BAB II

LANDASAN TEORI

Bab ini membahas beberapa pengertian dasar yang diperlukan pada pembahasan bab-bab berikutnya, yaitu mengenai teorema bayes, metode maksimum likelihood, algoritma EM (*Expectation-Maximization*), dan uji rasio likelihood.

2.1 Teorema Bayes

Misalkan $B_1, B_2, \ldots, B_T$ merupakan partisi dari ruang sampel S, $B_i \neq \varnothing, i = 1, 2, \ldots, T$, yang bersifat:

- $B_1 \cup B_2 \cup \ldots \cup B_T = S$
- $B_i \cap B_j = \varnothing$, untuk $i \neq j$.

Misalkan A adalah sembarang kejadian yang merupakan himpunan bagian S, yang bersifat $P(A) \neq 0$. Kejadian A dapat dipandang sebagai gabungan kejadian-kejadian $B_1 \cap A, B_2 \cap A, \ldots, B_T \cap A$ yang saling terpisah satu sama lain.

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \ldots \cup (B_T \cap A) \qquad (2.1.1)$$

Karena $B_1 \cap A, B_2 \cap A, \ldots, B_T \cap A$ merupakan himpunan-himpunan yang saling lepas, maka probabilitas kejadian A dapat dituliskan sebagai berikut:

$$\begin{aligned} P(A) &= P\left[(B_1 \cap A) \cup (B_2 \cap A) \cup \ldots \cup (B_T \cap A)\right] \\ &= P(B_1 \cap A) + P(B_2 \cap A) + \cdots + P(B_T \cap A) \\ &= P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + \cdots + P(B_T)P(A \mid B_T) \\ &= \sum_{i=1}^{T} P(B_i)P(A \mid B_i) \end{aligned} \tag{2.1.2}$$

Kemudian, menurut definisi peluang bersyarat diketahui bahwa:

$$\begin{aligned} P(B_r \mid A) &= \frac{P(B_r \cap A)}{P(A)} \\ &= \frac{P(B_r)P(A \mid B_r)}{P(A)} \end{aligned} \tag{2.1.3}$$

Dengan mensubstitusikan persamaan (2.1.2) ke persamaan (2.1.3), didapatkan:

$$P(B_r \mid A) = \frac{P(B_r)P(A \mid B_r)}{\sum_{i=1}^{T} P(B_i)P(A \mid B_i)} \tag{2.1.4}$$

Jadi, dapat dinyatakan jika terdapat kejadian-kejadian $B_1, B_2, \ldots, B_T$ merupakan partisi dari ruang sample S dengan $B_i \neq \emptyset, i = 1, 2, \ldots, T$, maka untuk sembarang kejadian *A*, dengan $P(A) \neq 0$, berlaku:

$$P(B_r \mid A) = \frac{P(B_r)P(A \mid B_r)}{\sum_{i=1}^{T} P(B_i)P(A \mid B_i)}$$

untuk *r* = 1, 2, …, *T*

2.2 Taksiran Maksimum Likelihood

Misalkan $X_1, X_2, \ldots, X_n$ adalah suatu sampel random berukuran n dari suatu distribusi dengan pdf $f(X;\theta)$, yang bergantung pada $\theta \in \Omega$, Ω disebut ruang parameter. Karena $X_1, X_2, \ldots, X_n$ merupakan sample random, pdf bersama dari $X_1, X_2, \ldots, X_n$ dapat dinyatakan sebagai:

$$f(X_1, X_2, \ldots, X_n;\theta) = f(X_1;\theta).f(X_2;\theta).\cdots.f(X_n;\theta) \qquad (2.2.1)$$

Pdf bersama dari $X_1, X_2, \ldots, X_n$ mengandung parameter θ, sehingga persamaan (2.2.1) dapat dituliskan sebagai suatu fungsi dari θ, sebut $L(\theta)$.

$$\begin{aligned} L(\theta) &= f(X_1, X_2, \ldots, X_n;\theta) \\ &= f(X_1;\theta).f(X_2;\theta).\cdots.f(X_n;\theta) \\ &= \prod_{i=1}^{n} f(X_i;\theta) \end{aligned} \qquad (2.2.2)$$

$L(\theta)$ disebut fungsi *likelihood*.

