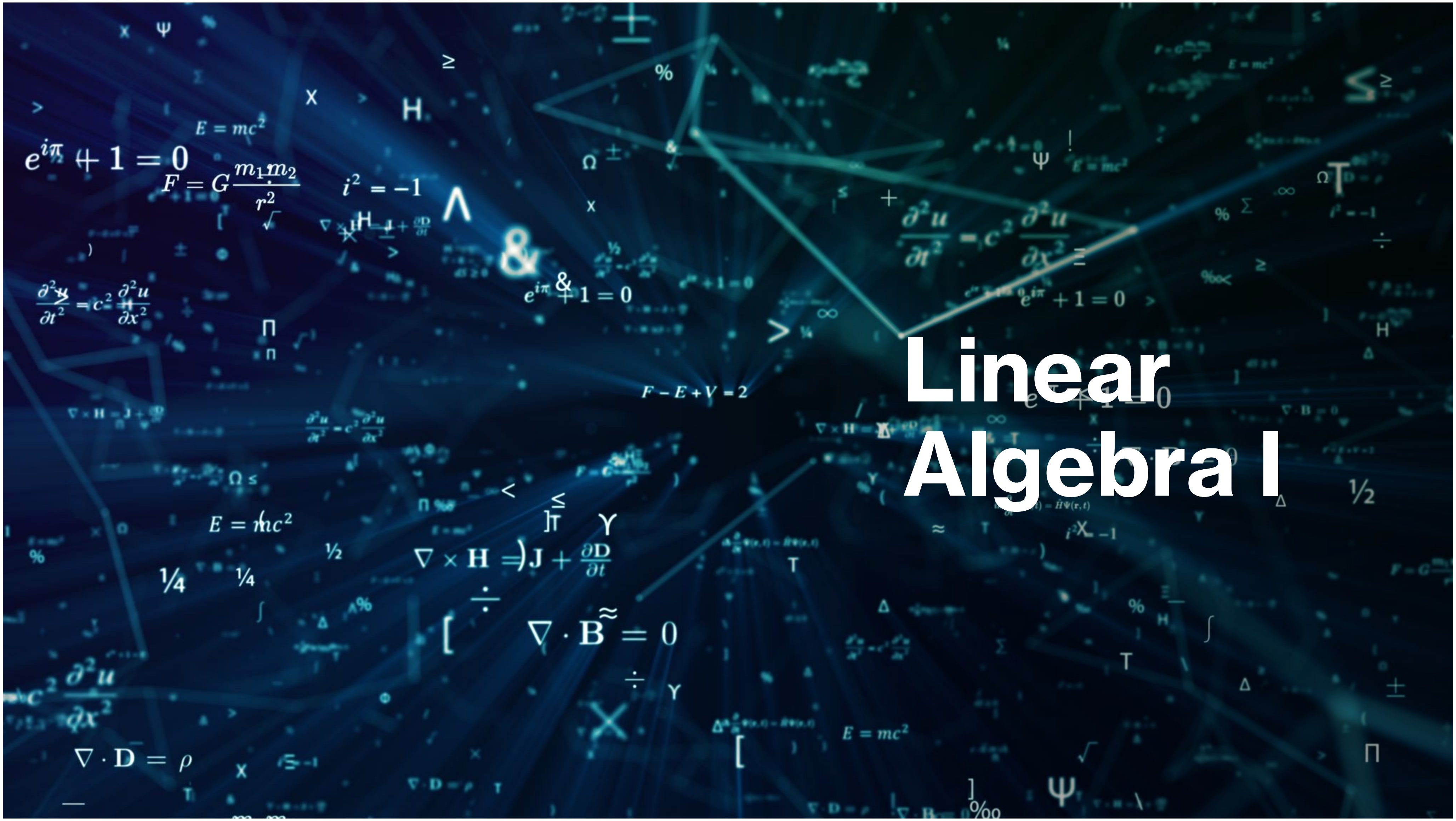


Linear Algebra I

Data Representation

- For example, We have the following data

Age	**GPA**	**Hours Studied**
20	3.4	15
21	3.6	10
19	3.2	18

- How do we represent the data?
- Every **row** is a vector representing one student.
  Every **column** is a vector representing a feature.

Data Representation

- Suppose we have the grades obtain for each student as well.

