BAB III

PENDEKATAN BAYESIAN PADA REGRESI LINEAR

3.1 Pendekatan Bayesian Pada Regresi Linear

Pada model linear harus memenuhi asumsi normalitas yakni variabel acak y dan X berdistribusi normal sehingga parameter θ juga berdistribusi normal. Model linear berbentuk y = Xθ + ε. Diberikan parameter θ' = (θ_1, …, θ_n) vektor acak berukuran p x 1, matriks X berukuran n x p diambil dari observasi x_1, …, x_n, vektor acak y berukuran n x 1 dan variabel acak bebas ε' = (ε_1, …, ε_n) yang berdistribusi normal, $N(0,\sigma^2)$.

Dalam mencari distribusi prior diperlukan fungsi *likelihood* dari parameter θ. Diketahui vektor $X\theta$ = E($y|\theta$) sebagai nilai rata-rata dan matriks kovarians D($y|\sigma^2$) = $\sigma^2 P^{-1}$ di mana P merupakan matriks definit positif.

Misalkan ($y|\theta,\ \sigma^2$) ~ N($X\theta,\ \sigma^2 P^{-1}$) dan σ^2 merupakan parameter yang tidak diketahui. Didefinisikan $\tau = \sigma^{-2}$. Karena ($y|\theta,\ \sigma^2$) ~ N($X\theta,\ \sigma^2 P^{-1}$) diperoleh $(\det \tau^{-1}P^{-1})^{-\frac{1}{2}} = (\det P)^{\frac{1}{2}}\tau^{\frac{n}{2}}$

Fungsi *likelihood*

$$p(y|\theta,\tau,C) = (2\pi)^{-\frac{n}{2}}(\det P)^{\frac{1}{2}}\tau^{\frac{n}{2}}\exp\left(-\frac{\tau}{2}(y-X\theta)'P(y-X\theta)\right)$$

$$\propto\ \tau^{\frac{n}{2}}\exp\left(-\frac{\tau}{2}(y-X\theta)'P(y-X\theta)\right) \qquad (3.1)$$

Distribusi prior untuk θ dan τ adalah distribusi normal-gamma

$\theta,\tau \sim \mathrm{NG}\ (\mu, \mathrm{V}, \mathrm{a}, \mathrm{b})$

dimana $\theta|\tau \sim \mathrm{N}\ (\mu, \mathrm{V})$ dan $\tau \sim \mathrm{G}\ (\mathrm{a}, \mathrm{b})$

Fungsi kepadatan peluang dari distribusi prior adalah :

$$p(\theta,\tau|C) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det V)^{\frac{1}{2}}}\frac{\mathrm{a}^b}{\Gamma(b)}\tau^{\frac{n}{2}+b-1}\ exp\left(-\frac{\tau}{2}\left(2\mathrm{a}+(\theta-\mu)'V^{-1}(\theta-\mu)\right)\right)$$

$$\propto\ \tau^{\frac{n}{2}+b-1}\ exp\left(-\frac{\tau}{2}(2\mathrm{a}+(\theta-\mu)'V^{-1}(\theta-\mu)\right) \qquad (3.2)$$

Selanjutnya akan dicari distribusi posterior dari θ dan τ dengan mengoperasikan distribusi prior (3.2) dengan fungsi *likelihood* (3.1),

$p(\theta,\tau|y,C) \propto p(\theta,\tau|C)\ x\ p(y|\theta,\tau,C)$

$$\propto \tau^{\frac{u}{2}+b-1} exp\left(-\frac{\tau}{2}(2\text{a} + (\theta-\mu)'V^{-1}(\theta-\mu))\right).$$

$$\tau^{\frac{n}{2}}\exp\left(-\frac{\tau}{2}(y-X\theta)'P(y-X\theta)\right)$$

$$\propto \tau^{\frac{n}{2}+\frac{u}{2}+b-1} exp\left(-\frac{\tau}{2}\Big(\mathbf{2}\text{a} + (\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{\theta}-\boldsymbol{\mu}\,) + \right.$$

$$\left.(\boldsymbol{y}-\boldsymbol{X\theta})'\boldsymbol{P}(\boldsymbol{y}-\boldsymbol{X\theta})\Big)\right) \quad (3.3)$$

