**Mathematics for Machine Learning**

# Data

* Individual pieces of factual information collected from various sources which is stored, processed for data analysis.

## Types of Data

* + Nominal Data- label variables without any measurable values. Eg: country, gender, race etc.
  + Ordinal Data - categorical data with a set order or scale to it. Eg: movie ratings, salary range etc.
  + Discrete - data with finite set of values which can be categorized. Eg:class strength, questions answered correctly etc
  + Continuous - data can take numerical values within a range. Eg: Water pressure, weight of a person.

A diagram of a data flow

Description automatically generated

# LINEAR ALGEBRA

A domain of mathematics concerning linear equations and their representations in vector spaces and through matrices.

A diagram of a triangle

Description automatically generated

Linear equation is an equation having a maximum order of one.A screenshot of a math problem

Description automatically generated

# Vectors

One dimensional matrix is considered as a vector.

A blackboard with different colored arrows

Description automatically generated with medium confidence

The span of a set of vectors is the set of all possible linear combinations of those vectors. In other words, if you have a set of vectors {v1,v2,…,vn}{, the span of these vectors, denoted as span(v1,v2,…,vn) is the set of all vectors that can be expressed as:

*c*1​**v**1 ​+ *c*2​**v**2 ​+ …+ *cn*​**v***n*​

where c1,c2,…,cn ​ are scalars.

To be a subspace, three conditions must be satisfied:

1. **Contains the zero vector**: The span of any set of vectors includes the zero vector because you can choose all the coefficients to be zero, resulting in the zero vector.
2. **Closed under addition**: If you take two vectors from the span and add them together, the result will still be in the span. This is because the sum of two linear combinations is itself a linear combination.
3. **Closed under scalar multiplication**: If you take any vector from the span and multiply it by a scalar, the result will still be in the span. This is because scaling each term of a linear combination still results in a linear combination.

These properties ensure that the span of a set of vectors forms a valid subspace of the vector space. Consequently, the span represents a subspace that can be described as the "smallest" subspace containing the original vectors. It's the subspace that is "generated" or "spanned" by those vectors.

**Linear Independence**: In linear algebra, a set of vectors is said to be linearly independent if no vector in the set can be represented as a linear combination of the others. Formally, a set of vectors {**v**1​,**v**2​,…,**v***n*​} is linearly independent if the only solution to the equation:

c1v1+c2v2+…+cnvn=0

is the trivial solution where all coefficients c1,c2,…,cn ​ are zero.

In simpler terms, no vector in a linearly independent set can be expressed as a combination of the others. Each vector in the set contributes in a unique way to the span of the set.

**Basis**: A basis for a vector space is a set of vectors that is both linearly independent and spans the entire space. In other words, a basis forms the "building blocks" of the space, and any vector in the space can be expressed as a unique linear combination of the basis vectors.

Formally, let *V* be a vector space. A set of vectors *B*={**v**1​,**v**2​,…,**v***n*​} is a basis for *V* if:

1. B spans V, i.e., every vector in V*V* can be expressed as a linear combination of the vectors in *B*.
2. B is linearly independent, i.e., no vector in *B* can be expressed as a linear combination of the others.

The number of vectors in a basis for a vector space is called the dimension of the vector space. If a vector space has dimension *n*, then any basis for that space will consist of *n* linearly independent vectors.

ANALOGY: **Linear Independence**: Imagine you have a group of friends, and each friend has their own favorite toy. They don't share their toys with each other. Now, let's say we look at all these toys together. If no toy can be made by combining other toys, we call these toys "linearly independent." It's like each toy has its own special superpower and can't be made by mixing others. If this happens, then we have a group of linearly independent toys!

**Basis**: Now, let's say we have a big box where we want to keep all our toys. We want to find some special toys to put in the box so that if we have any other toy, we can make it using a combination of the special ones. These special toys are like the leaders of our toy group. We call them the "basis."

Eigenvalues are scalars (usually represented by the Greek letter λ, lambda) associated with a square matrix. For a given matrix *A*, an eigenvalue *λ* is a scalar such that when multiplied by a corresponding eigenvector, the result is a scaled version of that eigenvector. Mathematically, it can be represented as:

*A***v**=*λ***v**

Where:

* *A* is the square matrix.
* **v** is the eigenvector associated with the eigenvalue λ.

Eigenvectors are non-zero vectors that, when multiplied by a square matrix, result in a scaled version of themselves. In other words, they point in a direction that is unchanged by the linear transformation represented by the matrix. Each eigenvector corresponds to an eigenvalue. Mathematically, for a given matrix *A* and eigenvalue *λ*, the equation can be written as:

*A***v**=*λ***v**

Where:

* *A* is the square matrix.
* **v** is the eigenvector.
* *λ* is the corresponding eigenvalue.

T(v)= Av

W is not eigen vector

V is eigen vector.

A is eigen value



In summary, the general proof of eigenvalues and eigenvectors involves:

1. Expressing the condition *A***v**=*λ***v**.
2. Rewriting it as (A−λI)v=0
3. Setting the determinant of *A*−*λI* to zero to find eigenvalues.

(*A*−*λI*)**v**=0

Where *I* is the identity matrix of size *n*×*n*.

For **v** to be a non-zero vector, the matrix *A*−*λI* must be singular (i.e., its determinant is zero), which means it has a non-trivial null space.

