Linear Algebra

Pritha Banerjee

University of Calcutta

banerjee.pritha74@gmail.com

March 19, 2025

Overview

Linear Independence (1/2)

- Given a dataset in the form of a matrix A of size $n \times p$, where n is the number of observations and p the number of attributes, identify the **number of linear relationships between attributes** from the data.
- **Linear combination of vectors:** Linear combination of vectors $\vec{a_1}, \vec{a_2}, \cdots \vec{a_n} \in V$ with scalars $x_1, x_2, \cdots, x_n \in S$ in vector space $(V, S, +, .)$ is the vector $b = x_1\vec{a_1} + x_2\vec{a_2} + \cdots + x_n\vec{a_n}$
- **Linear dependence:** The vectors $\vec{a_i} \in V, i = 1 \cdots n$ are **linearly dependent** if there exist n scalars $x_i \in S, i = 1 \cdots n$, at least one of them non-zero such that $x_1\vec{a_1} + x_2\vec{a_2} + \cdots + x_n\vec{a_n} = 0$, i.e,

  $$a_n = \sum_{i=1}^{n-1} x_i.a_i$$

  $$A = [\vec{a_1}\,\vec{a_2} \cdots \vec{a_n}]\,;\, x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- The vectors $\vec{a_1}, \vec{a_2}, \cdots \vec{a_n} \in V$ are **linearly independent** if scalars $x_1, x_2, \cdots, x_n \in S$ are all zero that satisfy $x_1\vec{a_1} + x_2\vec{a_2} + \cdots + x_n\vec{a_n} = 0$; In other words, $Ax = 0 \Rightarrow x = 0$

Linear Independence and Rank (2/2

- The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a **Linearly Independent** set.
- The **row rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of rows of A that constitute a **Linearly Independent** set.
- **column rank** of $A =$ **row rank** of A
- It helps to work with reduced set of variables as dependent attributes can be computed from the linear relation
- It is independent of size of dataset if data is taken from same data generation process
- Rank of following matrix $= 2$; since $C_2 = C_2 - 2C_1$ makes C_2 to be $[000]^T$.

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 0 \\ 3 & 6 & 0 \end{bmatrix}$$

- Finding rank of matrix: Convert Matrix into Echolon form using row/column transformation. The number of nonzero rows is the rank of matrix

Properties of Rank

- For $A \in \mathbb{R}^{m \times n}$, $rank(A) \leq min(m, n)$
- if $rank(A) = min(m, n)$, then A is said to be full rank.
- For $A \in \mathbb{R}^{m \times n}$, $rank(A) = rank(A^T)$
- For $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p} rank(AB) \leq min(rank(A), rank(B)$
- For $A, B \in \mathbb{R}^{m \times n}$, $rank(A + B) \leq rank(A) + rank(B)$

Row space and column space of matrix A

- **Vector space:** A vector space $(V, \mathbb{R}, +, .)$ is a set of vectors V with operations $+$ and $.$ that satisfies closure properties and eight axioms.
- **Subspace:** A subset W of vector space V if W itself satisfies closures under addition and multiplication;W is a subspace of V iff W is non-empty and $x + \alpha y \in W$, $\forall x, y \in W, \alpha \in R$
- **Span:** The span of a set of vectors $X = x_1, x_2, \cdots x_n$ from a Vector space is the subspace consisting of all linear combinations of the vectors in the set X; $v = \sum_{i=1}^{n} \alpha_i x_i, \alpha_i \in \mathbb{R}$
- Let A is $m \times n$ matrix. The space spanned by row vectors of A is **row space** of A, which is a subspace of $\mathbb{R}^n$
- The space spanned by columns of A is **range or column space** of A, which is a subspace of $\mathbb{R}^m$
- rank(A) $=$ dim(row space(A)) $=$ dim(column space(A))

Norm

A function $f : \mathbb{R}^n \to \mathbb{R}$ satisfying:

- $\forall x \in \mathbb{R}^n, f(x) \geq 0$ (non-negativity)
- $f(x) = 0$ iff $x = 0$ (definiteness)
- $\forall x \in \mathbb{R}^n, f(tx) = |t| f(x)$ (homogeneity)
- $\forall x, y \in \mathbb{R}^n, f(x+y) \leq f(x) + f(y)$ (triangle inequality)

Example

- l_p Norm: $||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$; $||x||_2 = (\sum_{i=1}^{n} |x_i|^2)^{1/2}$
- Norms for Matrix: Frobenius Norm $||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{tr(A^T A)}$, $tr()$: sum of diagonals of a square matrix.

