Notes on Linear Algebra by Suraj

Linear Algebra

Linear algebra is a branch of mathematics that deals with the study of vectors, matrices, and linear transformations.

Table of Contents

24. Tensors.
25. Random Matrix Theory.
26. Jordan Canonical Form.
27. Lyapunov Equations.
28. Matrix calculus.
29. Matrix Inequalities.

1. Vector spaces

A vector space is a collection of vectors that can be added together and scaled by scalars. The most common example of a vector space is the Euclidean space, which consists of vectors with n components, where n is a positive integer. In Euclidean space, vectors can be added component-wise and scaled by scalars.

A use case for vector spaces in data science is in machine learning, where datasets are often represented as vectors in a high-dimensional space. Vector spaces provide a framework for performing mathematical operations on these datasets, such as computing distances between vectors, projecting vectors onto subspaces, and performing linear transformations. Vector spaces are also used in deep learning, where neural networks can be thought of as functions that map vectors from one high-dimensional space to another.

2. Matrices

Matrices are a fundamental data structure in linear algebra, and are used to represent collections of numbers arranged in a grid or table. In Python, we can represent matrices using lists of lists or NumPy arrays.

Matrices are often used to represent data in machine learning and data science applications.

3. Linear equations

Linear equations are an important concept in linear algebra and are widely used in data science and machine learning. A linear equation is an equation that describes a linear relationship between two variables, typically represented as a straight line on a graph.

4. Eigen values and Eigen vectors

In linear algebra, an eigenvector of a square matrix is a non-zero vector that, when multiplied by the matrix, results in a scalar multiple of the original vector. This scalar multiple is called the eigenvalue of the matrix. In other words, if A is a square matrix, a non-zero vector v is an eigenvector of A if and only if $Av = \lambda v$, where λ is a scalar.

The eigenvectors and eigenvalues of a matrix have many important applications in various fields. One common application is in data analysis and dimensionality reduction. In particular, principal component analysis (PCA) is a technique that uses eigenvectors and eigenvalues to identify the most important features of a dataset and reduce its dimensionality.

For example, suppose you have a dataset with many features (columns) and you want to reduce it to a smaller set of features that capture most of the variation in the data. You can use PCA to compute the eigenvectors and eigenvalues of the covariance matrix of the data, and then select the top eigenvectors (those with the highest eigenvalues) as the new set of features. This can help simplify the data and improve the performance of machine learning algorithms that are applied to the data.

5. Determinants

Determinants are scalar values that can be calculated for square matrices. Determinants have many applications, such as testing whether a matrix is invertible, computing the volume of a parallelepiped in n-dimensional space, and solving systems of linear equations.

The determinant of the matrix A is zero. In linear algebra, a matrix is invertible if and only if its determinant is nonzero. Therefore, in this case, the matrix A is not invertible.

Given a system of equations in matrix form $Ax = b$, we can use the determinant of A to determine whether the system has a unique solution. If the determinant of A is nonzero, then the system has a unique solution. If the determinant is zero, then either the system has no solutions or infinitely many solutions.

6. Linear independence basis

Linear independence and basis are important concepts in linear algebra that are used to describe the structure of vector spaces.

Linear Independence:
A set of vectors in a vector space is said to be linearly independent if no vector in the set can be written as a linear

combination of the others. In other words, the only way to obtain the zero vector as a linear combination of the vectors in the set is by setting all the coefficients to zero.

Basis:
A basis for a vector space is a set of linearly independent vectors that span the entire space, meaning that any vector in the space can be written as a linear combination of the basis vectors.

7. Inner product spaces

An inner product space is a vector space with an additional structure that allows for the definition of a dot product. The dot product, or inner product, is a way of measuring the similarity between two vectors. In an inner product space, the dot product is defined in terms of an inner product function that satisfies certain properties.