Age	GPA	Hours Studied	Grades
20	3.4	15	80
21	3.6	10	50
19	3.2	18	90

- How can we train from this data to predict the grades of any students whose features are known?

Linear Algebra

- Linear algebra is the study of vectors and certain rules associated with vectors.

Vectors

- Linear algebra is the study of vectors and certain rules associated with vectors.
- The vectors many of us know from school are called “geometric vectors”, which are usually denoted by a small arrow above the letter, e.g., $\vec{x}$ and $\vec{y}$.
- We will discuss more general concepts of vectors and use a bold letter to represent them, e.g., $\boldsymbol{x}$ and $\boldsymbol{y}$.
- One major idea in mathematics is the idea of “closure”.
- This is the question: Does the resultant of the addition of two vectors belong from the same set or outside the set? Does the resultant of the scalar multiplication of a vectors with a scalar belong from the same set or outside the set?

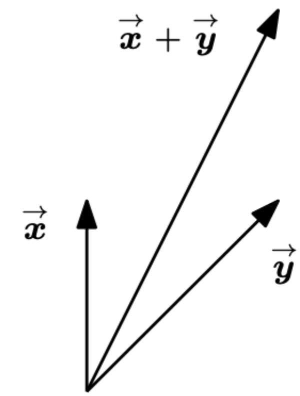
Vectors

- In general, vectors are special objects that can be added together and multiplied by scalars to produce another object of the same kind.

Vectors

Examples:

1. Geometric vectors. This example of a vector may be familiar from high school mathematics and physics. Two geometric vectors can be added, such that $\vec{x} + \vec{y} = \vec{z}$ is another geometric vector. Furthermore, multiplication by a scalar $\lambda\vec{x}$, $\lambda \in \mathbb{R}$, is also a geometric vector.

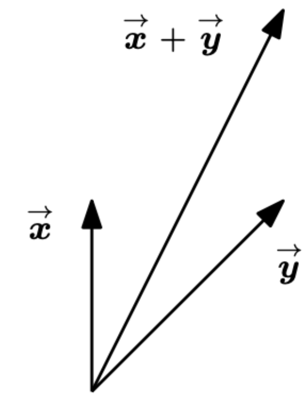


2. Polynomials are also vectors; see Figure 2.1(b): Two polynomials can be added together, which results in another polynomial; and they can be multiplied by a scalar $\lambda \in R$, and the result is a polynomial as well.

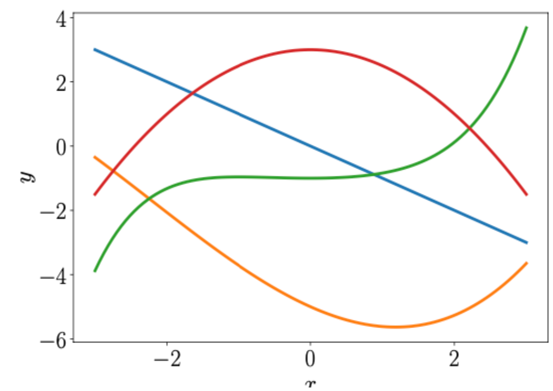
Vectors

Examples:

1. Geometric vectors. This example of a vector may be familiar from high school mathematics and physics. Two geometric vectors can be added, such that $\vec{x} + \vec{y} = \vec{z}$ is another geometric vector. Furthermore, multiplication by a scalar $\lambda\vec{x}$, $\lambda \in \mathbb{R}$, is also a geometric vector.



2. Polynomials are also vectors; see Figure 2.1(b): Two polynomials can be added together, which results in another polynomial; and they can be multiplied by a scalar λ ∈ R, and the result is a polynomial as well.



Vectors

Examples:

3. Audio signals are vectors. Audio signals are represented as a series of numbers. We can add audio signals together, and their sum is a new audio signal. If we scale an audio signal, we also obtain an audio signal. Therefore, audio signals are a type of vector, too.
4. Elements of $\mathbb{R}^n$ (tuples of n real numbers) are vectors. $\mathbb{R}^n$ is more abstract than polynomials. For instance,

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^n$$

is an example of a triplet of numbers. Adding two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ component-wise results in another vector: $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{c} \in \mathbb{R}^n$. Moreover, multiplying $\boldsymbol{a} \in \mathbb{R}^n$ by $\lambda \in \mathbb{R}^n$ results in a scaled vector $\lambda \boldsymbol{a} \in \mathbb{R}^n$.

