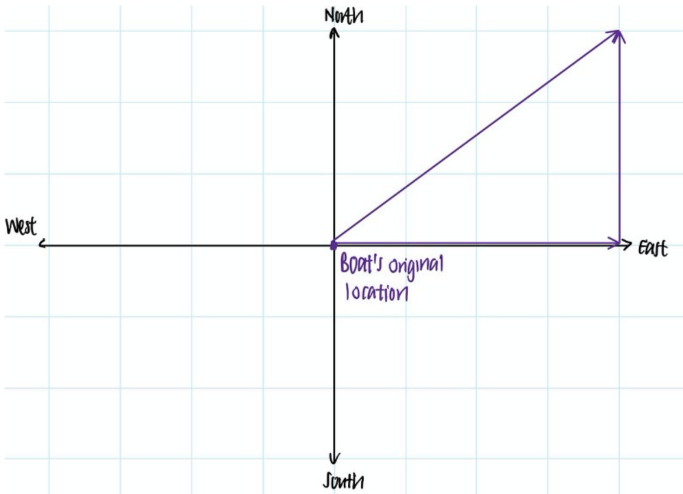
Linear Algebra for Data Science

Linear Algebra is a powerful mathematical tool that finds wide application in the field of data science and machine learning. It is a fundamental skill for machine learning practitioners, as many machine learning models can be represented and manipulated using matrices. In fact, datasets are often represented as matrices. Linear algebra plays a crucial role in tasks such as data preprocessing, data transformation, and model evaluation in machine learning workflows.

1. Basic Concepts in Linear Algebra in Data Science

- ### Vectors

A vector represents a quantity with direction. For example, a boat moving north-east at 5 miles per hour can be represented as a 5-unit vector in the north-west direction. Vectors exist in n-dimensional space, such as a 2-dimensional space for a boat's north-south and east-west movement. If the boat moved 3 miles north and 4 miles east in the past hour, its velocity vector can be represented as x = [3, 4]. Understanding vector properties, such as linear dependence or independence, and operations like dot and cross products, as well as the properties of vector spaces, is essential in learning linear algebra.



- ### Matrices

When we combine multiple vectors, we obtain a matrix, which can represent transformations such as scaling, rotation, or translation on vectors. For instance, to make a boat move twice as fast in the same direction (represented by the new vector y), we can scale the boat's velocity vector by two using the matrix A, according to the formula: y = A x

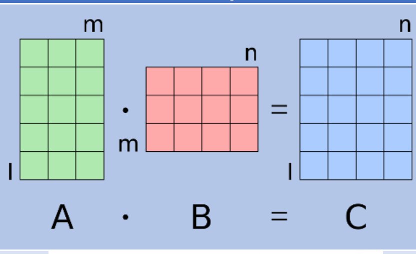
$$\begin{array}{rcl} & \text{3 columns} & \\ & \downarrow \;\; \downarrow \;\; \downarrow & \\ A = & \begin{bmatrix} -2 & 5 & 6 \\ 5 & 2 & 7 \end{bmatrix} & \leftarrow \text{2 rows} \end{array}$$

- ### Eigenvectors and Eigenvalues

Every matrix possesses unique properties that play a vital role in its behavior. Among these properties, the eigenvector stands out as a key element. It refers to a vector that remains unchanged in direction even after the matrix undergoes a transformation. Another critical property is the eigenvalue, which denotes the change in magnitude of the same vector following the transformation.

The eigenvector holds immense significance in advanced applications of the field, serving as a fundamental concept in various techniques such as Principal Component Analysis and Singular Value Decomposition. These methods leverage the power of eigenvectors to gain valuable insights and make meaningful interpretations in diverse domains.

