

Multiple Linear Regression

MR. HUYNH NAM

Objects in Linear Algebra

(11)

5	3	7
---	---	---

SCALAR

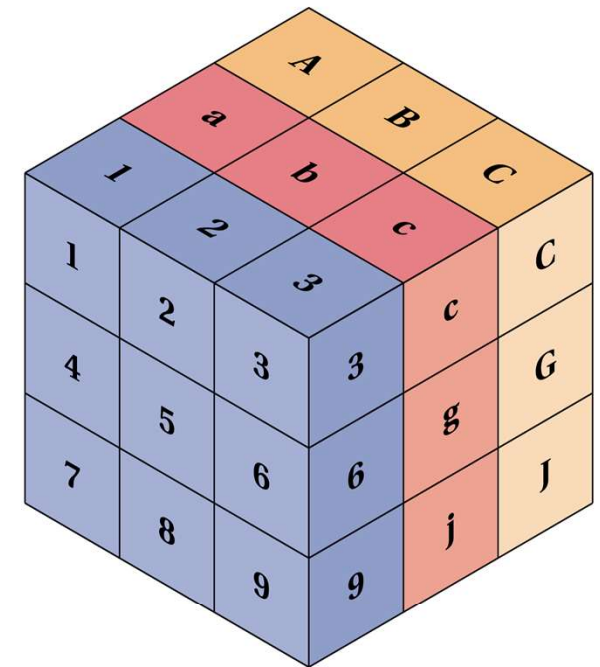
Row Vector
(shape 1x3)

5
1.5
2

Column Vector
(shape 3x1)

4	19	8
16	3	5

MATRIX



TENSOR

Định nghĩa

Cho $m, n \in \mathbb{N}$. Một bảng gồm $m.n$ số thực được xếp thành m hàng, n cột được gọi là một ma trận thực (ma trận) cấp (cỡ) $m \times n$. Ký hiệu:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix} \text{ hoặc } A = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix}$$

Ký hiệu tắt $A = [a_{ij}]_{m \times n}$ hoặc $A = (a_{ij})_{m \times n}$

Ký hiệu ma trận bởi các chữ in A, B, C, \dots

a_{ij} : ký hiệu của phần tử nằm ở hàng i , cột j của ma trận A .

Ma trận không

Ma trận cấp $m \times n$ có *tất cả các phần tử bằng 0* được gọi là ma trận không cấp $m \times n$, ký hiệu $O_{m \times n}$ hoặc O .

Ma trận vuông

Ma trận gồm n hàng, n cột được gọi là ma trận vuông cấp n , ký hiệu A_n thay cho $A_{n \times n}$.

Với $A = [a_{ij}]_{n \times n}$,

- $a_{11}, a_{22}, \dots, a_{nn}$: *các phần tử chéo* của ma trận A (nằm trên đường chéo chính);
- $\text{tr}(A) = a_{11} + a_{22} + \dots + a_{nn}$ gọi là vết của ma trận A ;
- nếu $a_{ij} = a_{ji}, \forall i, j$ thì A là *ma trận đối xứng*.

Ma trận chéo

Ma trận $A = [a_{ij}]_{n \times n}$ vuông cấp n có $a_{ij} = 0, \forall i \neq j$ được gọi là ma trận chéo cấp n .

Ma trận đơn vị

Ma trận vuông cấp n có *các phần tử chéo bằng 1, tất cả các phần tử còn lại bằng 0* được gọi là ma trận đơn vị cấp n , ký hiệu I_n hoặc đơn giản là I .

Ma trận chuyển vị

Ma trận $A = [a_{ij}]_{m \times n}$. Ma trận chuyển vị của A là $A^t = [a_{ji}]_{n \times m}$,

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ \textcolor{red}{a_{i1}} & \dots & \textcolor{red}{a_{ij}} & \dots & \textcolor{red}{a_{in}} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix} \longrightarrow A^t = \begin{bmatrix} a_{11} & \dots & \textcolor{red}{a_{i1}} & \dots & a_{m1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1j} & \dots & \textcolor{red}{a_{ij}} & \dots & a_{mj} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{1n} & \dots & \textcolor{red}{a_{in}} & \dots & a_{mn} \end{bmatrix}$$