Akan dicari θ yang memaksimumkan $L(\theta)$. Untuk mempermudah perhitungan dalam mencari nilai θ, $L(\theta)$ dapat dimodifikasi ke dalam bentuk ln, karena nilai θ yang memaksimumkan $\ln L(\theta)$ sama dengan nilai θ yang memaksimumkan $L(\theta)$.

Sehingga persamaan (2.2.2) dapat dimodifikasi menjadi:

$$\begin{aligned} \ln L(\boldsymbol{\theta}) &= \ln\left(\prod_{i=1}^{n} f(X_i;\boldsymbol{\theta})\right) \\ &= \sum_{i=1}^{n} \ln f(X_i;\boldsymbol{\theta}) \end{aligned} \tag{2.2.3}$$

Nilai $\boldsymbol{\theta}$ yang memaksimumkan $\ln L(\boldsymbol{\theta})$, diperoleh dengan mendifferensialkan $\ln L(\boldsymbol{\theta})$ terhadap $\boldsymbol{\theta}$ dan menyamakannya dengan 0. Nilai $\boldsymbol{\theta} = u(X_1, X_2, \ldots, X_n)$ yang memaksimumkan $\ln L(\boldsymbol{\theta})$ disebut sebagai taksiran *maximum likelihood* dari $\boldsymbol{\theta}$ dan dinotasikan dengan $\widehat{\boldsymbol{\theta}}$.

2.3 Algoritma EM (*Expectation-Maximization*)

Algoritma EM merupakan suatu algoritma yang bersifat iteratif yang dapat digunakan untuk mencari MLE dimana salah satu variabel dalam model merupakan variabel laten. Misalkan X adalah suatu variabel laten dan $Y_1, Y_2, \ldots, Y_n$ adalah *observed variable*, yang mempunyai joint pdf $p(\boldsymbol{y}, \theta)$. Misalkan $L(\boldsymbol{y}, \theta)$ adalah fungsi log likelihood dari **Y**.

$$L(\boldsymbol{y}, \theta) = \log\left[p(\boldsymbol{y}, \theta)\right] \tag{2.3.1}$$

Misalkan $p(\boldsymbol{y}, x, \theta)$ adalah pdf bersama dari **Y** dan X, dengan θ adalah parameter dalam model. Karena, seperti yang telah dinyatakan pada pemisalan awal, X adalah variabel laten, maka salah satu cara untuk mencari taksiran θ yang memaksimumkan fungsi likelihood dari **Y** adalah dengan menggunakan algoritma EM.

Prinsip dari algoritma EM dapat dijelaskan menjadi 2 bagian sebagai berikut:

1) E-*Step*

   E-*step* dilakukan untuk mencari $E\left[\mathbf{log}\left[p\left(x,y,\theta_t\right)\right]\middle| y,\hat{\boldsymbol{\theta}}_{t-1}\right]$, dimana:

   $\hat{\boldsymbol{\theta}}_{t-1}$ adalah taksiran $\boldsymbol{\theta}$ pada iterasi ke-(*t*-1).

   $\boldsymbol{\theta}_t$ adalah nilai $\boldsymbol{\theta}$ pada iterasi ke-(*t*).

   $\boldsymbol{\theta}_0$ adalah suatu nilai taksiran awal yang diberikan.

2) M-*Step*

   Pada M-*step*, maksimumkan $E\left[\mathbf{log}\left[p\left(x,y,\theta_t\right)\right]\middle| y,\hat{\boldsymbol{\theta}}_{t-1}\right]$ terhadap θ_t untuk mendapatkan taksiran θ_t pada iterasi ke-(t), sebut $\hat{\theta}_t$.

Proses E-*step* dan M-*step* ini akan dilakukan terus secara iteratif sampai didapatkan suatu estimasi untuk θ yang konvergen atau $\left|\hat{\theta}_t - \hat{\theta}_{t-1}\right|$ cukup kecil.

Dapat ditunjukkan bahwa iterasi algoritma EM seperti yang dijelaskan melalui E-*step* dan M-*step* diatas akan meningkatkan nilai $L(y,\theta)$ pada setiap iterasinya.

