# 6장, 7장 복습 문제

**〔1번 문제〕**

다음은 어느 실험에서 얻은 반응변수 $Y$와 설명변수 $X_1$, $X_2$, $X_3$의 자료이다.

$$Y=\begin{bmatrix}2\\3\\1\\2\\2\\1\\3\\3\end{bmatrix} \qquad X=\begin{bmatrix}1 & -1 & -1 & -1\\1 & +1 & -1 & -1\\1 & -1 & +1 & -1\\1 & +1 & +1 & -1\\1 & -1 & -1 & +1\\1 & +1 & -1 & +1\\1 & -1 & +1 & +1\\1 & +1 & +1 & +1\end{bmatrix}$$ ($X$ 행렬의 첫 열은 상수항임)

(1) 회귀모형 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\epsilon$을 적합시킬 때 회귀계수들의 최소제곱 추정치를 구하시오.

(2) 회귀모형 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\epsilon$을 적합시킬 때 회귀계수들의 최소제곱추정치를 구하시오.

(3) 위의 두 결과를 비교하고, 추정치들이 그와 같이 나온 이유를 설명하시오.

**〔2번 문제〕**

**안부점(saddle point)**을 이차원 평면에 반응표면을 그려서 예를 들어 설명하시오.

**〔3번 문제〕**

회귀분석에서는 오차항의 분산이 $\sigma^2$으로 일정하다고 가정한다. 만약 이 가정이 충족되지 않는 상황이라면, 이분산성 가정 하에 수행되는 최소제곱법이 데이터에 적합될 수 있다. 어떤 대각행렬 $W=diag(w_1,\ \cdots,w_n)$에 대해 $Cov(\epsilon)=\sigma^2W^{-1}$가 성립한다고 가정하자. 이 회귀모형에서 회귀계수 $\beta$의 최소제곱추정량과 $\hat{\beta}$의 분산-공분산행렬을 구하시오. (변수 변환을 사용하시오.)

**〔4번 문제〕**

능형 추정치와 LASSO 추정치가 제약식의 형태에 따라 어떤 차이를 보이는지 수식적 표현과 그림으로 비교•설명하시오.

**〔5번 문제〕**

$\beta$의 사전분포가 $\beta\sim N\left(0,\ \dfrac{\sigma^2}{\theta}I\right)$일 때 능형추정치 $\hat{\beta}(\theta)$는 제곱오차 손실함수 하에서 $\beta$의 베이스 추정치임을 보이시오.

**〔6번 문제〕**

Multiple Linear Regression Model에서 Ridge Regression은 다음과 같은 목표함수를 최소화하는 것으로 추정된다.

$$\hat{\beta}^{ridge} = \underset{\beta}{\arg\min} \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p}\beta_j^2 \le t$$

여기서 $t$는 임의의 양수이고 $\beta = (\beta_1, \cdots, \beta_p)^T$는 추정되어야 할 회귀계수이다.

Lagrange 방법을 이용하여 $\hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T y$가 됨을 증명하시오.

단, $X = \{x_{ij}\}_{i=1,\cdots,n;\, j=1,\cdots,p}$는 $n \times p$ 행렬이고, $y = (y_1, \cdots, y_n)^T$이다.

**〔7번 문제〕**

'주성분회귀'와 '부분최소제곱법'의 차이점을 설명하시오.