Multi-output linear regression

P.L.Green

November 10, 2020

1 Abstract

The maths behind multi-output linear regression.

2 Problem statement

We have N input-output pairs $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n=1}^N$ where $\boldsymbol{y}_n \in \mathbb{R}^K$. We propose the model

$$y_{n,i} = \boldsymbol{\theta}_i^T \boldsymbol{\phi}(\boldsymbol{x}_n) \tag{1}$$

where $\boldsymbol{\phi}$ is a D-dimensional basis function. In general, we are therefore proposing that

$$\begin{pmatrix} y_{n,1} \\ \vdots \\ y_{n,K} \end{pmatrix} = \begin{bmatrix} \boldsymbol{\theta}_1^T \\ \vdots \\ \boldsymbol{\theta}_K^T \end{bmatrix} \boldsymbol{\phi}(\boldsymbol{x}_n) \tag{2}$$

which we write as

$$\boldsymbol{y}_n = \boldsymbol{\Theta}^T \boldsymbol{\phi}(\boldsymbol{x}_n) \tag{3}$$

3 Least-squares solution

Objective function:

$$J = \frac{1}{2} \sum_{n=1}^{N} \sum_{i=1}^{K} \left(y_{n,i} - \boldsymbol{\theta}_i^T \boldsymbol{\phi}_n\right)^2 \tag{4}$$

$$= \frac{1}{2} \sum_n \sum_i \left(-2 y_{n,i} \boldsymbol{\theta}_i^T \boldsymbol{\phi}_n + \boldsymbol{\theta}_i^T \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T \boldsymbol{\theta}_i\right) + \text{const.} \tag{5}$$

where we have adopted the notation $\boldsymbol{\phi}_n \equiv \boldsymbol{\phi}(\boldsymbol{x}_n)$. Through standard vector calculus we can show that

$$\frac{\partial J}{\partial \boldsymbol{\theta}_j} = -\sum_n y_{n,j} \boldsymbol{\phi}_n + \left[\sum_n \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T\right] \boldsymbol{\theta}_j \tag{6}$$

With the aim of evaluating

$$\frac{\partial J}{\partial \boldsymbol{\Theta}} = \left[\begin{array}{ccc} \frac{\partial J}{\partial \boldsymbol{\theta}_1} & \frac{\partial J}{\partial \boldsymbol{\theta}_2} & \dots \end{array} \right] \tag{7}$$

we can now write that

$$\frac{\partial J}{\partial \mathbf{\Theta}} = \left[\; -\sum_n y_{n,1}\, \boldsymbol{\phi}_n + \left[\sum_n \boldsymbol{\phi}_n\, \boldsymbol{\phi}_n^T\right] \boldsymbol{\theta}_1 \quad -\sum_n y_{n,2}\, \boldsymbol{\phi}_n + \left[\sum_n \boldsymbol{\phi}_n\, \boldsymbol{\phi}_n^T\right] \boldsymbol{\theta}_2 \quad \ldots \; \right] \tag{8}$$

$$= -\sum_n \left[\; y_{n,1}\, \boldsymbol{\phi}_n \quad y_{n,2}\, \boldsymbol{\phi}_n \quad \ldots \; \right] + \left[\sum_n \boldsymbol{\phi}_n\, \boldsymbol{\phi}_n^T\right] \underbrace{\left[\; \boldsymbol{\theta}_1 \quad \boldsymbol{\theta}_2 \quad \ldots \; \right]}_{\mathbf{\Theta}} \tag{9}$$

Consequently, defining

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1^T \\ \vdots \\ \boldsymbol{y}_N^T \end{bmatrix} \qquad \mathbf{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1^T \\ \vdots \\ \boldsymbol{\phi}_N^T \end{bmatrix} \tag{10}$$

we can write that

$$\frac{\partial J}{\partial \mathbf{\Theta}} = -\, \mathbf{\Phi}^T\, \boldsymbol{Y} + \mathbf{\Phi}^T\, \mathbf{\Phi}\, \mathbf{\Theta} \tag{11}$$

Setting the above expression equal to zero and solving for the optimal parameter matrix we obtain

$$\mathbf{\Theta}^* = \left[\mathbf{\Phi}^T\, \mathbf{\Phi}\right]^{-1} \mathbf{\Phi}^T\, \boldsymbol{Y} \tag{12}$$

4 Regularised least-squares solution

If we use the classic regularisation term that is employed in ridge-regression then the objective function is

$$J = \frac{1}{2} \sum_{n=1}^{N} \sum_{i=1}^{K} \left(y_{n,i} - \boldsymbol{\theta}_i^T\, \boldsymbol{\phi}_n\right)^2 + \frac{\lambda}{2} \sum_{i=1}^{K} \boldsymbol{\theta}_i^T\, \boldsymbol{\theta}_i \tag{13}$$

then, following a procedure similar to the previous section, we can show that

$$\frac{\partial J}{\partial \boldsymbol{\theta}_j} = -\sum_n y_{n,j}\, \boldsymbol{\phi}_n + \left(\left[\sum_n \boldsymbol{\phi}_n\, \boldsymbol{\phi}_n^T\right] + \boldsymbol{I}\lambda\right) \boldsymbol{\theta}_j \tag{14}$$

and so

$$\frac{\partial J}{\partial \mathbf{\Theta}} = -\, \mathbf{\Phi}^T\, \boldsymbol{Y} + \left[\mathbf{\Phi}^T\, \mathbf{\Phi} + \boldsymbol{I}\lambda\right]^{-1} \mathbf{\Theta} \tag{15}$$

which implies that our optimal parameters are given by

$$\mathbf{\Theta}^* = \left[\mathbf{\Phi}^T\, \mathbf{\Phi} + \boldsymbol{I}\lambda\right]^{-1} \mathbf{\Phi}^T\, \boldsymbol{Y} \tag{16}$$