6.4.3 *Multiple linear regression*

So far we have considered datasets where the dependent variable y is a function of only one independent variable x. This functional dependence could be either linear or non-linear. However, in most practical applications, y may have a functional dependence on multiple independent attributes, i.e. each data variable $\mathbf{x}_i$ may be a $k = (m-1)$dimensional vector $(x_{i1}, x_{i2}, ..., x_{ik})$ for each of the $i = 1, 2, ..., n$ data points. In this section, we will seek a *multiple linear regression* model $M(\mathbf{x}) = p_0 + p_1 x_1 + p_2 x_2 + \cdots + p_k x_k$. This is similar in form to the model in section 6.4.2 if we consider the individual model functions to have a linear

dependence on the different components of the independent data vector $\mathbf{x}$. The least squares solution is given by

$$\mathbf{p} = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \cdot \\ \cdot \\ p_k \end{pmatrix}_{m\times 1} \tag{6.19}$$

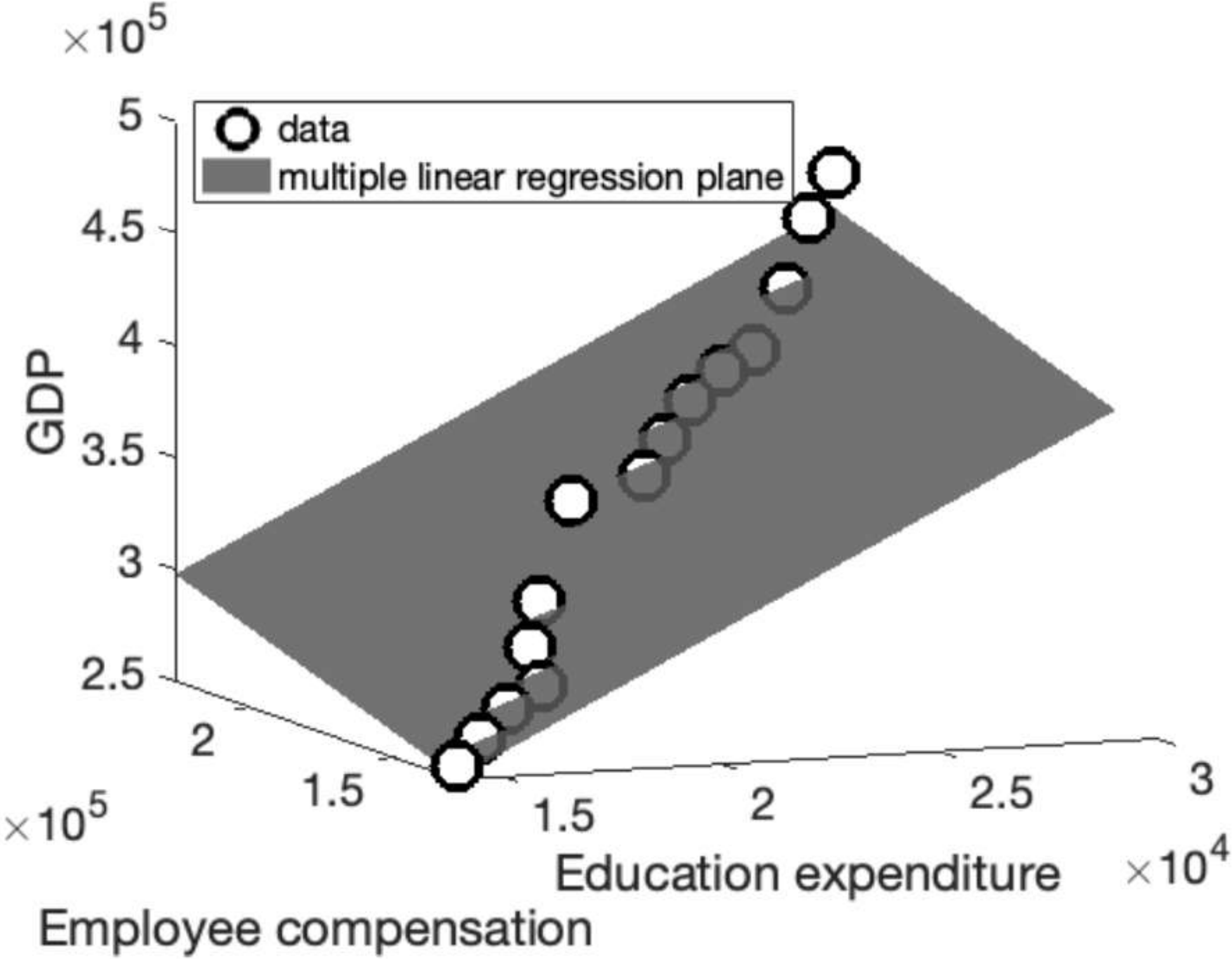
$$= \left(\Lambda^T \Lambda\right)^{-1} \Lambda^T \mathbf{q}, \tag{6.20}$$

where the matrix $\Lambda = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & x_{1k} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & x_{nk} \end{pmatrix}_{n\times m}$.

Example: mulitple linear regression for predicting multi-dimensional data

Consider the following data sourced from the Eurosat public repository. Here the GDP of Belgium (y) is a function of two independent indicators: (i) educational expenditure by the state (x_1), and (ii) employee compensation (x_2). All figures below are in million euros and have been rounded to the nearest thousand. Construct a linear regression model of this data.

GDP (y)	educational expenditure (x_1)	employee compensation (x_2)
256000	14000	129000
264000	15000	136000
273000	16000	142000
281000	17000	145000
296000	17000	149000
310000	18000	161000
344000	20000	180000
351000	22000	185000
363000	23000	193000
376000	24000	200000
386000	25000	204000
393000	26000	208000
417000	27000	212000
445000	28000	219000
461000	29000	226000


×10^5
5
4.5
4
3.5
3
2.5
GDP
data
multiple linear regression plane
2
1.5
×10^5
Employee compensation
1.5
2
2.5
3
×10^4
Education expenditure