One example of an inner product space is the space of real-valued n-dimensional vectors, denoted as R^n. In this space, the dot product of two vectors u and v is defined as:

u * v = u1v1 + u2v2 + ... + un*vn

where u1, u2, ..., un and v1, v2, ..., vn are the components of the vectors u and v, respectively.

8. Orthogonality

Orthogonality is a concept in linear algebra that refers to the perpendicularity of vectors or subspaces. Two vectors are said to be orthogonal if their dot product is zero. Orthogonality has many applications, such as projection of vectors onto subspaces etc.

In many applications, such as image processing and data analysis, the data is high-dimensional, which can make analysis and visualization difficult. One way to reduce the dimensionality of the data is to project it onto a lower-dimensional subspace. Orthogonality is useful in this context because it allows us to find a set of orthogonal vectors that span the subspace, which simplifies the computation and interpretation of the projection. For example, in principal component analysis (PCA), a widely used technique for dimensionality reduction, the goal is to find a set of orthogonal vectors that capture the maximum variance in the data.

9. Singular Value Decomposition(SVD)

Singular Value Decomposition (SVD) is a matrix factorization technique that factorizes a matrix into three matrices with specific properties. Given an $m \times n$ matrix A, SVD factorizes A into the product of three matrices: U, Σ, and V, where U and V are unitary matrices, and Σ is a diagonal matrix with non-negative real entries.

Use cases

Data compression - To compress a high dimensional data by representing it in a lower dimensional space while retaining as much information as possible.

Collaborative filtering - It can be used to build recommender systems by identifying latent features in a user-item interaction matrix.

PCA - It can be used to identify the most important features in a dataset and reduce its dimensionality.

Image processing - It can be used to perform image compression and denoising by representing images as a combination of basis images.

10. Vector norms

A vector norm is a mathematical function that measures the "size" or "magnitude" of a vector. There are many types of vector norms, but the most common ones are the L1 norm, the L2 norm, and the max norm.

The L1 norm of a vector x with n elements is defined as:

$||x||\_1 = |x\_1| + |x\_2| + ... + |x\_n|$

The L2 norm of a vector x with n elements is defined as:

$||x||\_2 = sqrt(x\_1^2 + x\_2^2 + ... + x\_n^2)$

The max norm of a vector x with n elements is defined as:

$||x||\_inf = max(|x\_1|, |x\_2|, ..., |x\_n|)$

Use cases

It is used in machine learning, especially L2 norm to regualarize the weights of a model which may prevent it from overfitting.

11 Orthogonal matrices

An orthogonal matrix is a square matrix where the columns (and rows) are orthonormal, meaning that each column has unit length and is orthogonal to every other column. In other words, an orthogonal matrix Q satisfies the equation $Q^T Q = I$, where Q^T is the transpose of Q and I is the identity matrix.

One use case for orthogonal matrices is in linear regression, where the least squares solution can be expressed as a matrix multiplication using the QR decomposition

12 Hermitian matrices.

A Hermitian matrix is a complex square matrix that is equal to its own conjugate transpose. In other words, if A is a Hermitian matrix, then $A = A^*$, where A^* represents the conjugate transpose of A.

13 Positive Definite Matrix

A positive definite matrix is a symmetric matrix in which all the eigenvalues are positive. In other words, for any vector x, the quadratic form $x^T Ax$ is positive, where A is a positive definite matrix and x^T denotes the transpose of x.

One use case for positive definite matrices is in optimization, where they are used to define convex functions.

Another use case for positive definite matrices is in Gaussian processes, which are a powerful tool in machine learning for modeling complex data distributions.

14 Sparse matrix

A sparse matrix is a matrix in which most of the elements are zero. Sparse matrices arise frequently in various applications, such as graph theory, finite element analysis, and image processing, where the matrices can be very large and have many zero entries. To store sparse matrices, we typically use specialized data structures that take advantage of their sparsity to reduce storage requirements and improve computational efficiency.

