Mathematics for Data Science

Dr. Indu Joshi

Assistant Professor at
Indian Institute of Technology Mandi

29 July 2024

Course Logistics

Major- 35%
Minor- 25%
Assignments- 30%
Tutorial and class participation- 10%

Motivation

Vectors and matrices are fundamental in machine learning due to their ability of:

- **Data Representation:** Vectors and matrices efficiently represent and organize large datasets, enabling the handling of complex and high-dimensional data structures.
- **Mathematical Operations:** They facilitate various mathematical operations, such as dot products and matrix multiplications, which are essential for algorithms like linear regression and neural networks.
- **Feature Transformation:** Vectors and matrices are used to transform features, enabling dimensionality reduction techniques like Principal Component Analysis (PCA) and feature scaling.

Motivation (contd)

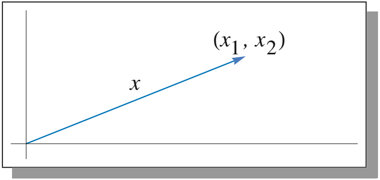
- **Model Parameters:** They store model parameters and weights, allowing for efficient computation and optimization in training processes, particularly in gradient descent and backpropagation.
- **Linear Algebra Applications:** Many machine learning algorithms rely on linear algebra concepts, which are naturally expressed using vectors and matrices, enhancing computational efficiency and algorithmic implementation.

Vector

- A *vector* is a mathematical object that encodes a length and direction.
- More formally these are elements of a vector space: a collection of objects that is closed under an addition rule and a rule for multiplication by scalars.
- A vector is often represented as a 1-dimensional array of numbers, referred to as components, and is displayed either in column form or row form.
- Represented geometrically, vectors typically represent coordinates within an n-dimensional space, where n is the number of dimensions.
- A simplistic representation of a vector might be an arrow in a vector space, with an origin, direction, and magnitude (length).

Vectors in R^n

A vector in R^n is an n-tuple $v = (v_1, v_2, ..., v_n)$, where $v_i \in R$



Elements of $\mathbf{R}^2$ can be thought of as points or as vectors.

Vector Algebra

- **Vector Addition:** Two vectors, u and v of same dimension can be added as: $u = [u_1, u_2, ..., u_n], v = [v_1, v_2, ..., v_n]$
  $u + v = [u_1 + v_1, u_2 + v_2, ..., u_n + v_n]$
  Example- In $\mathbb{R}^3$, $u = [1, 1, -1]$ and $v = [2, 3, 1]$
  $u + v = [3, 4, 0]$
- **Vector Subtraction:** Two vectors, u and v of same dimension can be subtracted as:
  $u = [u_1, u_2, ..., u_n], v = [v_1, v_2, ..., v_n]$
  $u - v = [u_1 - v_1, u_2 - v_2, ..., u_n - v_n]$
  Example- In $\mathbb{R}^3$, $u = [1, 1, -1]$ and $v = [2, 3, 1]$
  $u - v = [-1, -2, -2]$

Vector Algebra

- **Dot Product:** The dot product can be used to find the angle between two vectors. Furthermore, the dot product is useful in finding the projection of one vector onto another, i.e., to measure how much one vector extends in the direction of another.
  Dot product of two vectors, u and v of same dimension can be computed as: $u = [u_1, u_2, ..., u_n], v = [v_1, v_2, ..., v_n]$
  $u \cdot v = u_1 v_1 + u_2 v_2 + ... + u_n v_n$
  Example- In $\mathbb{R}^3$, $u = [1, 1, -1]$ and $v = [2, 3, 1]$
  $u \cdot v = 2 + 3 - 1 = 4$

