Mathematics for Machine Learning

Lecture 2
(18.04.2024)

Vector Spaces

Mahammad Namazov

Table of contents

Basic Concepts

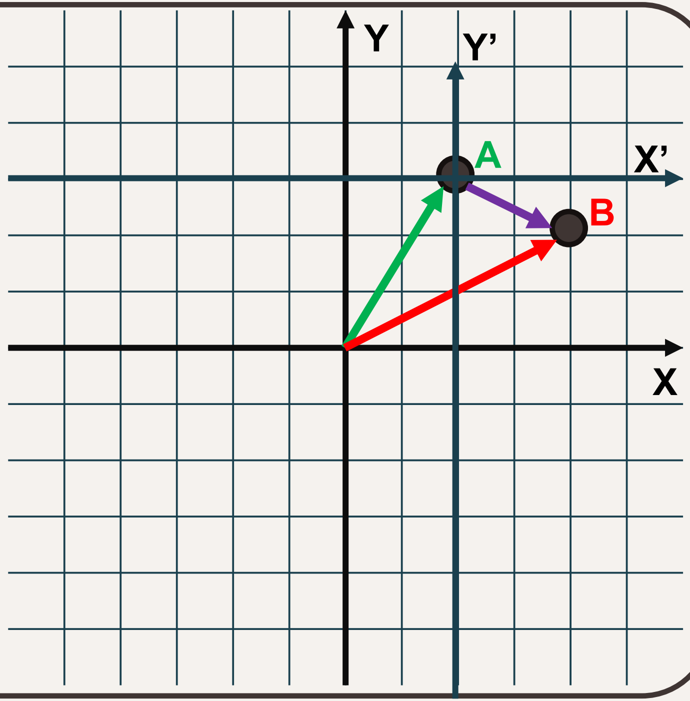
- Vectors
- Reference Systems
- Vector Representations
- Vector Operations

Vectors

- The mathematical objects with magnitude and directions;
- We will represent vectors with lower-case bold italic letters (e.g., $\boldsymbol{x}, \boldsymbol{y}$);
- Vectors are usually given in the form of a column, which might have several rows but one column (i.e., column vector);
- There is also row vectors, which have one row and several columns (depending on the dimension);
- Transpose of row vector is a column vector and vice versa;

Reference Systems

- Assume we have the following reference system:
- Assign the point A (2, 3)
- Assign the point B (4, 2)
- Representations with respect to origin
- Let's go from A to B
- Can we represent it as (2, -1) ?
- No! or Maybe Yes!



Vector Representation

- Vectors are usually represented as columns:

$$\boldsymbol{v} = \begin{bmatrix} v_{1,1} \\ v_{2,1} \\ \vdots \\ v_{n,1} \end{bmatrix}$$

- However, sometimes you might see transpose of it:

$$\boldsymbol{v}^{\boldsymbol{T}} = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,n} \end{bmatrix}$$

- Notations:
  - For each vector and matrix, each element has its row and column index;
  - Usually, i and j are used for row and column, respectively;
  - For instance, $v_{3,1}$ stands for an element at the 3^{rd} row and the 1^{st} column;
  - Vectors are special matrices, which has only one column;

Sum & Subtraction

Vector Operations

- Assume we have 2 vectors of a and b. To add them up or subtract one from another:
  - Number of rows of $\boldsymbol{a}$ and $\boldsymbol{b}$ must be equal;
  - Number of columns of $\boldsymbol{a}$ and $\boldsymbol{b}$ must be equal;
  - Result of sum (or subtraction) is always a vector;

- In other words, dimensionality must be preserved to perform these operations;

- In math representation:

$$\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}; \quad \forall \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u} \in \mathbb{R}^n, \forall n \in \mathbb{R}$$

  - $\boldsymbol{v}$ and $\boldsymbol{w}$ have n-dimensions

Scalar Multiplication

Vector Operations

- When you multiply a vector with scalar, you do one of three options:

$$\boldsymbol{w} = a\boldsymbol{v}, \qquad \forall a \in \mathbb{R}$$

  - Scale it up (i.e., lengthen): $a > 1$;
  - Don't change: $a = 1$;
  - Scale it down (i.e., shorten): $0 < a < 1$;

- None of these three does not change direction, but length!