Thus, we have:

det(A−λI)=0

This equation is called the characteristic equation of matrix *A*. Solving this equation for *λ* will give us the eigenvalues of matrix *A*. Once we have found the eigenvalues *λi*​, we can substitute them back into the equation (A−λiI)v=0 and solve for the corresponding eigenvectors **v***i*​.

A =



1 2

3 4

Neural network is just a giant differential equation. A screenshot of a computer

Description automatically generated

Where w is weight vector and x is features and y is labels/classes.

A screenshot of a computer

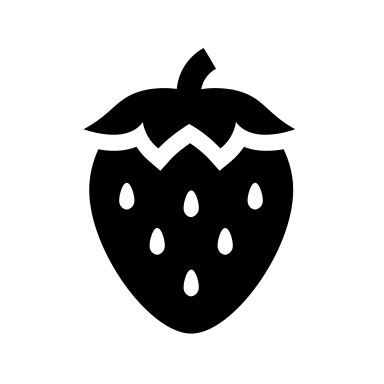
Description automatically generated

A screenshot of a computer

Description automatically generated

Non singular system

Singular System

S

(0,10)

Slope = -1;

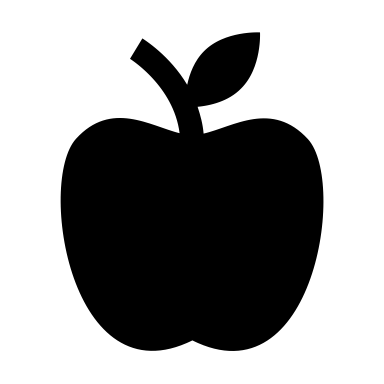
y-intercept=10

(4,6)

A+S=10

(8,2)

(10,0)

A

Linear equations can be represented in plane. System of linear equations can be represented as arrangement of lines in planeA graph of equations and equations

Description automatically generated



Constants in the system are not important to determine if an equation is singular or non-singular.

1)a+b=0

2)2a+2b=0

Singular matrix – infinitely many solutions

|  |  |
| --- | --- |
| 1 | 1 |
| 1 | 2 |

1)a+b=0

2)a+2b=0

|  |  |
| --- | --- |
| 1 | 1 |
| 1 | 2 |

Non-Singular matrix

System 3- singular

A+b+c = 10

A+b+2c=15

A+b+3c=18

no solutions-contradictory

System 2- singular

A+b+c = 10

A+b+3c=20

A+b+2c=15

Infinite solutions-redundant

System 4- singular

A+b+c = 10

2a+2b+2c=15

3a+3b+3c=12

Infinite solutions- redundant

System 1 – non singular

A+b+c = 10

A+2b+c=15

A+b+2c=12

Unique solution-complete

System 1 – non singular

A+b+c = 0

A+2b+c=0

A+b+2c=0

Unique solution-complete- a=b=c=0

System 3- singular

A+b+c = 0

A+b+2c=0

A+b+3c=0

no solutions-contradictory

System 2- singular

A+b+c = 0

A+b+3c=0

A+b+2c=0

Infinite solutions-redundant

System 4- singular

A+b+c = 0

2a+2b+2c=0

3a+3b+3c=0

Infinite solutions- redundant

System 2- By subtracting the second equation from the third equation, we get C=0 .Then substituting C=0 into any of the equations, we get A+B=0.

This implies that A=− *B*. So, the system can be rewritten as:

{A+B+0=0

A+B+0=0

A+B+0=0}

Since the equations are identical, they are essentially representing the same line in three-dimensional space. Hence, the system has infinitely many solutions because any value of *A* and *B* that satisfy the equation A=−*B* will satisfy all three equations.

This makes the system redundant because one equation can express all the information contained in the other two equations.

* a system of equation is singular if the second equation carries the same information as the first one. This is the concept of linear dependence.
* The reason this system is singular is because the second equation is a multiple of the first one. It’s two times the first one. That means if we take the left hand on the right hand and multiply everything by two in the first equation, you get the second equation.
* if you look at the corresponding matrix, then the second row is a multiple of the first one. By that, we mean that if you take every element in the first row and multiply them all by two, you get the second row. That means that the second row can be obtained from the first one, so the second row is dependent on the first one.
* the first row is dependent on the second one by taking the second row and multiplying everything by a half. In any way, either one depends on the other one or the other one depends on the one, they're both dependent, so they're linearly dependent.
* The second equation is not a multiple of the first one, or vice versa. There's no constant that I can multiply the first equation to get the second equation, or vice versa. This is why the system is non-singular. I cannot take a number and multiply one row entirely by the number to get the other row. That means the rows are linearly independent.

A screenshot of a math equation

Description automatically generated with medium confidence A diagram of a mathematical equation

Description automatically generated

A screenshot of a graph

Description automatically generated

The determinant – a quick formula to decide if a matrix is singular or non singlular. A screenshot of a math problem

Description automatically generated A screenshot of a math test

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Now, since everything below the diagonal is zero, all the terms in the determinant are going to contain one of these elements below the diagonal, except for the one that takes the entire main diagonal. Whenever you have a matrix where anything underneath the diagonal is zero, the determinant is going to be the product of the elements in the main diagonal.

Manipulating Equations

One way to manipulate equations is to multiply them by a constant.

Another way to manipulate equations is to add two equations.