Vector Algebra

- **Length/Magnitude of a Vector:** Length of a vector $v = [v_1, v_2, ..., v_n]$ is defined as:
  $|v| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + ... + v_n^2}$
  Example-
  $v = [2, 3, 1], |v| = \sqrt{2^2 + 3^2 + 1^2} = \sqrt{4 + 9 + 1} = \sqrt{14}$
- **Angle between two vectors:** The angle between two vectors u and v (θ) is defined as:
  $\theta = cos^{-1}(\frac{u \cdot v}{|u||v|})$

Linear Combination of Vectors

- Consider a set $S = \{v_1, v_2, ..., v_k\}$, then a vector $v = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_k v_k$ is called a *linear combination* of $v_1, v_2, ..., v_k$, where $\alpha_1, \alpha_2, ...\alpha_k$ are scalar.
- Example- $v_1 = [1, 2, -1], v_2 = [1, 1, 0], v_3 = [0, 1, -1]$
  $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = \alpha_1[1, 2, -1] + \alpha_2[1, 1, 0] + \alpha_3[0, 1, -1]$
  $[(\alpha_1 + \alpha_2), (2\alpha_1 + \alpha_2), (-\alpha_1 - \alpha_3)]$

Linearly Independent and Dependent Vectors

- A set of vectors $S = \{v_1, v_2, ..., v_n\}$ is linearly independent if the equation-
  $\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n = 0$
  holds only when $\alpha_1 = \alpha_2 = ...\alpha_n = 0$, otherwise the set S is linearly dependent.
- Example- $S = \{[1, 0], [1, 1]\}$ in $\mathbb{R}^2$
  $\alpha_1[1, 0] + \alpha_2[1, 1] = [0, 0] \implies \alpha_1 + \alpha_2 = 0, \alpha_2 = 0$
  $\implies \alpha_1 = 0.$ Therefore, S contains linearly independent vectors.
- $\hat{S} = \{[1, 1], [3, 3]\}$
  $(-3)[1, 1] + (1)[3, 3] = 0 \implies \alpha_1 = -3, \alpha_2 = 1.$
  Therefore, $\hat{S}$ contains linearly dependent vectors.

Linearly Independent and Dependent Vectors

- $\mathbb{R}^3 : S = \{[1, -1, 0], [1, 0, 1], [0, 1, 1]\}$
  $\alpha_1 = 1, \alpha_2 = -1, \alpha_3 = 1$
  $\alpha_1[1, -1, 0] + \alpha_2[1, 0, 1] + \alpha_3[0, 1, 1]$
  $= (1)[1, -1, 0] + (-1)[1, 0, 1] + (1)[0, 1, 1] = [0, 0, 0]$
  $v_2 = v_1 + v_3 = [1, -1, 0] + [0, 1, 1] = [1, 0, 1] = v_2$
- Example of linearly independent vectors in $\mathbb{R}^3$:
  $\{[1, 0, 0], [0, 1, 0], [0, 0, 1]\}$

Linearly Independent and Dependent Vectors

- In $\mathbb{R}^n$, a set of more than n vectors is linearly dependent.
- Any set containing zero vector is linearly dependent [because the zero vector can always be written as a non-trivial linear combination of the other vectors in the set]
  Example- Consider a set of vectors that includes the zero vector, for example, $\{v_1, v_2, 0, ..., v_n\}$
  If we take the scalar $\alpha_3 = 1$ for the zero vector 0 and set all other scalars $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$ we get:

$$0 \cdot v_1 + 0 \cdot v_2 + 1 \cdot 0 + ... + 0 \cdot v_n = 0$$

  This satisfies the condition for linear dependence because we found a non-trivial combination that results in the zero vector.

Orthogonal Vectors

- A set of vectors $\{v_1, v_2, ..., v_n\}$ are mutually (pairwise) orthogonal, if
  $v_i \cdot v_j = 0$, for $i \neq j$
- Example- $\mathbb{R}^3 : \{[1, 0, -1], [1, \sqrt{2}, 1], [1, -\sqrt{2}, 1]\}$
  $[1, 0, -1] \cdot [1, \sqrt{2}, 1] = 0$
  $[1, 0, -1] \cdot [1, -\sqrt{2}, 1] = 0$
  $[1, \sqrt{2}, 1] \cdot [1, -\sqrt{2}, 1] = 0$
- A set of orthogonal vectors is linearly independent.

