Linear Algebra

B Banerjee

Regression Analysis
Linear Algebra

Buddhananda Banerjee

Department of Mathematics
Centre for Excellence in Artificial Intelligence
Indian Institute of Technology Kharagpur

bbanerjee@maths.iitkgp.ac.in

Simple linear regression with Vector notation

- Consider a data set $D = \{(x_i, y_i) | x_i \in \mathbb{R}, y_i \in \mathbb{R}, \ \forall i = 1, 2, \cdots, n\}$
- x_is are non stochastic
- y_is are stochastic and realized values of random variable Y_is
- $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and $\mathbf{1} = (1, 1, \ldots, 1)^T$

Problem statement (Redefined)

We are interested to have a prediction vector

$$\hat{\mathbf{y}} = g(\mathbf{x}, \boldsymbol{\beta}) = [\mathbf{1} \ \ \mathbf{x}] \, \boldsymbol{\beta}$$

which will approximate well the observed vector $\mathbf{y}$ for known vector $\mathbf{x}$.

It is a problem in $\mathbb{R}^n$ now !!

Other uses of vector representation

- Weighted sum / Averaging
- Expectation of discrete random variable
- Combing audio signals for music composition
- Image representation in pic-cell.
- Principal component Analysis
- $\mathbb{P}_n$ = Polynomial up to degree n

Vector Space $(V, +, \cdot)$

Definition

A vector space $\mathbf{V}$ over real numbers $\mathbb{R}$ is a collection of vectors such that

1. $+ : \mathbf{V} \times \mathbf{V} \to \mathbf{V}$ [closed under vector addition]
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$, for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{V}$ [associative]
3. There exists $\mathbf{0} \in \mathbf{V}$ such that
   $\mathbf{0} + \mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in \mathbf{V}$ [identity element exists]
4. There exists $-\mathbf{x} \in \mathbf{V}$ for each $\mathbf{x}$ such that
   $(-\mathbf{x}) + \mathbf{x} = \mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ [inverse exists]
5. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ [commutative]
6. $a \cdot (b \cdot \mathbf{x}) = (ab) \cdot \mathbf{x}$ for all $a, b \in \mathbb{R}$ and $\mathbf{x} \in \mathbf{V}$
7. $1 \cdot \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbf{V}$
8. $(a + b) \cdot \mathbf{x} = (a \cdot \mathbf{x}) + (b \cdot \mathbf{x})$ for all $a, b \in \mathbb{R}$ and $\mathbf{x} \in \mathbf{V}$
9. $a \cdot (\mathbf{x} + \mathbf{y}) = a \cdot (\mathbf{x}) + a \cdot (\mathbf{y})$

Sub-Space $(S, +, \cdot)$

Definition

If a subset **S** of **V** is a vector space itself then **S** is celled subspace of **V**.

How to check S is a subspace of V?

(1) Whether $\mathbf{0} \in \mathbf{S}$?

(2) Whether $\mathbf{x} + a \cdot \mathbf{y} \in \mathbf{S}$? for all $\mathbf{x}, \mathbf{y} \in \mathbf{S}$ and $a \in \mathbb{R}$.

Example:

(1) All lines passing through $(0, 0)$ in $\mathbb{R}^2$.

(2) All planes passing through origin in $\mathbb{R}^n$.

(3) $\mathbb{P}_5$ in $\mathbb{P}_7$

Span

Definition

The span of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\} \in \mathbf{V}$ is the collection

$$Sp\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\} = \left\{ \sum_{i=1}^{k} c_i \mathbf{v}_i | c_i \in \mathbb{R} \right\}$$

which is the collection of all possible linear combinations of $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\}$.

Note: A span is always a subspace.

Example :

(a) $Sp\{(0,1),(1,1)\} = Sp\{(0,1),(1,0)\} = \mathbb{R}^2$

(b) $Sp\{(0,1,0),(1,1,0)\} = \mathbb{R} \times \mathbb{R} \times \{0\} = xy$ - pane in $\mathbb{R}^3$

In regression $\hat{\mathbf{y}} \in Sp\{\mathbf{1}, \mathbf{x}\}$ which is closest to $\mathbf{y} \in \mathbb{R}^n$

Linear Algebra

B Banerjee