Data Analysis* 21:661-682.

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