<u>Bukti</u>:

Misalkan $q(x)$ adalah suatu pdf sebarang dari X, dimana $\sum_x q(x) = 1$. Maka persamaan (2.3.1) dapat dituliskan sebagai:

$$
\begin{aligned}
L(\boldsymbol{y},\theta) &= \log\left[p(\boldsymbol{y},\theta)\right] \\
&= \sum_x q(x)\log\left[p(\boldsymbol{y},\theta)\right] \\
&= \sum_x q(x)\log\left[\frac{p(\boldsymbol{y},x,\theta)}{p(x\,|\,\boldsymbol{y},\theta)}\times\frac{q(x)}{q(x)}\right] \\
&= \sum_x q(x)\left\{\log\left[p(\boldsymbol{y},x,\theta)\right]-\log\left[q(x)\right]+\log\left[\frac{q(x)}{p(x\,|\,\boldsymbol{y},\theta)}\right]\right\} \\
&= \sum_x q(x)\log\left[p(x,\boldsymbol{y},\theta)\right]-\sum_x q(x)\log\left[q(x)\right]+\sum_x q(x)\log\left[\frac{q(x)}{p(x\,|\,\boldsymbol{y},\theta)}\right]
\end{aligned}
\tag{2.3.2}
$$

Definisikan:

$$Q\left(q\,||\,p_{\text{joint}}\right) = \sum_x q(x)\log\left[p(x,\boldsymbol{y},\theta)\right] \tag{2.3.3}$$

$$H(q\,||\,q) = -\sum_x q(x)\log\left[q(x)\right] \tag{2.3.4}$$

$$KL\left(q\,||\,p_{\text{post}}\right) = \sum_x q(x)\log\left[\frac{q(x)}{p(x\,|\,\boldsymbol{y},\theta)}\right] \tag{2.3.5}$$

Jadi, persamaan (2.3.1) dapat dituliskan kembali menjadi:

$$L(y,\theta) = Q\left(q\,||\,p_{\text{joint}}\right)+H(q\,||\,q)+KL\left(q\,||\,p_{\text{post}}\right) \tag{2.3.6}$$

Pandang persamaan (2.3.6), dapat dibuktikan bahwa bagian terakhir, KL, bersifat:

a) $KL\left(q \,||\, p_{\text{post}}\right) \geq 0 \qquad \forall q$

b) $\exists\, p$ sedemikian sehingga $KL\left(p \,||\, p_{\text{post}}\right) = 0$ (2.3.7)

Bukti:

a) Akan dibuktikan $KL\left(q \,||\, p_{\text{post}}\right) \geq 0$

$$\begin{aligned} KL\left(q \,||\, p_{\text{post}}\right) &= \sum_x q(x)\log\left[\frac{q(x)}{p(x\,|\,\boldsymbol{y},\theta)}\right] \\ &= \sum_x q(x)\left(-\log\left[\frac{p(x\,|\,\boldsymbol{y},\theta)}{q(x)}\right]\right) \end{aligned} \tag{2.3.8}$$

Berdasarkan pertidaksamaan Jensen, untuk f suatu fungsi konveks $E\left[f(x)\right] \geq f\left(E[x]\right)$, dan karena $-\log\left[\frac{p(x\,|\,\boldsymbol{y},\theta)}{q(x)}\right]$ adalah suatu fungsi konveks, maka berlaku:

$$\begin{aligned} E\left[-\log\left[\frac{p(x\,|\,\boldsymbol{y},\theta)}{q(x)}\right]\right] &\geq -\log\left(E\left[\frac{p(x\,|\,\boldsymbol{y},\theta)}{q(x)}\right]\right) \\ \sum_x q(x)\left(-\log\left[\frac{p(x\,|\,\boldsymbol{y},\theta)}{q(x)}\right]\right) &\geq -\log\left(\sum_x q(x)\left[\frac{p(x\,|\,\boldsymbol{y},\theta)}{q(x)}\right]\right) \end{aligned} \tag{2.3.9}$$

Berdasarkan (2.3.10) dan (2.3.8), maka $KL(q \,||\, p)$ dapat dituliskan sebagai berikut:

$$
\begin{aligned}
KL(q \,||\, p) &= \sum_x q(x) \log\left[\frac{q(x)}{p(x \,|\, \boldsymbol{y}, \theta)}\right] = \sum_x q(x)\left(-\log\left[\frac{p(x \,|\, \boldsymbol{y}, \theta)}{q(x)}\right]\right) \\
&\geq -\log\left[\sum_x q(x)\left(\frac{p(x \,|\, \boldsymbol{y}, \theta)}{q(x)}\right)\right] \\
&= -\log\left[\sum_x p(x \,|\, \boldsymbol{y}, \theta)\right] = -\log[1] \\
&= 0
\end{aligned}
\qquad (2.3.10)
$$