Akan diuraikan $(\mathbf{2}\text{a} + (\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{\theta}-\boldsymbol{\mu}) + (\boldsymbol{y}-\boldsymbol{X\theta})'\boldsymbol{P}(\boldsymbol{y}-\boldsymbol{X\theta})$

$$(\mathbf{2}\text{a} + (\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{\theta}-\boldsymbol{\mu}) + (\boldsymbol{y}-\boldsymbol{X\theta})'\boldsymbol{P}(\boldsymbol{y}-\boldsymbol{X\theta})$$

$$\Leftrightarrow\ 2\text{a} + \theta'V^{-1}\theta - \theta'V^{-1}\mu - \mu'V^{-1}\theta + \mu'V^{-1}\mu + y'Py - y'PX\theta$$
$$-\theta'X'Py + \theta'X'PX\theta$$

$$\Leftrightarrow\ 2\text{a} + y'Py + \mu'V^{-1}\mu - 2\theta'(X'Py + V^{-1}\mu) + \theta'(X'PX + V^{-1})\theta$$

$$\Leftrightarrow\ 2\text{a} + y'Py + \mu'V^{-1}\mu - \mu_0'(X'PX + V^{-1})\mu_0$$
$$+(\theta-\mu_0)'(X'PX + V^{-1})(\theta-\mu_0)$$

$$\Leftrightarrow\ 2\text{a} + y'Py + \mu'V^{-1}\mu - 2\mu_0'(X'Py + V^{-1}\mu) + \mu_0'(X'PX + V^{-1})\mu_0$$
$$+(\theta-\mu_0)'(X'PX + V^{-1})(\theta-\mu_0)$$

$$\Leftrightarrow\ 2\text{a} + y'Py + \mu'V^{-1}\mu - 2\mu_0'X'Py - 2\mu_0'V^{-1}\mu + \mu_0'X'PX\mu_0 + \mu_0'V^{-1}\mu_0$$
$$+(\theta-\mu_0)'(X'PX + V^{-1})(\theta-\mu_0)$$

$$\Leftrightarrow\ 2\text{a} + (\mu-\mu_0)'V^{-1}(\mu-\mu_0) + (y-X\mu_0)'P(y-X\mu_0)$$
$$+(\theta-\mu_0)'(X'PX + V^{-1})(\theta-\mu_0) \quad (3.4)$$

Substitusikan Persamaan (3.4) kedalam Persamaan (3.3) sehingga diperoleh

$$\propto \tau^{\frac{n}{2}+\frac{u}{2}+b-1} exp\left(-\frac{\tau}{2}[2\text{a} + (\mu-\mu_0)'V^{-1}(\mu-\mu_0) + (y-X\mu_0)'P(y-X\mu_0)\right.$$
$$\left. + (\theta-\mu_0)'(X'PX + V^{-1})(\theta-\mu_0)]\right)$$

$$\propto \tau^{\frac{n}{2}} exp\left\{-\frac{\tau}{2}[(\theta-\mu_0)'(X'PX + V^{-1})(\theta-\mu_0)]\right\}$$

$$\tau^{\frac{u}{2}+b-1}\exp\left\{-\frac{\tau}{2}[2\text{a} + (\mu-\mu_0)'V^{-1}(\mu-\mu_0)\right.$$
$$\left. + (y-X\mu_0)'P(y-X\mu_0)]\right\} \quad (3.5)$$

Distribusi posterior untuk θ dan τ pada (3.5) merupakan distribusi normal-gamma.

$\theta, \tau \sim \text{NG}\ (\mu_0, V_0, \text{a}_0, b_0)$

dimana $\theta|\tau \sim \text{N}\ (\mu_0, V_0)$ dan $\tau \sim \text{G}\ (\text{a}_0, b_0)$

dengan

$$\mu_0 = (X'PX + V^{-1})^{-1}(X'Py + V^{-1}\mu)$$
$$V_0 = (X'PX + V^{-1})^{-1}$$
$$a_0 = a + \frac{1}{2}[(\mu - \mu_0)'V^{-1}(\mu - \mu_0) + (y - X\mu_0)'P(y - X\mu_0)]$$
$$b_0 = \frac{1}{2}n + b$$

Selanjutnya distribusi posterior dari ε dapat diturunkan dengan bentuk $\varepsilon = y - Xe$ dan diketahui bahwa ε merupakan fungsi linear dari parameter θ. De Groot (2004) menjelaskan bahwa distribusi ε akan singular hanya pada ruang p-dimensi.