Generalization to $\mathbb{R}^n$

- Column vectors with n entries, i.e.

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \text{ are vectors in } \mathbb{R}^n$$

- All the vectors with n real valued entries are form $\mathbb{R}^n$
- Generally, we cannot have a geometric description for $\mathbb{R}^n$

Linear Operations in $\mathbb{R}^n$

Sum and scalar multiple?

Algebraic Properties in $\mathbb{R}^n$

For all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ and all scalars c and d:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$,
where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

Distance and Angles

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors given as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Dot Product

The dot product between $\boldsymbol{a}$ and $\boldsymbol{b}$ gives a scalar value and is defined as

$$\boldsymbol{a}.\boldsymbol{b} = \boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{b} = \begin{pmatrix} a_1 & a_2 & \dots & a_m \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_m b_m$$

Distance and Angles

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors given as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Length

The Euclidean norm or length of a vector $\boldsymbol{a}$ is defined as

$$\|\boldsymbol{a}\| = \sqrt{\boldsymbol{a}^T\boldsymbol{a}} = \sqrt{a_1^2 + a_2^2 + \cdots + a_m^2}$$

The Euclidean norm is a special case of a general class of norms, known as L_p-norm, defined as

$$\|\boldsymbol{a}\|_p = (|a_1|^p + |a_2|^p + \cdots + |a_m|^p)^{1/p}$$

Distance and Angles

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors given as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Distance

From Euclidean norm we can define the Euclidean distance between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ defined as

$$\|\boldsymbol{a} - \boldsymbol{b}\| = \sqrt{(\boldsymbol{a} - \boldsymbol{b})^{\boldsymbol{T}}(\boldsymbol{a} - \boldsymbol{b})} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_m - b_m)^2}$$

Similarly L_p-norm distance is defined as

$$\|\boldsymbol{a} - \boldsymbol{b}\|_p = (|a_1 - b_1|^p + |a_2 - b_2|^p + \cdots + |a_m - b_m|^p)^{1/p}$$

Distance and Angles

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors given as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Unit Vector

The unit vector in the direction of $\boldsymbol{a}$ is given as

$$\boldsymbol{u} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} = \left(\frac{1}{\|\boldsymbol{a}\|}\right)\boldsymbol{a}$$

This in turn makes $\|\boldsymbol{u}\| = 1$, and it is also called **normalized vector.**

Distance and Angles

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors given as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Angle

The cosine of the smallest angle between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is given as

$$\cos\theta = \frac{\boldsymbol{a}.\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} = \left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}\right)^{\boldsymbol{T}}\left(\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\right)$$

Distance and Angles

Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors given as

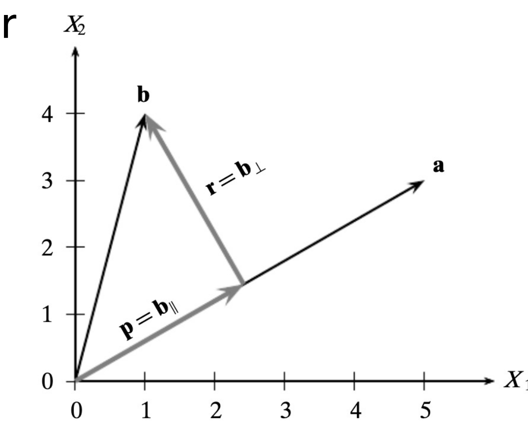
$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Orthogonality

Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are said to be orthogonal if and only if $\boldsymbol{a}.\boldsymbol{b} = \boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{b} = 0$, which in turn implies that $\cos\theta = 0$ and $90°$. In this case, we say that the vectors have no similarity.

Orthogonal Projection

Often in data mining we need to project a point or vector onto another vector, for example, to obtain a new point after a change of the basis vectors.



Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ be two m-dimensional vectors. An **orthogonal decomposition** of the vector $\boldsymbol{b}$ in the direction of another vector $\boldsymbol{a}$ is given as

$$\boldsymbol{b} = \boldsymbol{b}_\parallel + \boldsymbol{b}_\perp = \boldsymbol{p} + \boldsymbol{r}$$

where $\boldsymbol{p} = \boldsymbol{b}_\parallel$ is parallel to $\boldsymbol{a}$, and $\boldsymbol{r} = \boldsymbol{b}_\perp$ is perpendicular or orthogonal to $\boldsymbol{a}$.

The vector $\boldsymbol{p}$ is called the **orthogonal projection** or simply **projection of $\boldsymbol{b}$ on the vector $\boldsymbol{a}$,** denoted by $proj_{\boldsymbol{a}}(\boldsymbol{b})$.

Orthogonal Projection

- Note that the point $\boldsymbol{p} \in \mathbb{R}^{\boldsymbol{m}}$ is the point closest to $\boldsymbol{b}$ on the line passing through $\boldsymbol{a}$.
- Thus, the magnitude of the vector $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{p}$ gives the **perpendicular distance** between $\boldsymbol{b}$ and $\boldsymbol{a}$, which is often interpreted as the residual or error between the points $\boldsymbol{b}$ and $\boldsymbol{p}$.
- The vector $\boldsymbol{r}$ is also called the **error vector.**

Orthogonal Projection

We can derive an expression for $\boldsymbol{p} = proj_{\boldsymbol{a}}(\boldsymbol{b})$ by noting that $\boldsymbol{p} = c\boldsymbol{a}$ for some scalar c, as $\boldsymbol{p}$ is parallel to $\boldsymbol{a}$. Thus,

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - c\boldsymbol{a}$$

Because $\boldsymbol{p}$ and $\boldsymbol{r}$ are orthogonal, we have

$$\boldsymbol{p}.\boldsymbol{r} = \boldsymbol{p}^{\boldsymbol{T}}\boldsymbol{r} = 0$$

$$(c\boldsymbol{a})^{T}(\boldsymbol{b} - \boldsymbol{c}\boldsymbol{a}) = c\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{b} - c^{2}\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{a} = \boldsymbol{0}$$

$$c = \frac{\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{b}}{\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{a}} = \frac{\boldsymbol{a}.\boldsymbol{b}}{\boldsymbol{a}.\boldsymbol{a}} = \frac{\boldsymbol{a}.\boldsymbol{b}}{\|\boldsymbol{a}\|^{2}}$$

Therefore,

$$\boldsymbol{p} = proj_{\boldsymbol{a}}(\boldsymbol{b}) = \left(\frac{\boldsymbol{a}.\boldsymbol{b}}{\|\boldsymbol{a}\|^{2}}\right)\boldsymbol{a}$$

Matrices

An array of numbers is called a matrix.

A matrix with n number of rows and m number of columns is known as an $n \times m$ matrix.

Matrices

Example:

Data can often be represented or abstracted as an $n \times d$ data matrix, with n rows and d columns, where rows correspond to entities in the dataset, and columns represent attributes or properties of interest.

$$
\mathrm{D} = \begin{array}{c} \\ \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_n} \end{array}
\begin{array}{c} \begin{array}{cccc} X_1 & X_2 & \cdots & X_d \end{array} \\
\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \end{array}
$$

Matrices

Example:.

$$
\begin{array}{c c}
 & \begin{array}{cccc} X_1 & X_2 & \dots & X_d \end{array} \\
\mathrm{D} = \begin{array}{c} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_n} \end{array} & \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{pmatrix}
\end{array}
$$

where $\boldsymbol{x_i}$ denotes the ith row, which is a d-tuple given as

$$\boldsymbol{x_i} = (x_{i1}, x_{i2}, \dots, x_{id})$$

And $\boldsymbol{X_j}$ denotes the *j*th column, which is an n-tuple given as

$$\boldsymbol{X_j} = (x_{1j}, x_{2j}, \dots, x_{nj})$$

Matrices

- Depending on the application domain, rows may also be referred to as **entities, instances, examples, records, transactions, objects, points, feature-vectors, tuples**, and so on.
- Likewise, columns may also be called **attributes, properties, features, dimensions, variables, fields,** and so on.
- The number of instances n is referred to as the **size** of the data, whereas the number of attributes d is called the **dimensionality** of the data.
- The analysis of a single attribute is referred to as **univariate analysis**.
- The simultaneous analysis of two attributes is called **bivariate analysis.**
- The simultaneous analysis of more than two attributes is called **multivariate analysis.**

Matrices

The following are the operations of matrices

1. **Matrix addition** $M + N$ (only possible of both matrices have same size)
2. **Scalar multiplication** λM ($\lambda \in \mathbb{R}$)
3. **Matrix multiplication** MN (only possible if the number of column of the first matrix is equal to the number of rows of the second matrix).

