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 Independence

Definition

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\} \in \mathbf{V}$ are said to be linearly independent iff $\sum_{i=1}^{k} c_i\mathbf{v}_i = \mathbf{0} \implies c_1 = c_2 = \cdots = c_n = 0$. On the other hand if $\sum_{i=1}^{k} c_i\mathbf{v}_i = \mathbf{0}$ holds for some non zero $c_i \in \mathbb{R}$ the the vectors are called linearly dependent.

Example :
(a) $\{(0, 1), (1, 1)\}$ are independent
(b) $\{(0, 1), (1, 0)\}$ are independent
(c) $\{(0, 1), (1, 0), (1, 1)\}$ are dependent

Basis

Definition

If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\}$ are linearly independent then it is a basis of $Sp\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\}$, and the dimension of $Sp\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\}$ is the number of linearly independent elements in $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\}$.

Example :
(a) $\{(0,1),(1,1)\}$ is a basis of $\mathbb{R}^2$
(b) $\{(0,1),(1,0)\}$ is a basis of $\mathbb{R}^2$ also.
(c) $\{(0,1),(1,0),(1,1)\}$ is NOT a basis of $\mathbb{R}^2$

Note: Number of vectors in a basis of a vector space is known as the dimension of the vector space.

Definition

Orthogonal vectors

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbf{V}$ are said to be orthogonal if $\mathbf{u}^T\mathbf{v} = \sum_i u_i v_i = 0$

Orthogonal complement

If $\mathbf{S} \subseteq \mathbf{V}$ is a subspace then the orthogonal complement of $\mathbf{S}$ denoted by $\mathbf{S}^\perp$ is a collection

$$\mathbf{S}^\perp = \{\mathbf{v} | \mathbf{v} \in \mathbf{V}, \mathbf{u}^T\mathbf{v} = 0, \forall \mathbf{u} \in S\}$$

and $dim(\mathbf{S}^\perp) = dim(\mathbf{V}) - dim(\mathbf{S})$.

Examples:
(a) $Sp\{(1,0,0,0),(0,0,1,0)\} \perp Sp\{(0,1,0,0),(0,0,0,1)\}$
(b) $Sp\{(1,1,0,0),(0,1,1,0),(1,0,1,0),\} \perp Sp\{(0,0,0,1)\}$

Remarks

1. Basis is not unique.
2. Elements of a basis are need not be orthogonal to each other.
3. Linear independence need not imply orthogonality.
4. Orthogonality implies independence.
5. Orthogonal vectors with unit length are called orthonormal vectors.

Projection

Projection Matrix

If $\mathbf{S} \subseteq \mathbf{V}$ then the projection matrix of subspace $\mathbf{S}$ is P_s satisfying
(a) $P_s\mathbf{v} = \mathbf{v}$ if $\mathbf{v} \in \mathbf{S}$
(b) $P_s\mathbf{v} \in \mathbf{S}$ for all $\mathbf{v} \in \mathbf{V}$

Orthogonal Projection Matrix

A projection matrix P_s is an orthogonal projection matrix of subspace $\mathbf{S} \subseteq \mathbf{V}$ if $(\mathbf{I} - P_s)$ is a projection matrix of $\mathbf{S}^{\perp} \subseteq \mathbf{V}$ too.

Theorem

If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_k\}$ is an orthonormal basis of the subspace $\mathbf{S} \subseteq \mathbf{V}$ then the orthogonal projection matrix of $\mathbf{S}$ is $P_s = \sum_{i=1}^{k} v_i v_i^T$

Idempotency

Idempotent matrix

If a matrix P satisfies the relation that $P^2 = P$, then P is called an idempotent matrix.

Theorem

An idempotent matrix has eigen values 0 and 1.

Theorem

A projection matrix is an idempotent matrix.

In regression eventually $\hat{\mathbf{y}}$ becomes
the orthogonal projection of $\mathbf{y} \in \mathbb{R}^n$ in the subspace $S = Sp\{\mathbf{1}, \mathbf{x}\}$

Column Space

Definition

The column space of a matrix $A = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n]$ with columns $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ is

$$\mathcal{C}(A) = Sp\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\} = \{A\mathbf{x} |\ \mathbf{x} \in \mathbb{R}^n.\}$$

Hence, row-space of A denoted by $\mathcal{R}(A) = \mathcal{C}(A^T)$.

Properties :

1. $\mathcal{C}(A : B) = \mathcal{C}(A) + \mathcal{C}(B)$
2. $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$
3. $dim(\mathcal{C}(A)) = Rank(A)$
4. $\mathcal{C}(AA^T) = \mathcal{C}(A) \implies Rank(AA^T) = Rank(A)$
5. If A has n-rows then $dim(\mathcal{C}(A)^{\perp}) = n - Rank(A)$

Quadratic forms

Definition

A square matrix $\mathbf{A} = ((A_{ij}))_{n\times n}$ is said to be
(a) **positive definite (p.d.)** if

$$\mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x} > 0 \ \text{ for all } \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n.$$

(b) **positive semi-definite (p.s.d)** if

$$\mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x} \geq 0 \ \text{ for all } \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n.$$

[Also called non-negative definite (n.n.d.)]

Properties:
(a) If $\mathbf{A}$ is p.d. then $|\mathbf{A}| > 0$.
(b) If $\mathbf{A}$ is p.s.d. then $|\mathbf{A}| \geq 0$.

Generalized inverse

Definition

A matrix G is said to be a generalize inverse of a matrix A if $AGA = A$. Usually G is denoted by A^-.

Properties:

(1)If A is $m \times n$ then A^- is $n \times m$.

(2) A^- is not unique.

(3) For a matrix A the projection matrix of $\mathcal{C}(A)$ is AA^-

(4) For a matrix A **the orthogonal projection matrix of** $\mathcal{C}(A)$ **is**

$$A(A^TA)^-A^T.$$

In regression the prediction $\hat{\mathbf{y}}$ and
the error $\mathbf{y} - \hat{\mathbf{y}}$ are orthogonal to each other.

Reference

Jammalamadaka, S. Rao, and Debasis Sengupta. Linear models: an integrated approach. Vol. 6. World Scientific, 2003. Chapter 2

Bapat, Ravindra B. Linear algebra and linear models. Springer Science & Business Media, 2012. Chapter 1

Linear Algebra

B Banerjee