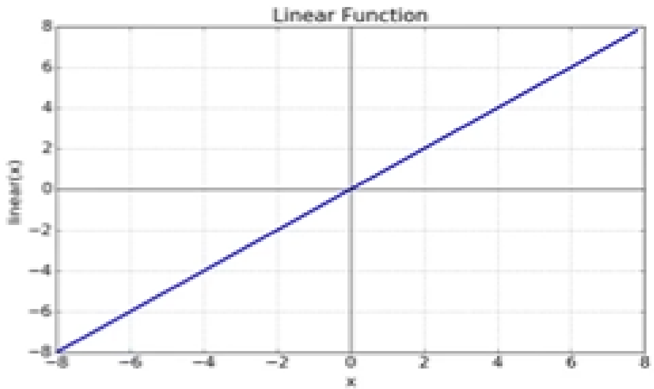
Linear algebra

- Linear algebra is a branch of mathematics that is widely used throughout science and engineering.
- A good understanding of linear algebra is essential for understanding and working with many machine learning algorithms.
- The study of linear algebra involves several types of mathematical objects:

Mathematical Objects needed for Deep Learning



Scalar

- Scalars are just numbers. They have magnitude but no direction.
- Scalars are commonly used to denote quantities such as temperature, time, mass, or distance.
- Any single value from the dataset would represent a scalar.

E.g., The feature x_1 in the dataset 'house price' would be a scalar.

- If dataset contains only single feature as an input, then it is represented as a scalar value.

```
print(house_price_df['bedrooms'].values[0])
3
```

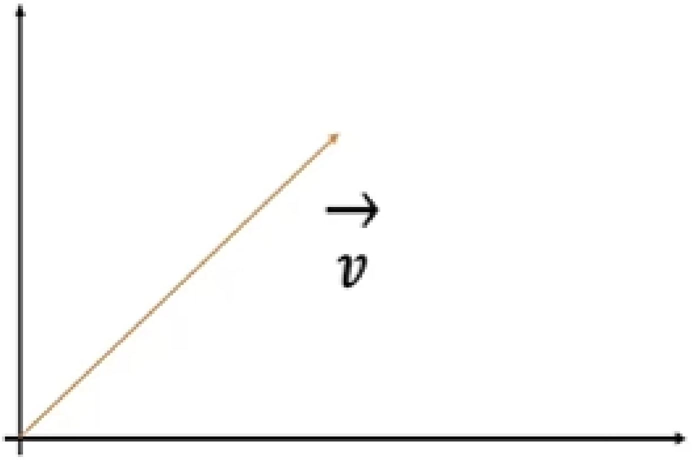
Scalar
3

Samples	x_1 bedrooms	y_i
s_1	**3**	1
s_2	2	0.5
…	…	…
s_n	4	5

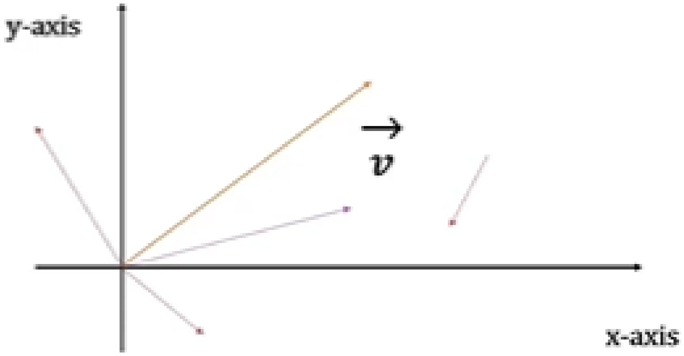
Vector

- Vector is an object with magnitude and direction.
- Arrow indicates the direction
- Line length represents the magnitude

Single vector starts from origin



Multiple vectors



Vector Usage

- A vector is a mathematical object that represents a collection of values **or** features.
- Vectors in deep learning are typically represented as column vectors i.e., written vertically.
- A vector in deep learning is denoted as $X = [x_1, x_2, x_3, \dots, x_n]^T$
- X represents the vector, and $x_1, x_2, x_3, \dots, x_n$ are the individual components of the vector.
- The superscript ^{T} indicates the transpose operation, which converts a row vector into a column vector.
- It is an ordered list of numbers. E.g. $\mathbf{s_1} = [\mathbf{3}, \mathbf{0}, \mathbf{1}, \dots, \mathbf{0}]$
- Each number corresponds to a specific component of the vector.