Jadi, $KL(q \,||\, p) \geq 0$**,** $\forall q$.

b) Akan dibuktikan $\exists\, p$ sedemikian sehingga $KL(p \,||\, p_{\text{post}}) = 0$

Misal pilih $q(x) = p(x \,|\, \boldsymbol{y}, \theta)$, maka $\sum_x p(x \,|\, \boldsymbol{y}, \theta) = 1$, dan:

$$
\begin{aligned}
\log\left[\frac{q(x)}{p(x \,|\, \boldsymbol{y}, \theta)}\right] &= \log\left[\frac{p(x \,|\, \boldsymbol{y}, \theta)}{p(x \,|\, \boldsymbol{y}, \theta)}\right] \\
&= \log[1] \\
&= 0
\end{aligned}
\qquad (2.3.11)
$$

Maka $KL(q \,\|\, p_{\text{post}})$ sesuai dengan persamaan (2.3.5), dapat dituliskan sebagai berikut:

$$
\begin{aligned}
KL(q \,||\, p_{\text{post}}) &= \sum_x q(x) \log\left[\frac{q(x)}{p(x \,|\, \boldsymbol{y}, \theta)}\right] \\
&= \sum_x p(x \,|\, \boldsymbol{y}, \theta) \times 0 \\
&= 0
\end{aligned}
$$

Jadi, untuk $q(x) = p(x \,|\, \boldsymbol{y}, \theta)$, nilai $KL(q \,\|\, p_{\text{post}}) = 0$

Misal $q(x) = p\left(x \mid \boldsymbol{y}, \hat{\theta}_{t-1}\right)$, maka $KL = 0$. Kemudian, sebut:

$$Q_{\hat{\theta}_{t-1}} = \sum_{x} p\left(x \mid \boldsymbol{y}, \hat{\theta}_{t-1}\right) \log\left[p\left(x, \boldsymbol{y}, \theta_t\right)\right] \tag{2.3.12}$$

$$H_{\hat{\theta}_{t-1}} = -\sum_{x} p\left(x \mid \boldsymbol{y}, \hat{\theta}_{t-1}\right) \log\left[p\left(x \mid \boldsymbol{y}, \hat{\theta}_{t-1}\right)\right] \tag{2.3.13}$$

Dengan mensubstitusikan (2.3.13), dan (2.3.14) ke (2.3.6), maka berlaku:

$$\begin{aligned} L\left(\boldsymbol{y}, \theta_t\right) &= Q_{\hat{\theta}_{t-1}} + H_{\hat{\theta}_{t-1}} + 0 \\ &= Q_{\hat{\theta}_{t-1}} + H_{\hat{\theta}_{t-1}} \end{aligned} \tag{2.3.14}$$

dimana $H_{\hat{\theta}_{t-1}}$ tidak bergantung pada θ_t (seperti terlihat pada (2.3.13)).

Misalkan $\hat{\theta}_t$ adalah taksiran θ yang memaksimumkan $Q_{\theta_{t-1}}$, maka:

$$\begin{aligned} L\left(\boldsymbol{y}, \theta_t\right) &= Q_{\hat{\theta}_{t-1}} + H_{\hat{\theta}_{t-1}} \\ &\leq Q_{\hat{\theta}_t} + H_{\hat{\theta}_t} \\ &= L\left(\boldsymbol{y}, \theta_{t+1}\right) - KL_{\theta_t} \\ &= L\left(\boldsymbol{y}, \theta_{t+1}\right) - 0 \\ &= L\left(\boldsymbol{y}, \theta_{t+1}\right) \end{aligned} \tag{2.3.16}$$

Jadi, $L\left(\boldsymbol{y}, \theta_t\right) \leq L\left(\boldsymbol{y}, \theta_{t+1}\right)$.

Jadi, terbukti bahwa dengan menggunakan algoritma EM akan didapatkan taksiran θ yang memaksimumkan fungsi likelihood dari **Y**.