A. Vector operations

We now define the math operations for vectors. The operations we can perform on vectors $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ are: addition, subtraction, scaling, norm (length), dot product, and cross product:

$$\vec{u} + \vec{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$$

$$\vec{u} - \vec{v} = (u_1 - v_1, u_2 - v_2, u_3 - v_3)$$

$$\alpha \vec{u} = (\alpha u_1, \alpha u_2, \alpha u_3)$$

$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + u_3^2}$$

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + u_3 v_3 \quad \vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos \theta.$$

$$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1)$$

$$\vec{u} \times \vec{v} \neq \vec{v} \times \vec{u}, \text{ in fact } \vec{u} \times \vec{v} = -\vec{v} \times \vec{u}.$$

Phép cộng các ma trận cùng cấp

Cho $A = [a_{ij}]_{m \times n}$ và $B = [b_{ij}]_{m \times n}$. Tổng của hai ma trận A và B là ma trận $C = A + B = [c_{ij}]_{m \times n}$ với $c_{ij} = a_{ij} + b_{ij}$

Phép nhân một số thực với một ma trận

Cho $\alpha \in \mathbb{R}$ và $A = [a_{ij}]_{m \times n}$. Tích của số thực α với ma trận A là ma trận cùng cấp với A , ký hiệu αA , có phần tử ở hàng i cột j là αa_{ij} .

Phép nhân ma trận hàng với ma trận cột

Cho $A = [a_{1j}]_{1 \times n}$ và $B = [b_{j1}]_{n \times 1}$. Tích của ma trận hàng A và ma trận cột B (theo thứ tự đó) là ma trận ký hiệu $A.B$ (hoặc AB) có duy nhất một phần tử là

$$c = a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} = \sum_{j=1}^n a_{1j}b_{j1}$$

Phép nhân hai ma trận

Cho $A = [a_{ij}]_{m \times n}$ và $B = [b_{jk}]_{n \times p}$ ($m, n, p \in \mathbb{N}$). Tích của ma trận A và ma trận B (theo thứ tự đó) là ma trận ký hiệu $A.B$ (hoặc AB) cấp $m \times p$, với phần tử ở hàng i cột k là

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^n a_{ij}b_{jk}$$

Chú ý

- Chỉ có tích hai ma trận AB khi
số cột của ma trận A = số hàng của ma trận B
- Khi A là ma trận cấp $m \times n$, B là ma trận cấp $n \times p$ thì tích AB là ma trận cấp $m \times p$.
- Tích của hai ma trận không có tính chất giao hoán.

Một số tính chất

- 1 $A + (B + C) = (A + B) + C, \forall A, B, C$ là các ma trận cùng cấp.
- 2 $A + B = B + A, \forall A, B$ là hai ma trận cùng cấp
- 3 $A + O = O + A = A, \forall A$
- 4 $\alpha(A + B) = \alpha A + \alpha B, \forall A, B$ cùng cấp và $\alpha \in \mathbb{R}$
- 5 $(\alpha + \beta)A = \alpha A + \beta A, \forall A, \forall \alpha, \beta \in \mathbb{R}$
- 6 $0.A = O; 1.A = A, \forall A$
- 7 $A(BC) = (AB)C, \forall A, B, C$ thỏa mãn điều kiện tồn tại tích ma trận.
- 8 $A(B + C) = AB + AC, \forall A, B, C$ thỏa mãn điều kiện tồn tại tổng, tích ma trận.
- 9 $O_{m \times n}.A_{n \times p} = O_{m \times p}; A_{m \times n}.O_{n \times p} = O_{m \times p}$
- 10 $I_m.A_{m \times n} = A_{m \times n} = A_{m \times n}.I_n$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} =$$

$$\begin{pmatrix} 1 & 2 & -1 \\ 3 & 2 & 0 \\ -4 & 0 & 2 \end{pmatrix} \begin{pmatrix} 3 & 4 & 2 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

Định nghĩa

Cho ma trận A vuông cấp n , I là ma trận đơn vị cấp n .

Nếu có ma trận B vuông cấp n sao cho $AB = BA = I$ thì nói ma trận A **khả nghịch** và gọi B là **ma trận nghịch đảo** của ma trận A .

Ký hiệu ma trận nghịch đảo của ma trận A là A^{-1} .

Ví dụ 1: Cho $A = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$ và $B = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$. Tính tích AB và BA .

Từ đó cho biết ma trận A có khả nghịch không? Chỉ ra ma trận nghịch đảo (nếu có) của ma trận A .

Định nghĩa định thức cấp n

Cho ma trận A vuông cấp n , $A = [a_{ij}]_{n \times n}$.

Định thức của ma trận A , ký hiệu $|A|$ hoặc $\det A$ hoặc $\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$, là

số cho bởi công thức

$$|A| = \det A = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \sum_{j=1}^n (-1)^{1+j} a_{1j} \det M_{1j}$$

(Nếu $\det A \neq 0$ ta nói ma trận A không suy biến)

2. Tính chất

❶ **Ma trận nghịch đảo** của ma trận vuông A (nếu có) là **duy nhất**.