Orthogonal Vectors

Let $\{v_1, v_2, ..., v_n\}$ be a set of orthogonal vectors. Suppose we have a linear combination of these vectors that equals the zero vector:

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n = 0$$

For each v_i (where i ranges from 1 to n), take the dot product of both sides of the equation with v_i:

$$v_i \cdot (\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n) = v_i \cdot 0$$

$$\alpha_1 \cdot (v_i \cdot v_1) + \alpha_2 \cdot (v_i \cdot v_2) + ... + \alpha_n \cdot (v_i \cdot v_n) = 0$$

Because the vectors are orthogonal, $v_i \cdot v_j = 0$ for $i \neq j$

$$\alpha_1 \cdot (0) + \alpha_2 \cdot (0) + ... + \alpha_i \cdot (v_i \cdot v_i) + ... + \alpha_n \cdot (v_i \cdot v_n) = 0$$

Orthogonal Vectors

$$\alpha_i \cdot (v_i \cdot v_i) = 0$$

The dot product $v_i \cdot v_i$ is the magnitude squared of v_i, which is always positive if $v_i \neq 0$

$$\alpha_i ||v_i||^2 = 0$$

Since $||v_i||^2 \geq 0 \implies \alpha_i = 0$. Therefore, each coefficient $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$. Hence, the set of orthogonal vectors is linearly independent.

Orthonormal Vectors

- A set of orthogonal vectors is orthonormal if each vector has length 1.
- Example- $\{[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}], [\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}]\}$

Matrix

- A matrix is a two-dimensional array of scalars with one or more columns and one or more rows.

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

- Usually we denote the elements of a matrix by $\{a_{ij}\}_{m \times n}$, where i is the index representing rows and varies from $1, 2, ..., m$, while j is the index representing columns and varies from $1, 2, ..., n$.

Diagonal and Triangular Matrix

- A square matrix whose all off-diagonal elements are zero is called a diagonal matrix. Example- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$
- A square matrix whose all elements below the main diagonal are zero is called an upper triangular matrix. Example- $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$
- A square matrix whose all elements above the main diagonal are zero is called a lower triangular matrix. Example- $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 0 & 5 \end{bmatrix}$

Identity Matrix

A diagonal matrix with all diagonal entries as 1 is called as an identity matrix. Example- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Matrix Algebra

Two matrices are called equal if their dimensions as well as all the corresponding elements are equal. Example- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ The last element (a_{33}) is different in the two matrices.

Therefore, these are not equal matrices.

Matrix Algebra

- The addition or subtraction of two Matrices X and Y of the same size returns a matrix Z of the same size

$$z_{ij} = x_{ij} \pm y_{ij} \quad \forall i, j$$

- Matrices of different sizes cannot be added or subtracted.
- **Commutativity:** $X \pm Y = Y \pm X$
- **Associativity:** $X \pm (Y \pm Z) = (X \pm Y) \pm Z = X \pm Y \pm Z$

Matrix Algebra

$$X = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 2 \end{bmatrix}$$

$$X + Y = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 2 & 0 \\ 5 & 3 & 3 \end{bmatrix} \quad X - Y = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 5 & -3 & -1 \end{bmatrix}$$