Identification of linear relationship among attributes

- Number of independent variables $\leq$ number of attributes
- Null space and Nullity identifies linear relations among attributes
- **Null space $N(A)$:** Null space of matrix $A \in \mathbb{R}^{n \times p}$ consists of all vectors $\vec{\beta}$ such that $A\vec{\beta} = 0$ and $\vec{\beta} \neq 0$, i.e,if $A = [\vec{x_1}\vec{x_2} \cdots \vec{x_p}]$, $\vec{\beta}$ satisfies

$$x_{11}\beta_1 + x_{12}\beta_2 + \cdots + x_{1p}\beta_p = 0$$
$$x_{21}\beta_1 + x_{22}\beta_2 + \cdots + x_{2p}\beta_p = 0$$
$$\cdots$$
$$x_{n1}\beta_1 + x_{n2}\beta_2 + \cdots + x_{np}\beta_p = 0$$

- **Nullity of matrix:** the number of vectors in the null space of matrix
- size of Null space gives the number of linear relations among attributes and each null space vector β is used to identify one linear relationship.
- **Rank Nullity theorem:** Nullity of A + Rank(A) = Total number of columns in A
- Example: Nullity of $A = 3 - 2 = 1$

Computing Null space of a matrix, Linear relationship

- Solve linear equation $A\vec{\beta} = 0$, Given A. Example:

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \\ 3 & 6 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$\beta_1 + 2\beta_2 = 0 \cdots \times 3$
$2.\beta_1 + 4\beta_2 = 0 \cdots \times 3$
$\beta_3 = 0; \beta_1 = -2\beta_2$

$$\begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^\top = \begin{bmatrix} -2.\beta_2 & \beta_2 & 0 \end{bmatrix}^\top = k \begin{bmatrix} -2 & 1 & 0 \end{bmatrix}^\top$$

- Different k gives null space vector.

Solving System of Linear Equations

- Let S a system of linear equation $Ax = b$ represented as augmented matrix $[A|b]$. Let the size of b, A and $[A|b]$ are $m \times 1, m \times n, m \times (n+1)$, then
  - S has no solution (inconsistent) iff $rank(A) \leq rank([A|b])$; approximate solution $(A^T.A)^{-1}A^Tb$
  - S has a unique solution iff rank(A) $=$ rank($[A|b]$) $=$ n; $x = A^{-1}b$
  - S has infinitely many solutions iff rank(A) $=$ rank($[A|b]$) $\leq n$; approximate solution $x = A^T(A.A^T)^{-1}b$

Unit vector, orthogonal vector

Unit Vector

- A unit vector is a vector with magnitude 1 (distance from origin)
- Unit vectors are used to define direction in coordinate system
- Any vector can be written as product of unit vector and scalar magnitude
- Example: $A = [3 \quad 4]^T$, unit vector $\hat{a} = \frac{A}{|A|}$
- $|A| = \sqrt{3^2 + 4^2}$, Thus $\hat{a} = [3/5 \quad 4/5]^T$

Orthogonal Vector

- Two vectors $A = [a_1 \quad a_2 \quad \cdots a_n]$ and $B = [b_1 \quad b_2 \quad \cdots b_n]$ are orthogonal to each other when their dot product $A.B = \sum_{i=1}^{n} a_i b_i = A^T B = 0$
- Example: vectors $v_1 = [1 \quad -2 \quad 4]$ and $v_2 = [2 \quad 5 \quad 2]$ are orthogonal since $v_1.{v_2}^T = 2 - 10 + 8 = 0$
- Orthogonal vectors with unit magnitude is called **orthonormal vectors**
- Example: $v_1 = \left[\frac{1}{\sqrt{21}} \quad \frac{-2}{\sqrt{21}} \quad \frac{4}{\sqrt{21}}\right]$ and $v_2 = \left[\frac{2}{\sqrt{33}} \quad \frac{5}{\sqrt{33}} \quad \frac{2}{\sqrt{33}}\right]$

Basis vector

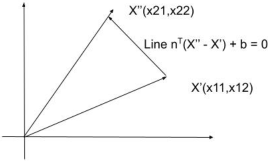
- **Basis vectors** are set of vectors that are linearly independent and span the space. i.e, any vector can be represented as a linear combination of basis vactors. Ex: [1 0] and [0 1] are basis vectors that span $\mathbb{R}^2$
- Basis vectors that span a space are not unique. However number of basis vectors that span a space is unique.
- To span $\mathbb{R}^2$, two basis vectors are required, for $\mathbb{R}^3$, three basis vectors are needed. To span a subspace, lesser number of basis vectors may be sufficient.
- **Rank of a matrix** gives a set of **linearly independent vectors** that are the **basis vectors** that **spans row/ column space of the matrix**.
- Basis vectors help in storage of large matrices. Store only linearly independent vectors and the scalar multiples to represent evey other vector of the matrix.
- Ex. matrix $A_{400\times 4}$ has 400 sample data with 4 attributes. Let rank of this matrix be 2. Then we need to store
  - $2 \times 4 = 8$ values for 2 basis vectors each having 4 attributes.
  - $2 \times 398 = 796$ scalar values to generate 398 other sample data, using 2 basis vectors
  - Total savings: $1600 - (796 + 8)$, approx. 50% savings of storage

Representation of line and plane

- General equation of a line: $n_1 x_1 + n_2 x_2 + b = 0$

$$\begin{bmatrix} n_1 & n_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b = 0 = n^T X + b = 0$$

- Point $X'(x_11, x_12$ and $X''(x21, x22)$ lies on line: $n^T X' + b = 0$ and $n^T X'' + b = 0$
- Subtracting 2 from 1, $n^T(X'' - X') = 0 \implies$ **orthogonal vectors**, n (normal vector) is perpendicular to $(X'' - X')$
- in 2D, $n^T = b = 0$ represent a line, where n is normal to the line.
- in 3D, $n^T = b = 0$ represent a plane, where n is normal to the plane.