E.g., **Row** Vector :$[\mathbf{3}, \mathbf{0}, \mathbf{1}, \dots, \mathbf{0}]$

Column Vector: $\begin{bmatrix} \mathbf{3} \\ \mathbf{0} \\ \mathbf{1} \\ \dots \\ \mathbf{0} \end{bmatrix}$

Samples	x_1	x_2	x_3	…	x_n	y_i
s_1	3	0	1	…	0	1
s_2	4	1	0	…	0	0
…	…	…	…	…	…	…
s_n	2	2	2	…	3	0

Vector Operations

- Element-wise multiplication involves multiplying each corresponding element of two vectors together.
- This operation is commonly used in various deep learning applications, such as **attention mechanisms, weighted computations**, and **element-wise interactions.**

Example 1:

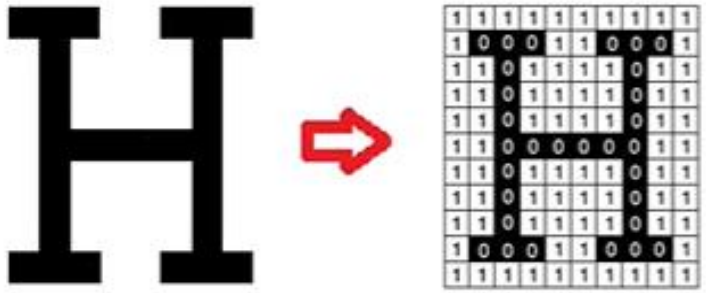
- Vector A: [2, 3, 4]   Vector B: [1, 2, 3]
- Element-wise multiplication, $A \odot B = [2*1, 3*2, 4*3] = [2, 6, 12]$

Example 2:

- Vector X: [1, 2, 3]   Vector Y: [4, 5, 6]
- Dot product $X \cdot Y = (1*4) + (2*5) + (3*6) = 4 + 10 + 18 = 32$

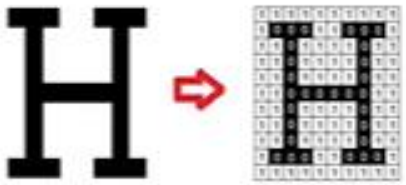
Matrix

- Matrices play a significant role to provide a structure to represent and manipulate the data.
- In deep learning, matrices are used to represent various components such as input data, weights, and gradients in neural networks.
- They enable efficient computation and manipulation of data within neural networks, supporting various applications across **image processing, natural language processing, recommender systems,** and **generative modeling**



Example

Image Processing: Used to represent images in deep learning for tasks like image classification, object detection, and image generation.



- Natural Language Processing: Matrices are employed to represent textual data in tasks such as sentiment analysis, language translation, and text generation.

E.g., : Consider the sentences:

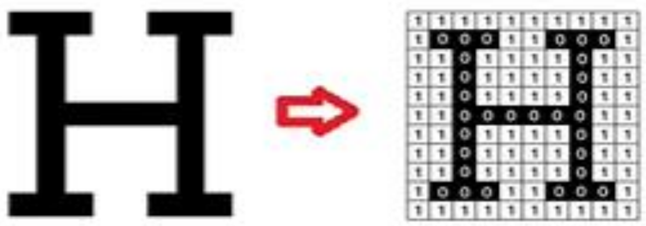
Corpus: I like deep learning. I like NLP. I enjoy flying.

From the above corpus, the list of unique words present are as follows:

Dictionary: ['I', 'like', 'enjoy', 'deep', 'learning', 'NLP', 'flying', '.']

counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

Matrix operations



1. Element-wise matrix multiplication (Hadamard product)
2. Matrix multiplication
3. Matrix Dot product

- **Matrix operations** are key tools in deep learning that enable efficient computation, manipulation, and transformations of data within neural networks.

Element-wise Multiplication (Hadamard product):

- **Element-wise matrix multiplication (Hadamard product):** Element-wise operations are performed on corresponding elements of two matrices **or** a matrix and a scalar.
- E.g., **np. multiply(array a, array b):**
- Multiplication of a matrix A of dimension $m * n$ and a matrix B of dimension m*n is given by:

$$C_{ij} = A_{ij} \odot B_{ij}$$

- For matrices of different dimensions (m × n and p × q, where m ≠ p or n ≠ q), the Hadamard product is undefined. Array **A** = [[3,2],[1,0]] Array **B** = [[0,4],[3,0]]
- E.g.,

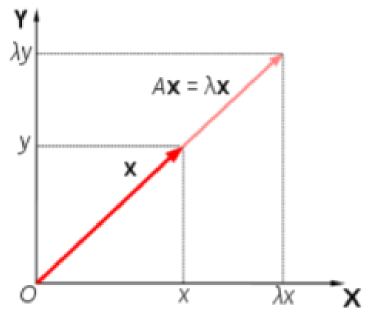
$$A_{ij} \odot B_{ij} = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix} \odot \begin{bmatrix} 0 & 4 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 3*0 & 2*4 \\ 1*3 & 0*0 \end{bmatrix} = \begin{bmatrix} 0 & 8 \\ 3 & 0 \end{bmatrix}$$

Eigenvalues and eigenvectors

- Eigenvalues and eigenvectors are concepts from linear algebra that are used to analyze and understand linear transformations, particularly those represented by square matrices.
- They are used in many different areas of mathematics, including machine learning and artificial intelligence.