Scalar Multiplication

$$X = \begin{bmatrix} x_{11} & x_{12} \dots & x_{1n} \\ x_{21} & x_{22} \dots & x_{2n} \\ \dots & \dots & \dots \\ x_{m1} & x_{m2} \dots & x_{mn} \end{bmatrix}$$

$$\gamma \in \mathbb{R} \ \ \mathit{or} \ \ \mathbb{C}$$

$$\gamma X = \begin{bmatrix} \gamma x_{11} & \gamma x_{12} \dots & \gamma x_{1n} \\ \gamma x_{21} & \gamma x_{22} \dots & \gamma x_{2n} \\ \dots & \dots & \dots \\ \gamma x_{m1} & \gamma x_{m2} \dots & \gamma x_{mn} \end{bmatrix}$$

$$X = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 5 & -3 & 1 \end{bmatrix}, \gamma = 5, \ \gamma X = \begin{bmatrix} 0 & 0 & 15 \\ 0 & 0 & 0 \\ 25 & -15 & 5 \end{bmatrix}$$

Matrix Multiplication

- The product of two matrices is a matrix.
- The necessary condition for multiplication of two matrices X and Y is that the number of columns in X must be equal to the number of rows in Y.

$$X_{m\times n} \times Y_{n\times p} = Z_{m\times p}$$

$$Y_{n\times p} \times X_{m\times n} \quad \text{– Multiplication Not Defined}$$

Matrix Multiplication

$$\begin{bmatrix} x_{11} & x_{12}\dots & x_{1n} \\ x_{21} & x_{22}\dots & x_{2n} \\ \dots & \dots & \dots \\ x_{m1} & x_{m2}\dots & x_{mn} \end{bmatrix} \begin{bmatrix} y_{11} & y_{12}\dots & y_{1p} \\ y_{21} & y_{22}\dots & y_{2p} \\ \dots & \dots & \dots \\ y_{n1} & y_{n2}\dots & y_{np} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12}\dots & z_{1p} \\ z_{21} & z_{22}\dots & z_{2p} \\ \dots & \dots & \dots \\ z_{m1} & z_{m2}\dots & z_{mp} \end{bmatrix}$$

$$z_{11} = x_{11}y_{11} + x_{12}y_{21} + \dots + x_{1n}y_{n1}$$

$$z_{ij} = [\mathrm{i}^{th}\ row\ of\ X] \cdot [\mathrm{j}^{th}\ column\ of\ Y]$$

Matrix Multiplication

$$X = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \; Y = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$Z = X \times Y$
$Z = \begin{bmatrix} 1 & -2 & -1 \\ 2 & -1 & 1 \end{bmatrix}$

$$z_{11} = [1 - 1] \cdot [10] = 1 - 0 = 1$$

$$z_{12} = [1 - 1] \cdot [-11] = -1 - 1 = -2$$

Transpose of a Matrix

$$X_{mn} = \{x_{ij}\} \quad \text{i= 1, 2,...m, j=1, 2, ...n}$$

$$X_{nm}^T = \{x_{ji}\}$$

$$X_{23} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \; X_{32}^T = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}$$

Transpose of a Matrix: Properties

1. $(A+B)^T = A^T + B^T$
2. $(AB)^T = B^T A^T$
3. $(kA)^T = kA^T$
4. $(A^T)^T = A$

Determinant of a Matrix

- Every square matrix has a determinant.
- Determinant of a matrix is a number.

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

$$|X| = det(X) = 3 - 2 = 1$$

Inverse of a Matrix

- Division of matrices is not defined since there may be $AB = AC$ while B and C may not be equal.
- Instead matrix inversion is used.
- Inverse of a square matrix, X, if it exists, is the unique matrix X^{-1} such that:

$$XX^{-1} = X^{-1}X = I$$

Inverse of a Matrix: Properties

1. $(AB)^{-1} = B^{-1}A^{-1}$
2. $(A^{-1})^{-1} = A$
3. $(A^{T})^{-1} = (A^{-1})^{T}$
4. $(kA)^{-1} = \frac{1}{k}A^{-1}$

- A square matrix that has an inverse is called a nonsingular matrix.
- A matrix that doesn't have an inverse is called a singular matrix.
- Square matrices have inverses except when the determinant is zero.
- When the determinant of a matrix is zero, that matrix is singular.

Thank You

Contact: indujoshi@iitmandi.ac.in