❷ Nếu ma trận A vuông cấp n có ma trận nghịch đảo A^{-1} thì

$$\det A \neq 0 \text{ và } \det A^{-1} = \frac{1}{\det A}$$

❸ Nếu ma trận A vuông có $\det A \neq 0$ thì A khả nghịch và

$$A^{-1} = \frac{1}{\det A} A^* = \frac{1}{\det A} \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}$$

Ma trận $A^* = [A_{ij}]^t$ được gọi là **ma trận phụ hợp** của ma trận A .

❹ Nếu A, B là hai ma trận vuông cùng cấp khả nghịch thì tích AB cũng khả nghịch và

$$(AB)^{-1} = B^{-1}A^{-1}$$

❺ Nếu A, B là hai ma trận vuông cùng cấp thỏa mãn $AB = I$ (hoặc $BA = I$) thì A khả nghịch và $B = A^{-1}$.

3. Cách tìm ma trận nghịch đảo

$A = [a_{ij}]$ vuông cấp n .

❶ Tính $\det A$.

Nếu $\det A = 0$, kết luận A không khả nghịch \rightarrow DỪNG.

Nếu $\det A \neq 0 \rightarrow$ chuyển sang bước 2.

❷ Tính các phần bù đại số $A_{ij} = (-1)^{i+j} \det M_{ij}$ của các phần tử a_{ij} .
Lập ma trận phụ hợp

$$A^* = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}$$

Suy ra

$$A^{-1} = \frac{1}{\det A} A^*$$

The multiple linear regression problem

Consider the body data again. To construct a more accurate model for predicting the weight of an individual (y), we may want to add other body measurements, such as head and waist circumferences, as additional predictors besides height (x_1), leading to multiple linear regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \quad (1)$$

where

- y : response, x_1, \dots, x_k : predictors
- $\beta_0, \beta_1, \dots, \beta_k$: coefficients
- ϵ : error term

Definition 1 .

Linear Model

A model is considered to be *linear* if it is linear in its *parameters*.

Consider the following models, and why they are linear/non-linear:

- $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$ is linear;
- $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}^2 + \epsilon_i$ is linear;
- $\log Y = \log C_0 L^\alpha K^\beta + \epsilon$ is considered linear since we can transform the model into $\log Y = C_0 + \alpha \log L + \beta \log K + \epsilon_i$;
- $Y = \frac{\beta_1}{\beta_1 - \beta_2} [e^{-\beta_2 X} - e^{\beta_1 X}] + \epsilon$ is non-linear.

A linear model therefore has the following underlying assumptions:

- The model form can be written $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \epsilon_i$ for all i ;
- $E[\epsilon_i] = 0$ for all i ;
- $\text{Var}(\epsilon_i) = \sigma^2$ for all i ;
- $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$.

