LiMo WiSe 16/17 Sheet 4: Ex 6 (Zusatzaufgabe)

Task:

Gegeben sei das Modell der linearen Einfachregression mit einem Achsenabschnitt und einem Prediktor x_i für $i = 1, \ldots, n$ Objekte. Zeigen Sie dass in der Hat-Matrix $H = X(X'X)^{-1}X'$ die Summe aller Elemente in jeder Zeile und in jeder Spalte gleich 1 ist.

Solution:

Wir wissen:

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

und

$$\begin{aligned}
(X'X)^{-1} &= \left(\begin{pmatrix} 1 & \ldots & n \\ x_1 & \ldots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \\
&= \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \\
&= \frac{1}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}
\end{aligned}$$

Das heißt H ist gegeben durch

$$
\begin{aligned}
H &= X(X'X)^{-1}X' \\
&= \underset{(n\times 2)}{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}} \overbrace{\frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}^{=a} \underset{(2\times 2)}{\begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}} \underset{(2\times n)}{\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}} \\
&= a \underset{(n\times 2)}{\begin{pmatrix} \sum_{i=1}^n x_i^2 - x_1 \sum_{i=1}^n x_i & -\sum_{i=1}^n x_i + nx_1 \\ \vdots & \vdots \\ \sum_{i=1}^n x_i^2 - x_n \sum_{i=1}^n x_i & -\sum_{i=1}^n x_i + nx_n \end{pmatrix}} \underset{(2\times n)}{\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix}} \\
&= a \underset{(n\times n)}{\begin{pmatrix} \sum_{i=1}^n x_i^2 - x_1 \sum_{i=1}^n x_i + x_1(-\sum_{i=1}^n x_i + nx_1) & \dots & \sum_{i=1}^n x_i^2 - x_1 \sum_{i=1}^n x_i - x_n(\sum_{i=1}^n x_i + nx_1) \\ \vdots & & \vdots \\ \sum_{i=1}^n x_i^2 - x_n \sum_{i=1}^n x_i + x_1(-\sum_{i=1}^n x_i + nx_n) & \dots & \sum_{i=1}^n x_i^2 - x_n \sum_{i=1}^n x_i - x_n(\sum_{i=1}^n x_i + nx_n) \end{pmatrix}} \\
&\Rightarrow \text{Summe über die erste Spalte} \\
&= a \left[\sum_{i=1}^n x_i^2 - x_1 \sum_{i=1}^n x_i + x_1 \left(-\sum_{i=1}^n x_i + nx_1 \right) + \dots + \sum_{i=1}^n x_i^2 - x_n \sum_{i=1}^n x_i - x_1 \left(\sum_{i=1}^n x_i + nx_n \right) \right] \\
&= a \left[n \sum_{i=1}^n x_i^2 - x_1 \sum_{i=1}^n x_i - \dots - x_n \sum_{i=1}^n x_i - nx_1 \sum_{i=1}^n x_i + \underbrace{nx_1x_1 + \dots + nx_1x_n}_{=nx_1 \sum_{i=1}^n x_i} \right] \\
&= a \left[n \sum_{i=1}^n x_i^2 - \underbrace{x_1 \sum_{i=1}^n x_i - \dots - x_n \sum_{i=1}^n x_i}_{=\sum_{i=1}^n x_i(x_1+\dots x_n)=(\sum_{i=1}^n x_i)^2} \right] \\
&= a \left[n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 \right] \\
&= \frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= 1
\end{aligned}
$$