One use case for sparse matrices is in graph theory. Graphs can be represented as sparse matrices, where the rows and columns correspond to the vertices, and the non-zero entries correspond to the edges.

Another use case for sparse matrices is in image processing. Images can be represented as matrices, and in many cases, most of the entries in the matrix are zero (corresponding to black pixels). By using sparse matrix representations and algorithms, we can efficiently manipulate and analyze images, such as in image compression or denoising.

15. Linear Regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In the context of linear algebra, linear regression can be viewed as a problem of finding the best linear approximation of the dependent variable using the independent variables.

16. Principal component analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique used to transform a dataset into a new coordinate system that better captures the variance of the data. PCA is often used in data science and machine learning to reduce the number of features in a dataset, making it easier to visualize and analyze.

In linear algebra, PCA can be understood as finding the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors represent the principal components of the data, and the corresponding eigenvalues represent the amount of variance captured by each component.

17, SVT (Singular value thresholding)

Singular Value Thresholding (SVT) is a matrix completion technique that uses singular value decomposition (SVD) to recover low-rank matrices from incomplete or corrupted data. In the context of linear algebra, SVT involves computing the SVD of a matrix, setting the singular values below a certain threshold to zero, and reconstructing the matrix using the modified singular values.

18.Eigen decomposition

Eigen decomposition, also known as spectral decomposition, is a fundamental concept in linear algebra that factorizes a matrix into a set of eigenvectors and eigenvalues. An eigenvector of a matrix A is a non-zero vector x such that the multiplication of A with x results in a scalar multiple of x, i.e., $Ax = \lambda x$, where λ is a scalar called the eigenvalue associated with the eigenvector x.

19. QR-Factorization

QR factorization is a matrix factorization technique that decomposes a matrix into the product of an orthogonal matrix Q and an upper triangular matrix R. This decomposition can be useful in solving systems of linear equations, computing eigenvalues and eigenvectors, and other applications.

20. Non Negative matrix factorization.

NNMF stands for Non-negative Matrix Factorization, which is a matrix factorization technique used for reducing the dimensionality of data. It decomposes a matrix into two non-negative matrices, which can be used for feature extraction and data compression. NNMF is widely used in various applications, such as image processing, text mining, and recommender systems.

One use case of NNMF is in image processing, where it can be used for feature extraction and dimensionality reduction. For example, consider a dataset of images of handwritten digits, where each image is represented as a matrix of pixel values. By performing NNMF on this dataset, we can extract a set of non-negative features that represent the underlying patterns in the images, which can be used for classification or clustering. This can lead to more efficient and accurate machine learning algorithms, as well as faster computation times.

21. Convex optimization.

Convex optimization is a subfield of optimization that deals with finding the minimum of a convex function over a convex set. Convex optimization has many applications, such as machine learning, signal processing, and control theory.

Convex optimization involves finding the minimum of a convex function f(x) over a convex set C, where x is a vector in n-dimensional space. A function f(x) is convex if the line segment between any two points on the graph of f(x) lies above the graph. A set C is convex if the line segment between any two points in C lies entirely within C.

Convex optimization problems have the form:

minimize f(x)

subject to x in C

where f(x) is a convex function and C is a convex set. Convex optimization problems have many desirable properties, such as the existence of a unique global minimum and efficient algorithms for solving them.

As an example, consider the following convex optimization problem:
minimize x^2 + y^2

subject to x + y >= 1

This problem involves finding the minimum of the function f(x,y) = x^2 + y^2 subject to the constraint x + y >= 1. The constraint defines a convex set C, which is the half-plane above the line x + y = 1. The function f(x,y) is a convex function, since it is the sum of two convex functions (x^2 and y^2).Convex optimization.

Convex optimization is a subfield of optimization that deals with finding the minimum of a convex function over a convex set. Convex optimization has many applications, such as machine learning, signal processing, and control theory.