- To change the direction of any vector, multiply it by -1;

Vector Spaces

Groups

Vector Spaces

Vector Subspaces

Groups

- **Definition:** Consider a set $\mathcal{G}$ and an operation $\otimes: \mathcal{G}\times\mathcal{G} \to \mathcal{G}$, defined on $\mathcal{G}$. Then $G := (\mathcal{G}, \otimes)$ is called a group if the following hold:
  - **Closure of $\mathcal{G}$ under $\otimes$:** $\forall x, y \in \mathcal{G}$: $x \otimes y \in \mathcal{G}$
  - **Associativity:** $\forall x, y \in \mathcal{G}$: $(x \otimes y) \otimes z = x \otimes (y \otimes z)$
  - **Neutral Element:** $\exists e \in \mathcal{G}\ \forall x \in \mathcal{G}$: $x \otimes e = x = e \otimes x$
  - **Inverse Element:** $\forall x \in \mathcal{G}\ \exists y \in \mathcal{G}$: $x \otimes y = e = y \otimes x$

  Notice: The inverse element is defined with respect to (w.r.t.) the operation $\otimes$ and does not necessarily mean $\frac{1}{x}$

- **Abelian Group:** If additionally, $\forall x, y \in \mathcal{G}: x \otimes y = y \otimes x$, then $G := (\mathcal{G}, \otimes)$is an Abelian group **(Commutativity)**

- **Some notes:**
  - $\times$ in $\mathcal{G}\times\mathcal{G} \to \mathcal{G}$ does not stand for multiplication, but mapping;
  - $\otimes$ does not stand for multiplication, but an operation (can be sum or multiplication)

Introduction

Vector Operations

- **In groups we considered inner operation which was applied to elements within the group;**
  - i.e., What happens in Vegas, it remains in Vegas!
- Consider this: Inner operation is addition, and outer operation is scaling (multiplication with scalar)
- Now what we know as ingredients:
  - What is a group?
  - What is an inner operation?
  - What is an outer operation?

Definition

Vector Spaces

- **Definition. A real valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations:**

$$+ \ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$
$$\cdot \ : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

- $(\mathcal{V}, +)$ is an **Abelian group**;
- **Distributivity:**
  - $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$: $\lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \cdot \boldsymbol{y}$
  - $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V}$: $\boldsymbol{x} \cdot (\lambda + \psi) = \boldsymbol{x} \cdot \lambda + \boldsymbol{x} \cdot \psi$
- **Associativity (outer):**

$$\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V}: \quad \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda \psi) \cdot \boldsymbol{x}$$

- **Neutral element w.r.t. outer operation:**

$$\forall \boldsymbol{x} \in \mathcal{V}: \quad 1 \cdot \boldsymbol{x} = \boldsymbol{x}$$

Vector Spaces

Vector Spaces

- Then vector space is a set of vectors with addition and scaling operations, where the followings hold for each vector belongs to it $(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}, \mathbf{0} \in V)$:
  - **Commutativity:**
    $$\boldsymbol{v} + \boldsymbol{u} = \boldsymbol{u} + \boldsymbol{v}$$
  - **Associativity:**
    $$(\boldsymbol{v} + \boldsymbol{u}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$$
  - **Additive Identity:**
    $$\mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$$
  - **Additive Inverse:**
    $$\boldsymbol{v} + \boldsymbol{w} = \mathbf{0}$$
  - **Multiplicative Identity:**
    $$1\boldsymbol{v} = \boldsymbol{v}$$
  - **Distributivity:**
    $$a(\boldsymbol{v} + \boldsymbol{u}) = a\boldsymbol{u} + a\boldsymbol{v}$$
    $$(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$$

Abelian Group $(\mathcal{V}, +)$:
Neutral Element $\mathbf{0} \in \mathcal{V}$

Complementary for $(\mathcal{V}, +, \cdot)$:
Neutral Element $1 \in \mathbb{R}$

Note: All operations must end up within the vector space: **Closure of Group**

Vector Subspaces

- We already defined vector space of $V = (\mathcal{V}, +, \cdot)$;

- Now assume there is a subset of $\mathcal{V}$ which is not an empty set:

$$\mathcal{U} \subseteq \mathcal{V}, \qquad \mathcal{U} \neq \emptyset$$

- Then $U = (\mathcal{U}, +, \cdot)$ is called a vector subspace of V (or linear subspace) if
  - Operations of this vector space are + and ·
  - And applications of these operations are restricted with $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$

- The vector subspace of vector space can be shown as:

$$U \subseteq V$$

- We need to show following details to determine if any $(\mathcal{U}, +, \cdot)$ is a subspace of V:
  - $\mathcal{U} \neq \emptyset$, in particular $\mathbf{0} \in \mathcal{U}$;
  - Closure of U:
    - With respect to the outer operation: $\forall \lambda \in \mathbb{R}, \forall\, \boldsymbol{x} \in \mathcal{U} : \lambda \boldsymbol{x} \in \mathcal{U}$
    - With respect to the inner operation: $\forall \lambda \in \mathbb{R}, \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} : \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$

