STATS 3001 / STATS 4101 / STATS 7054
Statistical Modelling III
Tutorial 3
2022
Solutions

QUESTIONS:

1. The purpose of this tutorial is to derive the maximum likelihood estimates for the multiple regression model. It is intended partly to help with understanding the derivation of the Box-Cox profile likelihood.

   Consider the multiple regression model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$$

   where $E_1, E_2, \ldots, E_n$ are independent with

$$E_i \sim N(0, \sigma^2)$$

   (a) Show that the model can be expressed equivalently as

$$Y_i \sim N(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma^2)$$

   independently for $i = 1, 2, \ldots, n$ where $\boldsymbol{x}_i^T$ is the ith row of X.

   (b) Write down the log-likelihood function, $\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y})$.

   (c) Show that for any value of $\sigma^2 > 0$ the log-likelihood is maximized with respect to $\boldsymbol{\beta}$ by the ordinary least squares estimate $\hat{\boldsymbol{\beta}}$.

   (d) Hence show that the maximum likelihood estimate for σ^2 is

$$\hat{\sigma}^2 = \frac{RSS(\boldsymbol{y})}{n}$$

   where the residual sum of squares (RSS) is defined by

$$RSS(\boldsymbol{y}) = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}})^2.$$

   (e) If c is a scalar then show that

$$RSS(c\boldsymbol{y}) = c^2 RSS(\boldsymbol{y}).$$

SOLUTIONS:

(a) Observe first that the ith element of $X\boldsymbol{\beta}$ is $\boldsymbol{x}_i^T\boldsymbol{\beta}$. Also, since $\boldsymbol{x}_i^T\boldsymbol{\beta}$ is constant for each i, it follows that

$$Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \mathcal{E}_i \sim N(\boldsymbol{x}_i^T\boldsymbol{\beta}, \sigma^2)$$

independently for $i = 1, 2, \ldots, n$.

(b)

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y}) &= \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2\right)\right) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}Q(\boldsymbol{\beta})
\end{aligned}$$

where

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2.$$

(c) Since ℓ depends on $\boldsymbol{\beta}$ through $-Q(\boldsymbol{\beta})/(2\sigma^2)$, it follows that ℓ is maximised with respect to $\boldsymbol{\beta}$ when $Q(\boldsymbol{\beta})$ is minimised. That is, when

$$\boldsymbol{\beta} = (X^TX)^{-1}X^T\boldsymbol{y}.$$

(d)

$$\frac{\partial \ell(\hat{\boldsymbol{\beta}}, \sigma^2; \boldsymbol{y})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}Q(\hat{\boldsymbol{\beta}}).$$

Solving for σ^2 in $\frac{\partial \ell}{\partial \sigma^2}$ yields

$$\hat{\sigma}^2 = \frac{Q(\hat{\boldsymbol{\beta}})}{n}$$

as required.

(e) Observe first that

$$RSS(\boldsymbol{y}) = \|(I-P)\boldsymbol{y}\|^2.$$

Hence

$$\begin{aligned} RSS(c\boldsymbol{y}) &= \|(I-P)c\boldsymbol{y}\|^2 \\ &= \|c(I-P)\boldsymbol{y}\|^2 \\ &= c^2\|(I-P)\boldsymbol{y}\|^2 \end{aligned}$$

as required.
