Multivariate Linear Regression (a, b, c) closed form derivation

Sonam Kumar
sonam [dot] ibab [at] gmail [dot] com

Q : Closed form of multivariate linear regression.

Soln. : Let the given data be of the following form:

$$\left\{\begin{matrix} x_1^{(1)}, x_2^{(1)}, x_3^{(1)} \dots x_N^{(1)} \\ x_1^{(2)}, x_2^{(2)}, x_3^{(2)} \dots x_N^{(2)} \\ y_1, y_2, y_3 \dots y_n \end{matrix}\right\} \equiv \left\{x_i^{(1)}, x_i^{(2)}, y_i\right\}_{i=1}^{N}$$

We want to fit the data as a plane $y = a + bx^{(1)} + cx^{(2)}$ in the three dimensional $(x^{(1)}, x^{(2)}, y)$ coordinate system, by estimating the parameters a, b, c using the data described by features $\{x_i^{(1)}\}_{i=1}^N$, $\{x_i^{(2)}\}_{i=1}^N$ and their linear combinations $\{y_i\}_{i=1}^N$ from the given data.

Let the predicted model be described by a function 'f' such that:

$$ypred_i = f(x_i^{(1)}, x_i^{(2)}) \text{ for all } i = 1, \dots, N$$

or equivalently,

$$ypred_i = a + bx_i^{(1)} + cx_i^{(2)} \quad \forall\, i = 1, \dots, N$$

as a linear combination of $\{1, x_i^{(1)}, x_i^{(2)}\}$ for $i = 1, \dots, N$

Let $\{ypred_i\}_{i=1}^N$ be the features of predicted model. Then for each data point $(x_i^{(1)}, x_i^{(2)}, y_i)$ when $i = 1, \dots, N$, the error is:

$$(y_i - ypred_i)^2$$
$$= (y_i - a - bx_i^{(1)} - cx_i^{(2)})^2$$

Then, the total error as approximated by what we call the "loss function" is as below:

$$J(a, b, c) = \sum_{i=1}^{N} \left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right)^2$$

Where a, b, c are the parameters to be computed.

Note that this is the phase when we transition from the data space to the model space. Both the data space and the model space are three dimensional and that a plane $y = a' + b'x^{(1)} + c'x^{(2)}$ will be the "plane of fit" (as the line of fit in case of $y = a + bx$ fit of the data $\{x_i, y_i\}_i$) in data space corresponding to point (a', b', c') in model space. A representative figure is depicted below:

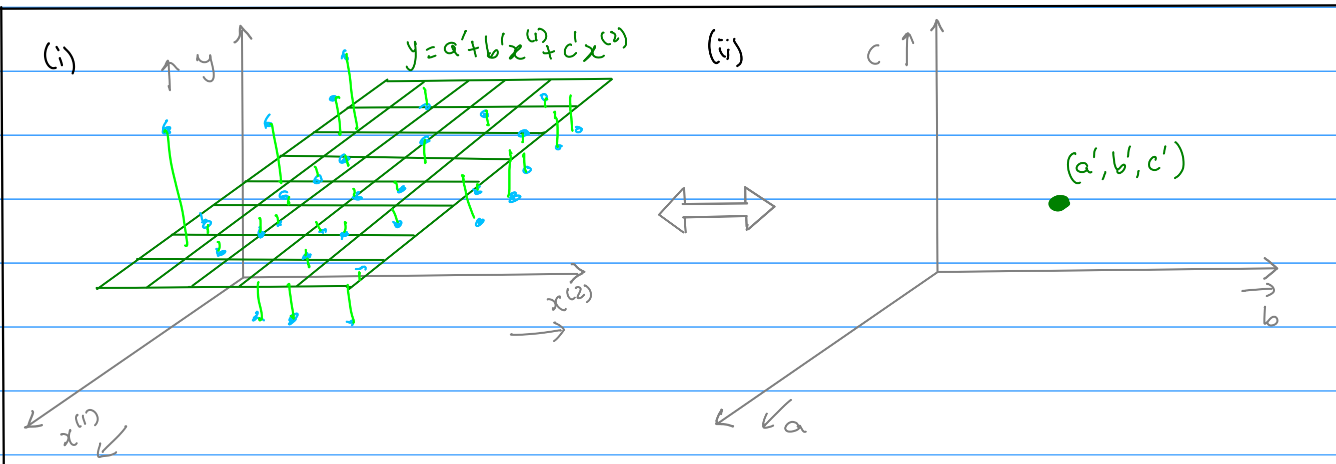

Fig. 1 Coordinate spaces denoting a particular predicted model of given data, namely $y\text{pred}_i = a' + b'x_i^{(1)} + c'x_i^{(2)}$ for all $i = 1, \dots, N$.
(i) DATA SPACE   (ii) MODEL SPACE

The aim of a closed form solution is to minimize the cost function such that we obtain an optimal set of parameters that would yield a model fitting the data.

In order to minimize the cost function, find the gradient and equate that to zero:

$$\nabla_{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} J(a,b,c) = 0$$

This can be achieved by equating all partial derivatives to zero:

$$\frac{\partial}{\partial a} J(a,b,c) = 0, \quad \frac{\partial}{\partial b} J(a,b,c) = 0, \quad \frac{\partial}{\partial c} J(a,b,c) = 0$$

Computing partial differentiation of $J(a,b,c)$ w.r.t. a considering b and c as constant gives:

$$\sum_{i=1}^{N} 2\left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right)(-1) = 0$$

or,
$$\sum_{i=1}^{N} \left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right) = 0$$

or,
$$Na + b\sum_{i=1}^{N} x_i^{(1)} + c\sum_{i=1}^{N} x_i^{(2)} = \sum_{i=1}^{N} y_i \quad \dots (I)$$

Computing partial differentiation of $J(a,b,c)$ w.r.t. b considering a and c as constant gives:

$$\sum_{i=1}^{N} 2\left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right)\left(-x_i^{(1)}\right) = 0$$

or, $$\sum_{i=1}^{N} \left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right) x_i^{(1)} = 0$$

or, $$\sum_{i=1}^{N} \left\{ x_i^{(1)} y_i - a x_i^{(1)} - b\left(x_i^{(1)}\right)^2 - c\, x_i^{(1)} x_i^{(2)} \right\} = 0$$

or, $$a \sum_{i=1}^{N} x_i^{(1)} + b \sum_{i=1}^{N} \left(x_i^{(1)}\right)^2 + c \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} = \sum_{i=1}^{N} x_i^{(1)} y_i \quad \cdots (\mathrm{II})$$

Computing partial differentiation of $J(a,b,c)$ w.r.t. c considering b and c as constant gives:

$$\sum_{i=1}^{N} 2\left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right)\left(-x_i^{(2)}\right) = 0$$

or, $$\sum_{i=1}^{N} \left(y_i - a - bx_i^{(1)} - cx_i^{(2)}\right) x_i^{(2)} = 0$$

or, $$\sum_{i=1}^{N} \left\{ x_i^{(2)} y_i - a x_i^{(2)} - b\, x_i^{(1)} x_i^{(2)} - c\left(x_i^{(2)}\right)^2 \right\} = 0$$

or, $$a \sum_{i=1}^{N} x_i^{(2)} + b \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} + c \sum_{i=1}^{N} \left(x_i^{(2)}\right)^2 = \sum_{i=1}^{N} x_i^{(2)} y_i \quad \cdots (\mathrm{III})$$

The linear system of equations (I), (II), (III) is obtained where the model parameters a, b, c are unknown whereas the following quantities can be computed from given data (as column sum of an excel sheet, maybe):

$$\sum_{i=1}^{N} x_i^{(1)}, \quad \sum_{i=1}^{N} x_i^{(2)}, \quad \sum_{i=1}^{N} y_i$$

$$\sum_{i=1}^{N} \left(x_i^{(1)}\right)^2, \quad \sum_{i=1}^{N} \left(x_i^{(2)}\right)^2, \quad \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)}$$

$$\sum_{i=1}^{N} x_i^{(1)} y_i, \quad \sum_{i=1}^{N} x_i^{(2)} y_i$$

The system of linear equations can hence be rewritten in matrix form as below:

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i^{(1)} & \sum_{i=1}^{N} x_i^{(2)} \\ \sum_{i=1}^{N} x_i^{(1)} & \sum_{i=1}^{N} \left(x_i^{(1)}\right)^2 & \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} \\ \sum_{i=1}^{N} x_i^{(2)} & \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} & \sum_{i=1}^{N} \left(x_i^{(2)}\right)^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i^{(1)} y_i \\ \sum_{i=1}^{N} x_i^{(2)} y_i \end{bmatrix}$$

<u>Notation</u>: Represent the above matrix equation by $\mathbf{M}\mathbf{u} = \mathbf{v}$, where $\mathbf{M}$ is a 3×3 matrix and $\mathbf{u}$ and $\mathbf{v}$ are 3×1 column vectors.

The column vector $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is obtained by multiplying the inverse of the 3×3 matrix $\mathbf{M}$ on both sides of the equation, which gives:

$$\mathbf{M}^{-1}\mathbf{M}\mathbf{u} = \mathbf{M}^{-1}\mathbf{v}$$

or,

$$\mathbf{u} = \mathbf{M}^{-1}\mathbf{v}$$

Explicitly, the matrix equation becomes:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^{N} x_i^{(1)} & \sum_{i=1}^{N} x_i^{(2)} \\ \sum_{i=1}^{N} x_i^{(1)} & \sum_{i=1}^{N} \left(x_i^{(1)}\right)^2 & \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} \\ \sum_{i=1}^{N} x_i^{(2)} & \sum_{i=1}^{N} x_i^{(1)} x_i^{(2)} & \sum_{i=1}^{N} \left(x_i^{(2)}\right)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i^{(1)} y_i \\ \sum_{i=1}^{N} x_i^{(2)} y_i \end{bmatrix}$$

Recall that the given data is in the form $\{x_i^{(1)}, x_i^{(2)}, y_i\}_{i=1}^{N}$ and thus the matrix $\mathbf{M}$ and vector $\mathbf{v}$ can be computed.

Consequently, inverse $\mathbf{M}^{-1}$ of the matrix $\mathbf{M}$ can be calculated by either

Using the augmented matrix (Gauss-Jordan elimination) method or by calculating the reciprocal determinant of the adjoint (adjoint is the transposed cofactor matrix).

Thus, the column vector $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is obtained from which the optimal parameters are obtained as the scalars described below:

$$a_{opt} = u_{11}$$
$$b_{opt} = u_{21}$$
$$c_{opt} = u_{31}$$

Therefore, the optimal predicted multivariate linear regression model for the data $\{x_i^{(1)}, x_i^{(2)}, y_i\}_{i=1}^N$ is as below:

$$\boxed{y = a_{opt} + b_{opt}\, x^{(1)} + c_{opt}\, x^{(2)}}$$

COROLLARY:

The formulation of multivariate linear regression as described above can be extended to a quadratic regression of the data $\{x_i, y_i\}_{i=1}^N$ as:

$$y = a_{opt} + b_{opt}\, x + c_{opt}\, x^2,$$

where the parameters are defined by the following matrix relation:

$$\begin{bmatrix} a_{opt} \\ b_{opt} \\ c_{opt} \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 \\ \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 & \sum_{i=1}^N x_i^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \\ \sum_{i=1}^N x_i^2 y_i \end{bmatrix}$$

The R.H.S. can be computed from the data $\{x_i, y_i\}_{i=1}^N$ by calculating the summations over $i = 1, \dots, N$ of the following:

$$x_i,\ x_i y_i,\ x_i^2,\ x_i^2 y,\ x_i^3,\ x_i^4$$

DONE!!!