Eigenvalues and Eigenvectors

Given a square matrix $A \in \mathbb{R}^{n \times n}$, if following holds, $Ax = \lambda x, \quad x \neq 0$,
λ is said to be its **eigenvalue** and x is its corresponding **eigenvector**.



Intuitively, this means that multiplying the matrix A to x results in a new vector that is pointing to the same direction as x, scaled by the factor of λ.

Eigenvalues and Eigenvectors

The equation $Ax = \lambda x, \quad x \neq 0$ can be rewritten into the following equation:
$(\lambda I - A)x = 0, \quad x \neq 0$

However, the above equation has a non-zero solution to x if and only if it has a non-empty nullspace. That is, $(\lambda I - A)$ is singular.

$|(\lambda I - A)| = 0$

Can obtain eigenvalues with the above equation.

Corresponding eigenvectors for each λ_i can be obtained by solving $(\lambda_i I - A)x = 0$

(However, these are not practical methods used for obtaining eigenvalues and eigenvectors)

Steps for calculating Eigenvalues

Step 1. Form the matrix $B = (A - \lambda I)$

Step 2. Create an equation using: Determinant of B = 0. It is a polynomial equation in λ, $p(\lambda) = 0$

Step 3. Solve $p(\lambda) = 0$. These are the eigenvalues of the matrix A

Example: Find eigenvalues

- Find eigenvalues of the 2 x 2 matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$.

Step 1. Form Matrix

$A - \lambda I = \begin{bmatrix} 1 - \lambda & 2 \\ 3 & 0 - \lambda \end{bmatrix}$.

Step 2. Create Equation

$\det(A - \lambda I) = 0$ gives the equation:

$(1 - \lambda)(- \lambda) - 6 = 0$

Example: Find eigenvalues

The 2×2 matrix, resulted in a second-degree polynomial in our equation.

$\lambda^2 - \lambda - 6 = 0$

Step 3. Solve

$(\lambda - 3)(\lambda + 2) = 0$

$\lambda = 3$ or $\lambda = -2$

Eigenvalues are 3 and -2

Steps for Calculating Eigenvectors

We can easily solve the original equation $Ax = \lambda x$ for eigenvectors using the eigenvalue.

Step 1. Find eigenvalues λ of A

Step 2. For **each** λ, form homogeneous system of linear equations $(A - I\lambda)x = 0$.

Step 3. Solve the above equations to get eigenvectors for λ

Example: Find eigenvectors

- Find eigenvectors of the 2 x 2 matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$.

Step 1. Find Eigenvalues

We have already calculated the eigenvalues as -2 and 3.

First we find the eigenvector for $\lambda = -2$.

Step 2. Form Equations for $\lambda = -2$

$$(A - I\lambda) = \begin{bmatrix} 1-(-2) & 2 \\ 3 & 0-(-2) \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix}$$

Example: Find eigenvectors

- **Eigenvector for $\lambda = 3$:**

We now solve $(A - I\lambda)x = 0 \rightarrow \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} = 0$

We get the equation: $-2x_1 + 2x_2 = 0 \rightarrow x1=x2$

Again, we can choose any value for x_2 to calculate an eigenvector.

For $x_2 = 1$ we get $x_1 = 1$

Eigenvector is $(1, 1)^T$

Example: Find eigenvectors

- **Step 3. Solve Equations**

We now solve $(A - I\lambda)x = 0$ ➔ $\begin{bmatrix} 3 & 2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \end{bmatrix} = 0$

We get the equation: $3x_1 + 2x_2 = 0$ ➔ $x1 = \frac{-2}{3} x2$

We can choose any value for x_2 to calculate the eigenvector.

For $x_2 = 3$ we get $x_1 = -2$

Eigenvector is $(-2, 3)^T$

Any value of x_2 results in an eigenvector. They are all scaled by a factor. **Any scalar multiple of an eigenvector is also an eigenvector.**

Examples for each

- Transpose of a matrix
- Inverse
- Trace of a matrix
- Determinant
- Symmetric matrix