Section 2: Advanced Topics in Linear Algebra

Vector Spaces and Subspaces

Definition and Properties

Vector Space:
A vector space V is a set of vectors along with two operations: vector addition and scalar multiplication, satisfying certain axioms. The vectors in V can be added together and multiplied by scalars (real or complex numbers), and the result will still be in V.

Subspace:
A subspace W of a vector space V is a subset of V that is itself a vector space under the same operations as V.

Basis and Dimension

Basis:
A basis of a vector space V is a set of vectors in V that are linearly independent and span V. This means every vector in V can be uniquely written as a linear combination of the basis vectors.

Example:
In $\mathbb{R}^3$, the standard basis is:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Any vector $\mathbf{v} \in \mathbb{R}^3$ can be written as $\mathbf{v} = a\mathbf{e}_1 + b\mathbf{e}_2 + c\mathbf{e}_3$.

Dimension:
The dimension of a vector space V is the number of vectors in a basis of V.

Example:
The dimension of $\mathbb{R}^3$ is 3 because it has 3 basis vectors.

Null Space, Column Space, and Row Space

Null Space (Kernel):
The null space of a matrix $\mathbf{A}$ is the set of all vectors $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$. It is a subspace of $\mathbb{R}^n$.

Example:
For $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, the null space consists of all vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ that satisfy $\mathbf{Ax} = \mathbf{0}$.

Column Space (Image):
The column space of a matrix $\mathbf{A}$ is the set of all linear combinations of the columns of $\mathbf{A}$. It is a subspace of $\mathbb{R}^m$.

Example:
For $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, the column space is spanned by the vectors $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$.

Row Space:
The row space of a matrix $\mathbf{A}$ is the set of all linear combinations of the rows of $\mathbf{A}$. It is a subspace of $\mathbb{R}^n$.

Example:
For $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, the row space is spanned by the vectors $\begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 3 & 4 \end{bmatrix}$.

Linear Transformations

Definition and Examples

Definition:
A linear transformation T from a vector space V to a vector space W is a function that maps vectors in V to vectors in W such that for all $\mathbf{u}, \mathbf{v} \in V$ and scalars c:

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (Additivity)
2. $T(c\mathbf{u}) = cT(\mathbf{u})$ (Homogeneity)

Example:
The function $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x \\ 3y \end{bmatrix}$ is a linear transformation.

Matrix Representation of Linear Transformations

Any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a matrix $\mathbf{A}$ such that $T(\mathbf{x}) = \mathbf{Ax}$.

Example:
For $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x + y \\ 3x - 2y \end{bmatrix}$, the matrix representation is:

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

Properties of Linear Transformations

1. **Kernel (Null Space)**: The set of vectors that are mapped to the zero vector.
2. **Image (Range)**: The set of vectors that can be written as $T(\mathbf{x})$ for some $\mathbf{x} \in V$.
3. **Rank**: The dimension of the image of T.
4. **Nullity**

: The dimension of the kernel of T.

Eigenvalues and Eigenvectors

Definition and Calculation

Definition:
For a square matrix $\mathbf{A}$, an eigenvector $\mathbf{v}$ and eigenvalue λ satisfy the equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.

Calculation:
To find eigenvalues, solve the characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

Example:
For $\mathbf{A} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$, solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$:

$$\begin{vmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{vmatrix} = (4-\lambda)(3-\lambda) - 2 = \lambda^2 - 7\lambda + 10 = 0$$

The eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = 2$.

To find eigenvectors, substitute each eigenvalue back into $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ and solve for $\mathbf{v}$.

Diagonalization of Matrices

A matrix $\mathbf{A}$ is diagonalizable if there exists a matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix.

Example:
For $\mathbf{A} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$ with eigenvalues $\lambda_1 = 5$ and $\lambda_2 = 2$ and corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, the matrix $\mathbf{P}$ is formed by eigenvectors:

$$\mathbf{P} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

Applications in Data Science (e.g., PCA)

Principal Component Analysis (PCA) is used to reduce the dimensionality of data while preserving as much variance as possible. Eigenvalues and eigenvectors of the covariance matrix of the data are used to identify principal components.

Singular Value Decomposition (SVD)

Definition and Computation

Definition:
SVD decomposes a matrix $\mathbf{A}$ into three matrices $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}^T$:

$$\mathbf{A} = \mathbf{U\Sigma V}^T$$

Where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\mathbf{\Sigma}$ is a diagonal matrix containing singular values.

Interpretation of SVD

- $\mathbf{U}$: Contains the left singular vectors (orthonormal columns).
- $\mathbf{\Sigma}$: Contains the singular values (diagonal elements) sorted in descending order.
- $\mathbf{V}$: Contains the right singular vectors (orthonormal columns).

Applications in Data Science (e.g., Noise Reduction, Data Compression)

Noise Reduction:
By truncating the smaller singular values, we can reduce noise in data. For example, in image processing, SVD can be used to remove noise from an image by reconstructing it using only the largest singular values.

Data Compression:
SVD is used in data compression techniques by approximating a matrix with lower rank. For example,

in recommender systems, the user-item matrix can be approximated with fewer singular values to reduce storage requirements.

Principal Component Analysis (PCA)

Concept and Purpose

PCA is a dimensionality reduction technique that transforms data into a new coordinate system where the greatest variances by any projection of the data come to lie on the first coordinates (called principal components), the second greatest variances on the second coordinates, and so on.

Step-by-Step Procedure

1. **Standardize the Data**: Subtract the mean of each feature and divide by the standard deviation.
2. **Compute the Covariance Matrix**: Calculate the covariance matrix of the standardized data.
3. **Compute the Eigenvalues and Eigenvectors**: Find the eigenvalues and eigenvectors of the covariance matrix.
4. **Sort Eigenvalues and Eigenvectors**: Sort the eigenvalues in descending order and arrange the corresponding eigenvectors.
5. **Select Principal Components**: Choose the top k eigenvectors as the principal components.
6. **Transform the Data**: Project the data onto the selected principal components to obtain the transformed dataset.

Interpretation of Principal Components

- The principal components are the directions in which the data varies the most.
- Each principal component corresponds to an eigenvector of the covariance matrix.
- The amount of variance captured by each principal component is proportional to its corresponding eigenvalue.

Applications in Dimensionality Reduction

- **Visualization**: Reducing high-dimensional data to 2 or 3 dimensions for visualization.
- **Noise Reduction**: By keeping only the principal components with the highest variance, we can filter out noise.
- **Feature Extraction**: Transforming data into a lower-dimensional space can help in extracting important features and improving the performance of machine learning models.