Linear Independence

Introduction

Linear Combination

Linear Independence

Basis and Span

Definition

- If we add vectors (from same Vector Space) and scale any of them, we will end up with a vector in the given Vector Space.
  - In other words, Closure property is still on play
- Thanks to ***basis vectors*** of any Vector Space, we can reach any point in the Vector Space, by adding and scaling them (or by doing both at the same time).
- To understand operations better, we need to follow this order:
  - Linear Combination;
  - Linear Independence;
  - Basis Vectors

Linear Combination

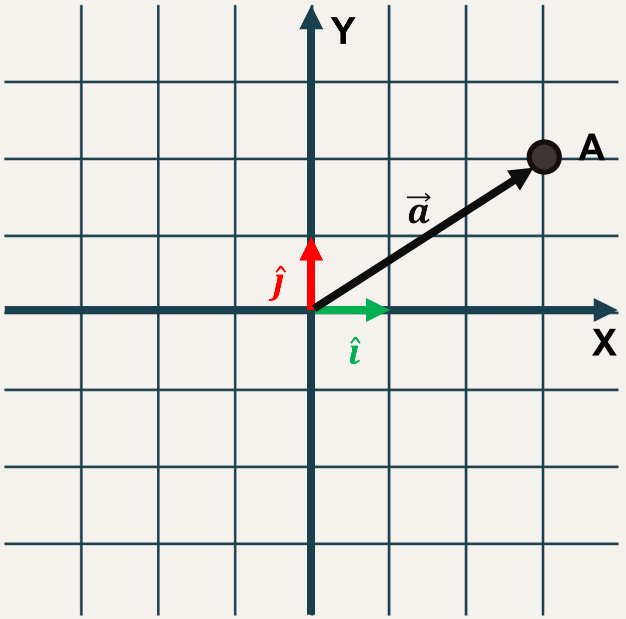
- Basis vectors (no need to understand for now):

$$\hat{\imath} = \begin{bmatrix}1\\0\end{bmatrix}, \hat{\jmath} = \begin{bmatrix}0\\1\end{bmatrix}$$

- Remember point A?
- From origin to A:

$$\vec{a} = \begin{bmatrix}3\\2\end{bmatrix}$$

- OR as a linear combination of basis vectors:



Linear Combination

- Basis vectors (no need to understand for now):

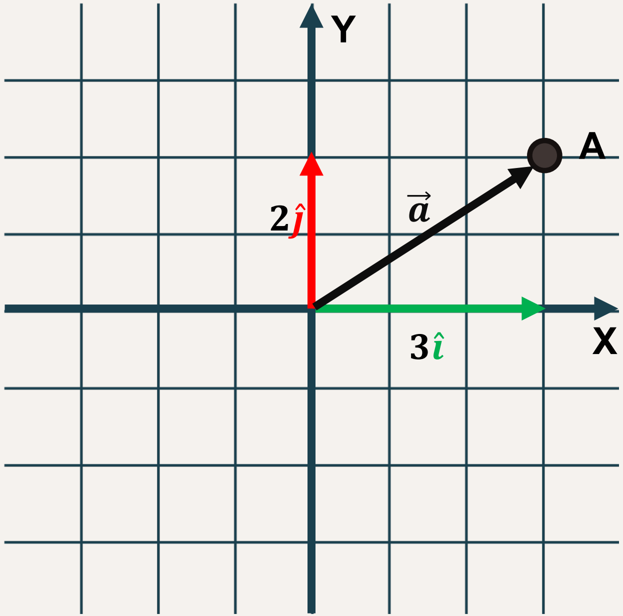
$$\hat{\imath} = \begin{bmatrix}1\\0\end{bmatrix},\ \hat{\jmath} = \begin{bmatrix}0\\1\end{bmatrix}$$

- Remember point A?
- From origin to A:

$$\vec{a} = \begin{bmatrix}3\\2\end{bmatrix}$$

- OR as a linear combination of basis vectors:

$$\vec{a} = 3\hat{\imath} + 2\hat{\jmath} = 3\begin{bmatrix}1\\0\end{bmatrix} + 2\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}3\\2\end{bmatrix}$$



Linear Combination Scenarios

Linear Combination

- **Definition:**
  Consider a vector space V, and a finite number of vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_k} \in V$. Then, every $\boldsymbol{v} \in V$ of the form:
  $$\boldsymbol{v} = \boldsymbol{\lambda_1 x_1} + \dots + \boldsymbol{\lambda_k x_k} = \sum_{i=1}^{k} \boldsymbol{\lambda_i x_i} \in V$$
  With $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\boldsymbol{x_1}, \dots, \boldsymbol{x_k}$

- **Scenario 1:** When 2 vectors are linearly independent:
  $$\boldsymbol{a} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$
  - We can reach any point on 2D with them (since the example is given in 2D);