2.4 Uji Rasio Likelihood

Terdapat variabel random $X_1, X_2, \ldots, X_n$ yang memiliki pdf $f_i\left(X_i; \theta_1, \theta_2, \ldots, \theta_m\right)$ untuk i = 1, 2, …, n. Parameter-parameter dari populasi $\theta_1, \theta_2, ..., \theta_m$ dimisalkan berada pada ruang parameter Ω. Misalkan ω merupakan subset dari Ω dan akan diuji hipotesis sebagai berikut:

$H_0 : \left(\theta_1, \theta_2, \ldots, \theta_m\right) \in \omega$
H_1 : tidak demikian

Definisikan fungsi likelihood sebagai berikut:

- Sebut joint pdf dari $X_1, X_2, \ldots, X_n$ pada saat $\left(\theta_1, \theta_2, ..., \theta_m\right) \in \omega$:

$$\mathrm{L}(\omega)$$

- Sebut joint pdf dari $X_1, X_2, \ldots, X_n$ pada saat $\left(\theta_1, \theta_2, ..., \theta_m\right) \in \Omega$:

$$L(\Omega)$$

Pandang rasio likelihood dari kedua fungsi likelihood sebagai berikut:

$$\lambda^* = \frac{L(\omega)}{L(\Omega)}$$

Nilai λ^* tidak dapat digunakan sebagai statistik uji untuk menguji H_0 dan H_1 karena nilai $L(\omega)$ dan $L(\Omega)$ biasanya tidak dapat ditetapkan dengan lengkap, Misalkan $\hat{\omega}$ merupakan taksiran maksimum likelihood untuk ω dan $\hat{\Omega}$ merupakan taksiran maksimum likelihood untuk Ω.

Pandang:

$$\lambda\left(x_1,x_2,\ldots,x_n\right)=\lambda=\frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

Nilai λ dapat dicari. λ dapat digunakan sebagai statistik uji untuk menguji H_0 dan H_1.

Perhatikan $\lambda=\dfrac{L(\hat{\omega})}{L(\hat{\Omega})}$.

$L(\hat{\omega})$ dan $L(\hat{\Omega})$ bernilai positif sehingga λ juga akan bernilai positif, atau $\lambda \geq 0$. Kemudian karena $\hat{\omega}$ adalah subset dari $\hat{\Omega}$, maka $L(\hat{\omega}) \leq L(\hat{\Omega})$.

Karena $L(\hat{\omega}) \leq L(\hat{\Omega})$, maka nilai $\lambda \leq 1$. Jadi, diperoleh $0 \leq \lambda = \dfrac{L(\hat{\omega})}{L(\hat{\Omega})} \leq 1$.

Apabila $\lambda = 0$ maka $L(\hat{\omega}) = 0$, H_0 ditolak.

Jadi, H_0 akan ditolak apabila λ bernilai kecil (mendekati 0).

Misalkan λ_0 adalah suatu bilangan positif sedemikian sehingga $H_0:\left(\theta_1,\theta_2,\ldots,\theta_m\right)\in\omega$ akan ditolak apabila $\lambda\left(x_1,x_2,\ldots,x_n\right)=\lambda\leq\lambda_0$.

Misalkan α adalah tingkat signifikansi yang dipakai dalam pengujian.

$$\alpha=\Pr\left[\lambda\left(x_1,x_2,\ldots x_n\right)\leq\lambda_0;H_0\right].$$

Jika pdf dari statistik $\lambda=\lambda\left(X_1,X_2,\ldots,X_n\right)$ dapat diketahui untuk H_0 benar, λ_0 dapat ditentukan, sedemikian sehingga $\alpha=\Pr\left[\lambda\left(x_1,x_2,\ldots x_n\right)\leq\lambda_0;H_0\right]$.

Namun, seringkali, di bawah H_0 benar, sulit untuk menentukan distribusi dari

$\lambda = \lambda\left(X_1, X_2, ..., X_n\right)$. Oleh karena itu tidak mungkin untuk menemukan λ_0 sedemikian sehingga $\alpha = \Pr\left[\lambda\left(x_1, x_2, ...x_n\right) \leq \lambda_0; H_0\right]$.

Karena distribusi dari statistik $\lambda = \lambda\left(X_1, X_2, ..., X_n\right)$ sulit ditentukan, maka dicari statistik uji lain yaitu $L^2 = -2\ln\lambda$ (log rasio likelihood). Dapat dibuktikan bahwa $-2\ln\lambda$ akan berdistribusi χ^2_r, dimana r = dimensi $\boldsymbol{\Omega}$ – dimensi $\boldsymbol{\omega}$.