An example of a regression model with $k = 2$ predictors

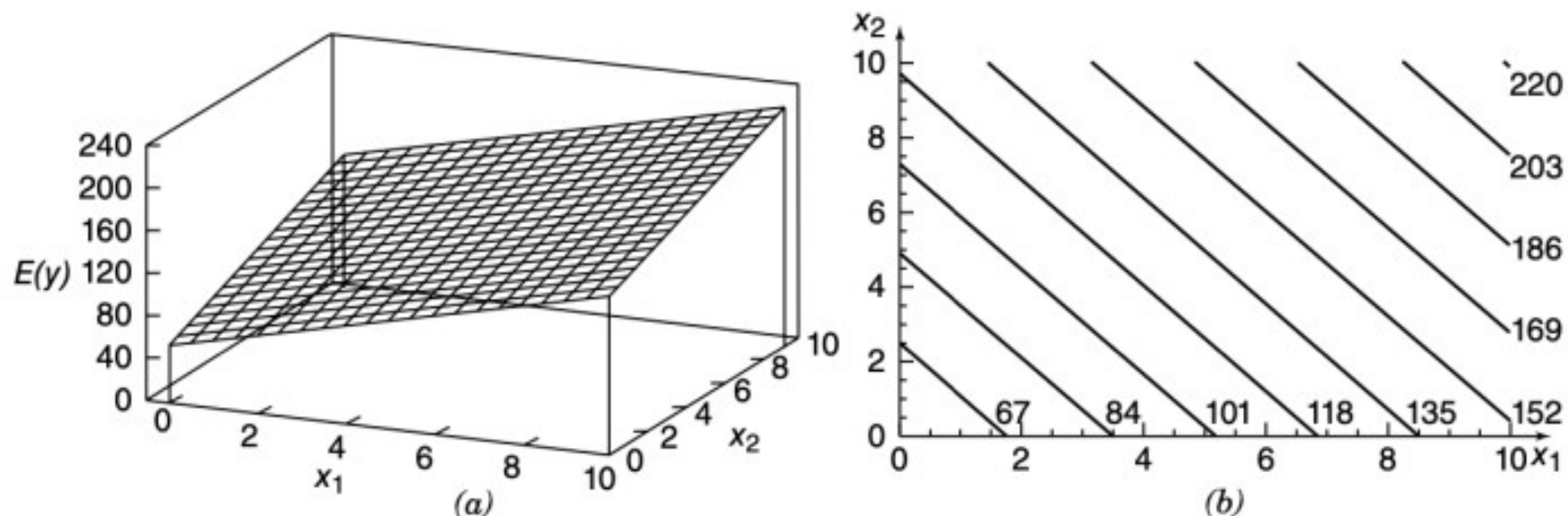


Figure 3.1 (a) The regression plane for the model $E(y) = 50 + 10x_1 + 7x_2$. (b) The contour plot.

An example of a full quadratic model

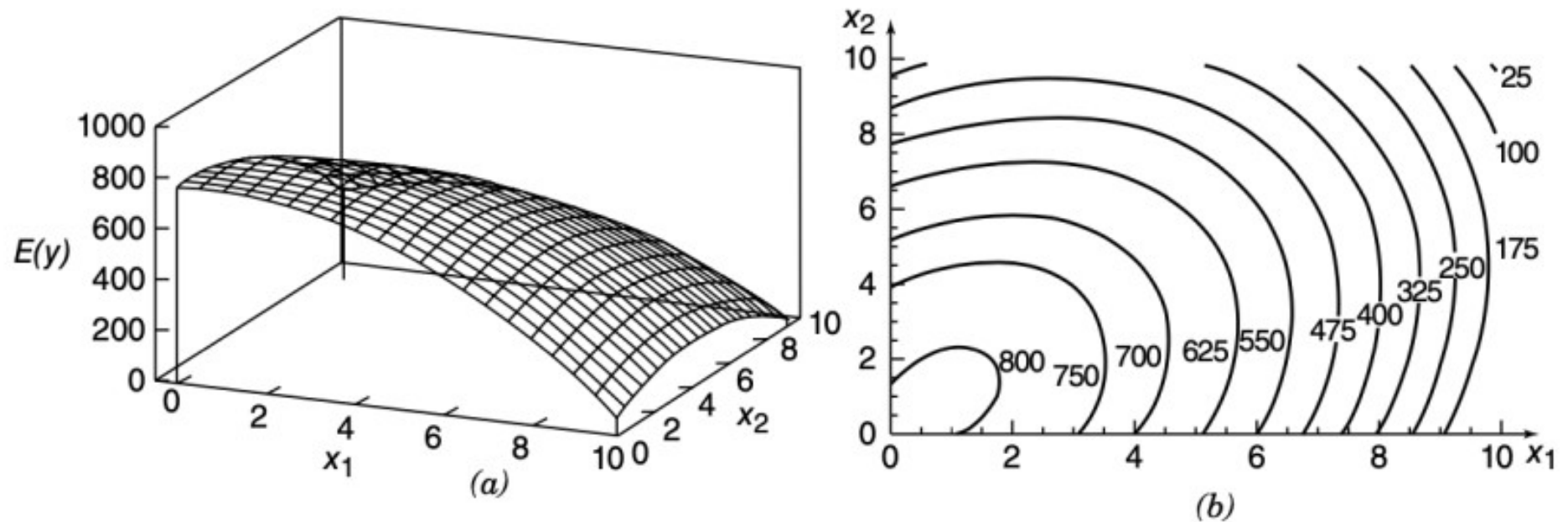


Figure 3.3 (a) Three-dimensional plot of the regression model $E(y) = 800 + 10x_1 + 7x_2 - 8.5x_1^2 - 5x_2^2 + 4x_1x_2$, (b) The contour plot.

