전산통계

Chapter 4.3 뉴턴 알고리즘

가우스-뉴턴 알고리즘

- 모수 θ의 최소제곱추정량 (least square estimator, LSE)을 구할 때 사용
- 선형회귀모형 (linear regression model) : 하나의 반응변수 y와 p개의 설명변수들과의 인과관계를 알아내기 위한 모형
- $y_t = E\left(y_t \middle| x_{1t}, x_{2t}, \cdots x_{pt}\right) + \varepsilon_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_p x_{pt} + \varepsilon_t$
- $t = 1,2, \cdots, n$, ε_t : i.i.d.
- $y = X\beta + \varepsilon$, X : 자료행렬
- 회귀계수벡터의 최소제곱추정량 $\hat{\beta}_{LSE} = (X^T X)^{-1} X^T y$

- 비선형 회귀모형 $y_t = f(x_t, \beta) + \varepsilon_t,\ t = 1,2,\cdots,n,\ \varepsilon_t$: i.i.d.
- 순환공식을 이용한 뉴턴 알고리즘을 통해 LSE추정 가능

- $y_t = f(x_t, \hat{\beta}) + e_t \ \ = f(x_t, b_m) + \sum_{j=1}^{p} \left[\frac{\partial}{\partial \beta_j} f(x_t, \beta)\right]_{\beta = b_m} \times (\widehat{\beta_j} - b_{jm}) + e_t^*$

$$\Rightarrow y_t - f(x_t, b_m) = \sum_{j=1}^{p} \left[\frac{\partial}{\partial \beta_j} f(x_t, \beta)\right]_{\beta = b_m} \times (\widehat{\beta_j} - b_{jm}) + e_t^*$$

$$\Rightarrow \mathrm{y} - f_m = D_m(\hat{\beta} - b_m) + e_m$$

$$x_t = (x_{1t}, x_{2t}, \cdots x_{pt})^T, b_m = (b_{1m}, b_{2m}, \cdots b_{pm})^T, D_m = \left[\frac{\partial}{\partial \beta_j} f(x_t, \beta)\right]_{n \times p}$$

$$Var(e_m) = I\sigma^2$$

- $\left(\widehat{\hat{\beta} - b_m}\right) = (D_m^T D_m)^{-1} D_m^T (\mathrm{y} - f_m)$

  $\Rightarrow \theta_m = \hat{\beta} - b_m$

  $\Rightarrow \hat{\beta} = b_m + \theta_m \risingdotseq b_m + \widehat{\theta_m} \equiv b_{m+1}$

  $\Rightarrow b_{m+1} = b_m + \widehat{\theta_m} = b_m + (D_m^T D_m)^{-1} D_m^T (\mathrm{y} - f_m)$

- 1단계) 초기화 단계 : 초기치 b_0, 허용한계치 ϵ
- 2단계) 반복 단계 : $m = 0,1,2,\cdots$에 대하여 $b_{m+1} = b_m + (D_m^T D_m)^{-1} D_m^T (\mathrm{y} - f_m)$
- 3단계) 종료 단계 : $\frac{||b_{m+1} - b_m||}{||b_m||} < \varepsilon$이면 반복을 중지하고 b_{m+1}을 LSE로 사용

[예] 비선형모형

① model: $y = e^{\alpha + \beta x} + \epsilon$

② $\frac{\partial}{\partial\alpha} e^{\alpha+\beta x} = e^{\alpha+\beta x}, \; \frac{\partial}{\partial\beta} e^{\alpha+\beta x} = x\, e^{\alpha+\beta x}$

③ Gauss-Newton method

[Step 1] initial value: $b_0 = (\alpha_{0,} \beta_0)$

[Step 2] iteration: $b_{m+1} = b_m + (D_m{}' D_m)^{-1} D_m{}' (y - f_m)$, $m = 0, 1, 2, \cdots$

$$\text{where } D_m = \begin{pmatrix} \exp(\alpha_m + \beta_m x_1) & x_1 \exp(\alpha_m + \beta_m x_1) \\ \exp(\alpha_m + \beta_m x_2) & x_2 \exp(\alpha_m + \beta_m x_2) \\ \cdots & \cdots \\ \exp(\alpha_m + \beta_m x_n) & x_n \exp(\alpha_m + \beta_m x_n) \end{pmatrix} \quad \text{and} \quad f_m = \begin{pmatrix} \exp(\alpha_m + \beta_m x_1) \\ \exp(\alpha_m + \beta_m x_2) \\ \cdots \\ \exp(\alpha_m + \beta_m x_n) \end{pmatrix}$$

뉴턴-랩슨 알고리즘

- 모수 θ의 최대가능도 (maximum likelihood estimate, MLE)을 구할 때 사용
- 모수공간 $\theta = \{\theta_1, \theta_2, \theta_3\}$, 표본공간 $S = \{x_1, x_2, x_3\}$

$f(x\|\theta)$	x_1	x_2	x_3
θ_1	0.4	0.6	0.0
θ_2	0.0	0.8	0.2
θ_3	0.8	0.2	0.0

- 확률밀도함수 $f(x|\theta)$, 가능도함수(likelihood function) $L(\theta|x)$
- 최대가능도원리 : 가능도함수의 가장 큰 값을 제공하는 θ를 MLE로 한다.
- MLE = $L(\theta|x)$을 최대로 하는 θ
- 일반적으로 로그-가능도함수 $l(\theta) = logL(\theta)$를 최대로 하는 θ를 추정
- $l'(\theta) = 0$의 해를 구하는 문제

- $\hat{\theta}$: $l'(\theta) = 0$의 해
- $0 = \left[\frac{d}{d\theta} l(\theta)\right]_{\theta=\hat{\theta}} \equiv \frac{d}{d\theta} l\left(\hat{\theta}\right) = \frac{d}{d\theta}\left\{l(\theta_k) + \left(\hat{\theta} - \theta_k\right)\frac{d}{d\theta} l(\theta_k) + \left(\hat{\theta} - \theta_k\right)^2 \frac{d^2}{d\theta^2} l(\theta_k) + \right.$

- 1단계) 초기화 단계 : 초기치θ_0, 허용한계치 ϵ

- 2단계) 반복 단계 : $k = 0,1,2,\cdots$에 대하여 $\theta_{k+1} = \theta_k + \left(-\frac{d^2}{d\theta^2} l(\theta_k)\right)^{-1} \frac{d}{d\theta} l(\theta_k)$

- 3단계) 종료 단계 : $\left|\frac{\theta_{k+1}-\theta_k}{\theta_k}\right| < \varepsilon$이면 반복을 중지하고 θ_{k+1}을 MLE로 사용

- 1단계) 초기화 단계 : 초기치θ_0, 허용한계치 ϵ

- 2단계) 반복 단계 : $k = 0,1,2,\cdots$에 대하여 $\theta_{k+1} = \theta_k + \left(-\frac{d^2}{d\theta^2} l(\theta_k)\right)^{-1} \frac{d}{d\theta} l(\theta_k)$

- 3단계) 종료 단계 : $\left|\frac{\theta_{k+1}-\theta_k}{\theta_k}\right| < \varepsilon$이면 반복을 중지하고 θ_{k+1}을 MLE로 사용

- $f(x) = \frac{1}{4}x^4 - x^2$
- $f(x|\lambda) = \lambda e^{-\lambda x}$