If M is a $m \times p$ matrix and N is a $p \times n$ matrix then MN will be a $m \times n$ matrix

Example:

The multiplication of matrices

$$M = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 4 \\ 6 & 7 & 9 \end{bmatrix}, N = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Is

$$MN = \begin{bmatrix} 3 & 3 \\ 7 & 7 \\ 22 & 22 \end{bmatrix}$$

Example:

For a linear system of linear equations

$$2x + 3y = 2$$
$$2x + 2y = 1$$

The matrix equation is written as $A\boldsymbol{x} = \boldsymbol{b}$

$$\begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Why?

Because when we multiply

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x + 3y \\ 2x + 2y \end{bmatrix}$$

Which can also be written as

$$x \begin{bmatrix} 2 \\ 2 \end{bmatrix} + y \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Which is a **linear combination** of the vectors $\boldsymbol{v_1} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\boldsymbol{v_2} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. Here the vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ are called **column vectors** of the matrix A.

We will come back to discuss the solution of system linear equation.

Linear Combinations of Vectors

- Given a set of vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots., \boldsymbol{v_k}$, where k is an integer,
- then the vector $\boldsymbol{y}$ defined as

$$\boldsymbol{y} = x_1\boldsymbol{v_1} + x_2\boldsymbol{v_2} + \cdots + x_k\boldsymbol{v_k}$$

  is a **linear combination** of vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots., \boldsymbol{v_k}$, where $x_{1,} x_2 \ \ldots., x_k \in \mathbb{R}$.
- Or equivalently, the vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots., \boldsymbol{v_k}$ said to **generate** the vector $\boldsymbol{y}$.

Example

$$2\begin{bmatrix}-1\\1\end{bmatrix} - 0.5\begin{bmatrix}2\\4\end{bmatrix} + 2\begin{bmatrix}0.5\\-0.5\end{bmatrix} = \begin{bmatrix}-2\\-1\end{bmatrix}$$

Example

Let $\boldsymbol{a_1} = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \boldsymbol{a_2} = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$

- Determine whether $\boldsymbol{b}$ can be generated (or written as a linear combination) of $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$.
- That is, determine whether weights x_1 and x_2 exist such that
$$x_1\boldsymbol{a_1} + x_2\boldsymbol{a_2} = \boldsymbol{b}.$$
- The above equation is known as a vector equation.
- Find if this vector equation has a solution.

Solution

Use the definitions of scalar multiplication and vector addition to rewrite the vector equation

$$x_1\boldsymbol{a_1} + x_2\boldsymbol{a_2} = \boldsymbol{b}$$

$$x_1\begin{bmatrix}1\\-2\\-5\end{bmatrix} + x_2\begin{bmatrix}2\\5\\6\end{bmatrix} = \begin{bmatrix}7\\4\\-3\end{bmatrix}$$

$$\begin{bmatrix}x_1\\-2x_1\\-5x_1\end{bmatrix} + \begin{bmatrix}2x_2\\5x_2\\6x_2\end{bmatrix} = \begin{bmatrix}7\\4\\-3\end{bmatrix}$$

$$\begin{bmatrix}x_1 + 2x_2\\-2x_1 + 5x_2\\-5x_1 + 6x_2\end{bmatrix} = \begin{bmatrix}7\\4\\-3\end{bmatrix}$$

Questions:

- What do we mean by $A\boldsymbol{x} = \boldsymbol{b}$?
- What is the solution of $A\boldsymbol{x} = \boldsymbol{b}$?
- Why do we care about the solution?
- How do we solve the system?
- What does it mean if there is no solution to the system?
- What is a linear combination?
- Can we interpret $A\boldsymbol{x} = \boldsymbol{b}$ in terms of a linear combination?
- In terms of linear combination, what is the meaning of Ax=b being consistent or being inconsistent?