- **Scenario 2:** When 2 vectors are linearly dependent:
  $$\boldsymbol{a} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 6 \\ 4 \end{bmatrix} = 2 \times \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$
  - We can reach any point on the line, but cannot reach any point on the plane;

- **Scenario 3:** When 2 vectors are zero vectors:
  - We can represent only the origin, with them

Basics

Linear Independence

- Two vectors are linearly independent, if one cannot be expressed by some scalar multiplication of the other;

$$\boldsymbol{a} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

- Three vectors are linearly independent, if one cannot be represented as a linear combination of the others;

$$\boldsymbol{a} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}, \boldsymbol{c} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

- Let’s test it, before generic formulations;

Definition

Linear Independence

- We have again the same vector space V and $k \in \mathbb{N}$ vectors.

- Assume, there is such linear combination that brings us to the origin (i.e., provides zero vector $\mathbf{0}$)

$$\mathbf{0} = \boldsymbol{\lambda_1 x_1} + \cdots + \boldsymbol{\lambda_k x_k}$$

- If at least one of the coefficients (i.e., lambdas) is not zero ($\lambda_i \neq 0$), then these vectors are linearly dependent;

- If the equation is correct if and only if all coefficients are zero ($\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$), then these vectors are linear independent.

- WHY?

Basis and Span

- Generating Set:
  - We are again in the same vector space $V = (\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_k}\} \in V$. If every vector $\boldsymbol{v} \in V$ can be represented using these vectors, then we call $\mathcal{A}$ as generating set

- Span:
  - The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans the vector space V, we write:
$$V = span[\mathcal{A}] \text{ or } V = span[x_1, x_2, \dots, x_k]$$
  - Notice that, generating sets are sets of vectors that span vector (sub)spaces;

- Basis:
  - We already defined vector space V and $\mathcal{A}$, and we know that $\mathcal{A}$ is also a subspace of V ($\mathcal{A} \subseteq V$);
  - $\mathcal{A}$ is called minimal if there exists no smaller set $\bar{\mathcal{A}} \subseteq \mathcal{A} \subseteq V$ that spans V;
  - Every **linearly independent** generating set of V is **minimal** and is called a **basis** of V.

Matrices

Linear Transformation

Linear Mapping

Significance

Linear Transformation

- Suppose we have 2 vector spaces V and W;
- Dimensionality is not our problem for now;
- Assume we have:
  - $\boldsymbol{v} \in V$ that we want to transform it into W;
  - There is such linear transformation L that:

$$L\colon V \to W$$

- Properties for L:
  - Additivity:

$$L(\boldsymbol{v} + \boldsymbol{w}) = L(\boldsymbol{v}) + L(\boldsymbol{w})$$

  - Homogeneity:

$$L(c\boldsymbol{v}) = cL(\boldsymbol{v})$$

Linear Mapping

- In Linear Algebra Language:
  - Assuming V and W are two vector spaces, and $L: V \rightarrow W$, then homogeneity and additivity will hold for any vector in these vector spaces;
- Chance to represent any vector $\boldsymbol{v} \in V$ in the W;
- In AI, hidden layers work with this principle;
- You have an image, a word, anything that can be represented as a vector;
  - Current representation is okay for you, but ask to computer;
  - Represent it in different dimensions, different spaces;
  - Voila, it recognizes how to understand the information

Significance

- These linear transformations are represented with matrices;
- Matrices are mathematical objects that:
  - Has m rows;
  - Has n columns;
- A matrix $M_{m \times n}$ actually:
  - Maps vectors from m-dimensional vector space into n-dimensional vector space;
  - M shows from where, n shows to where;
- More details, next week!

Conclusion

Summary

Takeaways

References

Summary

- When you define vector in some space, pay attention to reference system;
- We cannot visualize when dimension of the vector is higher than 3;
- Vector spaces are set of vectors with specific operations;
- If vector space is m-dimensional, m linear independent vectors are enough to represent it;
- Basis vectors play significant role to represent vector space;
- Linear Transformation is a mapping tool between vector spaces;

Takeaways

- What we have seen today may not be seem crucial, but think as a building;
- We need to know what are vectors to define Vector Spaces;
- We need to know Vector Spaces to make linear transformation among or between them;
- We need to know Linear Transformation to see what AI models do actually;
- Are these enough? No, we have a lot to talk!

References

- For further information to read more detailed:
  - Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
  - Chapter 2, Sections: 4, 5, 6

- For imagination what is going on:
  - https://www.youtube.com/watch?v=fNk_zzaMoSs&ab_channel=3Blue1Brown
  - https://www.youtube.com/watch?v=k7RM-ot2NWY&ab_channel=3Blue1Brown

The End

Thanks for your attention and patience!

Mahammad Namazov