Convex optimization involves finding the minimum of a convex function f(x) over a convex set C, where x is a vector in n-dimensional space. A function f(x) is convex if the line segment between any two points on the graph of f(x) lies above the graph. A set C is convex if the line segment between any two points in C lies entirely within C.

Convex optimization problems have the form:

minimize f(x)

subject to x in C

where f(x) is a convex function and C is a convex set. Convex optimization problems have many desirable properties, such as the existence of a unique global minimum and efficient algorithms for solving them.

As an example, consider the following convex optimization problem:

minimize $x^2 + y^2$

subject to $x + y >= 1$

This problem involves finding the minimum of the function $f(x,y) = x^2 + y^2$ subject to the constraint $x + y >= 1$. The constraint defines a convex set C, which is the half-plane above the line x + y = 1. The function f(x,y) is a convex function, since it is the sum of two convex functions (x^2 and y^2).

22. Graph theory

Graph theory is a branch of mathematics that deals with the study of graphs, which are structures consisting of nodes (also called vertices) and edges that connect these nodes. Graphs can be used to represent many real-world situations, such as social networks, transportation systems, and computer networks.

Suppose you want to model a social network of users and their connections. You can represent this network using a graph, where each user is a node and each connection between users is an edge.

23. Differential Equations

Differential equations are mathematical equations that describe the relationships between a function and its derivatives.

We will use the example of a simple harmonic oscillator, which is a system that oscillates back and forth around a stable equilibrium point. The equation that describes the motion of a simple harmonic oscillator is a second-order differential equation:

d^2x/dt^2 + kx/m = 0

24 Tensors

Tensors are mathematical objects that generalize scalars, vectors, and matrices to higher dimensions. In Python, tensors can be represented using PyTorch, NumPy arrays, which provide efficient array operations and numerical computations.

25 Random Matrix Theory

Random Matrix Theory (RMT) is a branch of mathematics that studies the statistical properties of matrices with random entries. In RMT, the entries of a matrix are assumed to be random variables with certain statistical properties, such as Gaussian distributions.

RMT has applications in many fields, such as quantum mechanics, wireless communications, and machine learning.

One common use case of RMT is in analyzing the performance of random matrix-based algorithms in machine learning, such as principal component analysis (PCA) and random projection. RMT provides a theoretical framework for understanding the behavior of these algorithms and predicting their performance under different scenarios.

26. Jordan Canonical Form

The Jordan canonical form is a way to decompose a matrix into simpler parts that can be easier to work with. Specifically, it transforms a matrix into a block diagonal form where each block corresponds to a Jordan block, which is a matrix that has a specific structure.

The Jordan blocks are constructed by taking the eigenvalues of the matrix and constructing a block for each eigenvalue. Each block has the eigenvalue along its diagonal, and 1's above the diagonal. If there are multiple blocks corresponding to the same eigenvalue, they are arranged in a specific way that depends on the dimension of the blocks and the number of blocks.

27. Lyapunov Equations

Lyapunov equations are a type of matrix equation that arise in control theory and signal processing. In particular, they are used to analyze the stability of linear dynamical systems. Given a matrix A, the Lyapunov equation is of the form:

$$AX + XA^T = -Q$$

where X is a matrix to be solved for, and Q is a given symmetric positive definite matrix.

The solution to the Lyapunov equation can be used to determine the stability of the linear dynamical system described by A. If the solution X is positive definite, then the system is asymptotically stable.

28. Matrix calculus

It is a branch of calculus that deals with derivatives and integrals of matrices. It is used in many applications, such as machine learning, control theory, and optimization.

29. Matrix Inequalities

Matrix inequalities are mathematical expressions that relate matrices and their eigenvalues.

The use case of matrix inequalities such as the Rayleigh-Ritz theorem is primarily in linear algebra and numerical analysis. These inequalities provide useful bounds on eigenvalues and other matrix properties that can be used to analyze the behavior of linear systems, such as in control theory and signal processing.