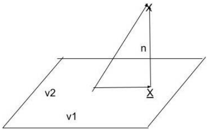


Projection

- Want to represent data through smaller set. Projection is to approximate the data through smaller set.
- 2 basis vector is used to represent a plane, ie. any point on the plane is linear combination of the basis vectors representing the plane
- Any line on the plane is linear combination of basis vectors v_1, v_2
- **Projection:** projection $\hat{X}$ of a vector X onto a lower dimension is $\hat{X} = c_1.v_1 + c_2.v_2$. Need to obtain the scalars c_1, c_2
- Use vector additin: $X = c_1.v_1 + c_2.v_2 + n$ such that $n^T v_1 = v_1^T n = 0$ and $n^T v_2 = v_2^T n = 0$ (since orthogonal)
- Projection onto general orthogonal direction: substitute X in $v_1^T n = 0$

$$v_1^T (X - c_1.v_1 - c_2.v_2) = 0$$

$$v_1^T X = c_1 v_1^T v_1 \implies c_1 = \frac{v_1^T X}{v_1^T v_1}$$



- $\hat{X} = \frac{v_1^T X}{v_1^T v_1}.v_1 + \frac{v_2^T X}{v_2^T v_2} v_2$, using $v_2^T n = 0$ to obtain c_2

Generalized Projection

- Consider projection of X onto a space spanned by k linearly independent vectors. $\hat{X} = \sum_{j=1}^{k} c_j v_j, X \in \mathbb{R}^n, v_i \in \mathbb{R}^n$
- $\hat{X} = Vc, V = [v_1 \; v_2 \; \cdots v_k]_{n \times k}$ and $c = [c_1 \; c_2 \; \cdots c_k]_{k \times 1}$
- using orthogonality $X = \hat{X} + n \implies n^T(X - \hat{X}) = 0$:

$$V^T(X - \hat{X}) = V^T(X - Vc) = 0$$

$$V^T X - V^T V c = 0$$

$$c = (V^T V)^{-1} V^T X$$

inverse exists as V are basis vectors, thus linearly independent, has non-zero rank.

$$\hat{X} = V(V^T V)^{-1} V^T X$$

Projection: Example

- Project $X = [1\ 2\ 3]^T$ onto space spanned by $v_1 = [1\ \ -1\ \ -2]^T$ and $v_2 = [2\ 0\ 1]^T$
- $\hat{X} =$

$$\frac{1}{6}\begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}-1\\-1\\-2\end{bmatrix}\begin{bmatrix}1\\-1\\-2\end{bmatrix}+\frac{1}{5}\begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}2\\0\\1\end{bmatrix}\begin{bmatrix}2\\0\\1\end{bmatrix}$$

- $\hat{X} =$

$$\begin{bmatrix}5/6\\7/6\\20/6\end{bmatrix}$$

Orthogonal Matrices

- A square matrix $U \in \mathbb{R}^{n \times n}$, is **orthogonal** iff
  - all columns are mutually orthogonal $v_i^T v_j = 0 \forall v_i \neq v_j$
  - All columns are normalized, $v_i^T v_i = 1, \forall i$
- Orthogonal matrix do not change the Euclidean Norm of a vector when the matrix operate on it. $||U_x||_2 = ||x||_2$
- multiplication by an orthogonal matrix is like a rotation, that only changes the direction of the vector, but not the magnitude.

Quadratic form of Matrices and their properties

- Given a square matrix $U \in \mathbb{R}^{n \times n}$ and a vector $x \in \mathbb{R}^n$, scalar value $x^T Ax$ is called quadratic form
- A **symmetric matrix** A is **positive definite (negative)**, if for all non-zero vectors $x \in \mathbb{R}^n, x^T Ax > 0 (< 0)$
- A **symmetric matrix** A is **positive semi definite (negative)**, if for all non-zero vectors $x \in \mathbb{R}^n, x^T Ax \geq 0 (\leq 0)$
- **Positive (negative) definite** matrices are always **full rank**, hence **inverse** exists
- **Gram Matrix:** Given any matrix $A \in \mathbb{R}^{m \times n}$, matrix $G = A^T A$ is always **positive semidefinite**
- if $m \geq n$, G is positive definite