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20

Suppose we have the following dataset with one response variable y (*food price: USD / ton*) and two predictor variables X_1 (gold price: USD / milli ounce) and X_2 (fuel price: USD / gallon)

Letting

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

we can rewrite the sample regression model in matrix form

$$\underbrace{\mathbf{y}}_{n \times 1} = \underbrace{\mathbf{X}}_{n \times p} \cdot \underbrace{\boldsymbol{\beta}}_{p \times 1} + \underbrace{\boldsymbol{\epsilon}}_{n \times 1} \quad (3)$$

where $p = k + 1$ represents the number of regression parameters (note that k is the number of predictors in the model).

Least squares (LS) estimation

The LS criterion can still be used to fit a multiple regression model

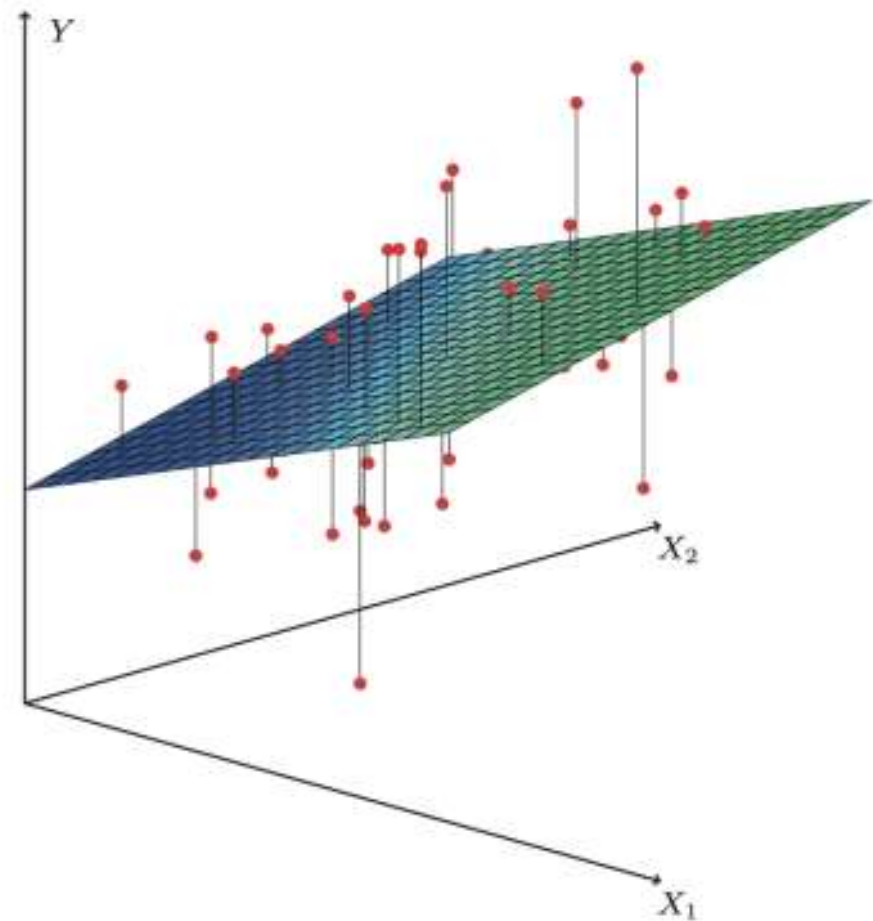
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

to the data as follows:

$$\min_{\hat{\beta}} S(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$$

where for each $1 \leq i \leq n$,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$$



Let $\mathbf{e} = (e_i) \in \mathbb{R}^n$ and $\hat{\mathbf{y}} = (\hat{y}_i) = \mathbf{X}\hat{\boldsymbol{\beta}} \in \mathbb{R}^n$. Then $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$.
Correspondingly the above problem becomes

$$\min_{\hat{\boldsymbol{\beta}}} S(\hat{\boldsymbol{\beta}}) = \|\mathbf{e}\|^2 = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$$

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}'\mathbf{v}$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}'\mathbf{a}) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{a}'\mathbf{x}) = \mathbf{a}$$

$$\frac{\partial}{\partial \mathbf{x}} (\|\mathbf{x}\|^2) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}'\mathbf{x}) = 2\mathbf{x}$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}'\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$$

$$\frac{\partial}{\partial \mathbf{x}} (\|\mathbf{B}\mathbf{x}\|^2) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}'\mathbf{B}'\mathbf{B}\mathbf{x}) = 2\mathbf{B}'\mathbf{B}\mathbf{x}$$

fuel	food
0	1
1	0
2	2

y	X₁	X₂
140	60	22
155	62	25
159	67	24
179	70	20

Suppose we have the following dataset with one response variable y (*food price: USD / ton*) and two predictor variables X_1 (gold price: USD / milli ounce) and X_2 (fuel price: USD / gallon)

THANK YOU