Misalkan $H = X(R + X'X)^{-1}X$, maka distribusi posterior dari ε yang bersyarat τ akan menghasilkan distribusi normal multivariat singular dengan rata-rata $\hat{\varepsilon} = y - Xm_1$ dan matrik kovarians $\tau^{-1}H$.

Dengan memisalkan h_{ij} merupakan elemen-elemen dari H, maka setiap ε_i(untuk i = 1, 2, …, n) memiliki distribusi t dengan lokasi $\hat{\varepsilon}_i$, ketelitian $\frac{a_1}{b_1 h_{ii}}$ dan $2a_1$ sebagai derajat kebebasan. (DeGroot, 2004, hlm. 42). Matrik kovarians dari ε merupakan proporsi dari H.

Distribusi prior improper dari Zellner (1996) $p(\theta, \tau|C) = \tau^{-1}$, misalkan $R \to 0$, $a \to -\frac{1}{2}p$ dan $b \to 0$. Misalkan

$$\hat{\theta} = (X'X)^{-1}X'y \ , s^2 = \frac{(y - X\hat{\theta})'(y - X\hat{\theta})}{(n-p)}$$

Selanjutnya DeGroot (2004, hlm. 252) menunjukkan bahwa distribusi posterior dari ε merupakan distribusi t multivariate pada ruang p-dimensi dengan

$$\hat{\varepsilon} = y - X\hat{\theta} \, , H = X(X'X)^{-1}X \, , \qquad a_1 = \frac{1}{2}(n-p), \qquad b_1 = \frac{1}{2}(n-p)s^2$$

Untuk mendeteksi observasi yang merupakan pencilan, didefinisikan peluang p_i menjadi $pr(|\varepsilon_i| > k\sigma|y)$, peluang posterior pada observasi ke-i yang didefinisikan sebagai sebuah pencilan.

Misalkan $\Phi(z)$ merupakan fungsi distribusi normal standar dengan

$$z_1 = \frac{(k - \hat{\varepsilon}_i\sqrt{\tau}\,)}{\sqrt{h_{ii}}} \ , \qquad z_2 = \frac{(-k - \hat{\varepsilon}_i\sqrt{\tau})}{\sqrt{h_{ii}}} \quad (3.6)$$

Selanjutnya diperoleh

$$p_i = pr(|\varepsilon_i| > k\sigma|y) = \int \{1 - \Phi(z_1) + \Phi(z_2)\}\, p(\tau|y)d\tau \quad (3.7)$$

Hasil dari p_i dapat dibandingkan dengan peluang prior $2\Phi(-k)$. Nilai dengan peluang posterior p_i yang tinggi yang menjadi pencilan akan memiliki $|\hat{\varepsilon}_i|$ yang besar atau h_{ii} yang besar ataupun keduanya. Ketika $|\hat{\varepsilon}_i|$ membesar akibatnya $|\varepsilon_i|$ juga ikut membesar, namun ketika h_{ii} membesar ada ketidakpastian untuk ε_i. Banyaknya h_{ii} sering menunjukkan sebagai *leverage*.

3.2 Penentuan Nilai k

Nilai k dapat dipilih sehingga peluang prior yang tidak memiliki pencilan besar, misalkan 0,95. Ini memberikan $k = \Phi^{-1}\{0,5 + \frac{1}{2}(0,95^{1/n})\}$, dengan mensubstitusikan nilai n dan melihat Daftar Distribusi Normal Baku (lampiran 1).

Cara lain yang dapat digunakan apabila dalam pertimbangannya model diperlukan untuk menggambarkan data bukan terhadap model yang stokastik, maka nilai k = 2 dapat digunakan untuk menemukan observasi yang tidak baik dijelaskan oleh data maupun ukuran sampel.

Dalam Chaloner (1988), observasi dinamakan sebagai pencilan ketika peluang posterior (3.7) lebih besar dari peluang prior $2\Phi(-k)$.