2. Different Linear Algebra Operations Used in Data Science

Operation	Description	Applications	Example
Matrix Multiplication	Multiplying two matrices to produce a third matrix	Data transformation, solving systems of linear equations, linear regression, deep learning	
Determinant	Scalar value computed from a square matrix	Finding the inverse of a matrix, computing condition number, solving systems of linear equations	$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\lvert A \rvert = ad - bc$
Inverse	Another matrix that, when multiplied with the original matrix, yields the identity matrix	Solving systems of linear equations, data transformation, regularization in machine learning	$A^{-1} = \frac{1}{\lvert A \rvert}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
Transpose	Swapping rows and columns of a matrix to produce a new matrix	Data manipulation, data transformation, computing covariance matrices	$\begin{bmatrix} 2 & 4 & -1 \\ -10 & 5 & 11 \\ 18 & -7 & 6 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & -10 & 18 \\ 4 & 5 & -7 \\ -1 & 11 & 6 \end{bmatrix}$
Rank	Number of linearly independent rows (or columns) in a matrix	Determining linear independence of vectors, identifying dimensionality of datasets, solving systems of linear equations	$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ Rank (A) = 4 Nullity (A) = 1
Eigenvalues and Eigenvectors	Values and vectors that remain unchanged (up to scaling) when a matrix is applied as a linear transformation	Dimensionality reduction (e.g. PCA), feature extraction, data compression, image processing	$\underbrace{\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}}_{\text{Original Matrix}} = \underbrace{\begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix}}_{\text{Eigenvectors Matrix}} \underbrace{\begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}}_{\text{Eigenvalues Matrix}} \underbrace{\begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}}_{\text{Inverse of Eigenvectors Matrix}}$

3. Different Methods for Solving Linear Systems of Equations

In mathematics, a system of linear equations (or linear system) is a collection of one or more linear equations involving the same variables. For example, $\begin{cases} 3x + 2y - z = 1 \\ 2x - 2y + 4z = -2 \\ -x + \frac{1}{2}y - z = 0 \end{cases}$

is a system of three equations in the three variables x, y, z. A solution to a linear system is an assignment of values to the variables such that all the equations are simultaneously satisfied. A solution to the system above is given by the ordered triple $(x, y, z) = (1, -2, -2),$

since it makes all three equations valid. The word "system" indicates that the equations are to be considered collectively, rather than individually.

Some methods for solving Linear Systems of Equations:

Method	Description	Benefits
Gaussian Elimination	Performs row operations on the coefficient matrix to transform it into row-echelon or reduced row-echelon form, and then uses back-substitution to obtain the solution.	• Simple to implement. • Applicable to square and rectangular matrices • Provides exact solutions
LU Decomposition	Factors the coefficient matrix into lower and upper triangular matrices using Gaussian elimination, and then uses forward and backward substitution to obtain the solution.	• Efficient for solving multiple linear systems with the same coefficient matrix but different right-hand side vectors. • Avoids redundant computations of LU factorization for each system
Matrix Inversion	Computes the inverse of the coefficient matrix and then multiplies it with the right-hand side vector to obtain the solution.	• Provides exact solutions. • Applicable to small matrices with non-singular coefficient matrices • Useful for theoretical analysis of linear systems
Iterative Methods	Approximates the solution iteratively, often used for large sparse systems of equations.	• Suitable for large sparse systems of equations • Can handle systems with millions of unknowns

4. Dimensionality Reduction

In machine learning, classification problems can involve numerous features that are used to make the final classification decision. These features are essentially variables that can make visualizing and working with the training set challenging, especially when many of them are correlated and redundant. This is where dimensionality reduction algorithms come into play, as they help reduce the number of variables by identifying a set of principal variables. Dimensionality reduction can be achieved through feature selection or feature extraction techniques.

Let's consider an example of dimensionality reduction using e-mail classification. In this problem, we need to determine if an e-mail is spam or not, based on various features like the title, content, and template usage. However, some of these features may overlap or be highly correlated, such as humidity and rainfall in a weather classification problem. By reducing the number of features, we can simplify the problem and make it easier to visualize. For instance, a 3-D problem can be mapped to a 2-D space, and a 2-D problem to a simple line. The diagram below illustrates this concept, where a 3-D feature space is split into two 2-D spaces, and further reduction is possible if correlated features are identified.