Mathematics for Machine Learning

Lecture 3
(25.04.2024)

Linear Transformation

Mahammad Namazov

Table of contents

Recap

Vectors

Vector Spaces

Matrices

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 ($\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{0} \in \boldsymbol{V}$):
  - **Commutativity:**
  $$\boldsymbol{v} + \boldsymbol{u} = \boldsymbol{u} + \boldsymbol{v}$$
  - **Associativity:**
  $$(\boldsymbol{v} + \boldsymbol{u}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$$
  - **Additive Identity:**
  $$\boldsymbol{0} + \boldsymbol{v} = \boldsymbol{v}$$
  - **Additive Inverse:**
  $$\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{0}$$
  - **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 $\boldsymbol{0} \in \mathcal{V}$

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

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

Matrices

- Mapping from one vector space to another:

$$L: V \rightarrow W$$

- For transforming from m-dimensional space to n-dimensional space we need a matrix M, which has:
  - m rows;
  - n columns;

- Have significant role in AI applications;

- About details, we will dive deeper now;

Linear Transformation

Analysis

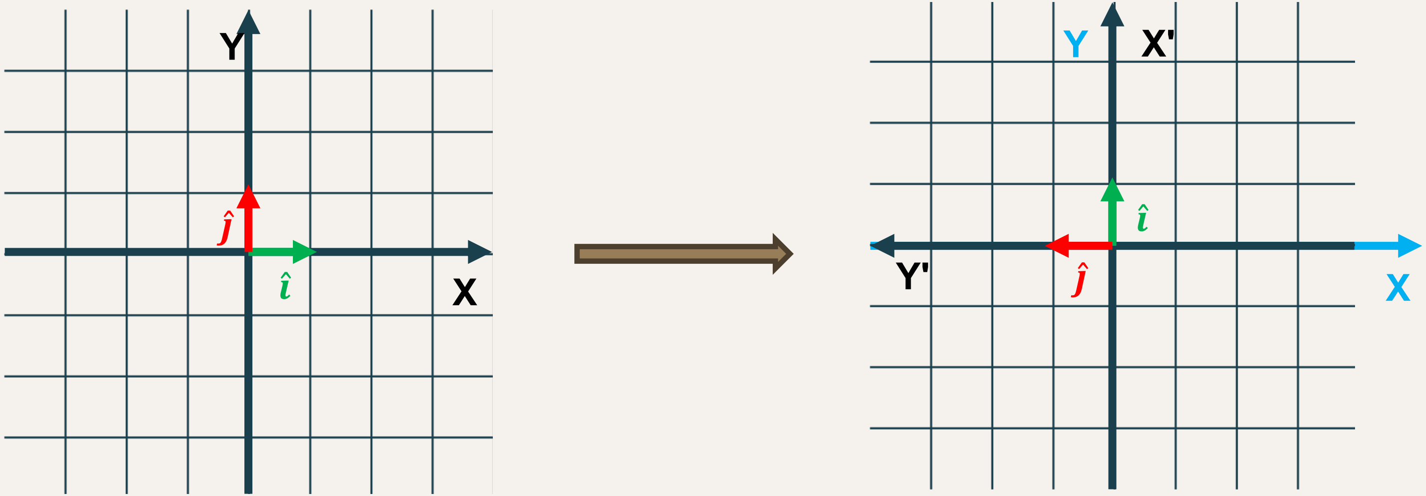
Matrix Multiplication

Composition

Properties

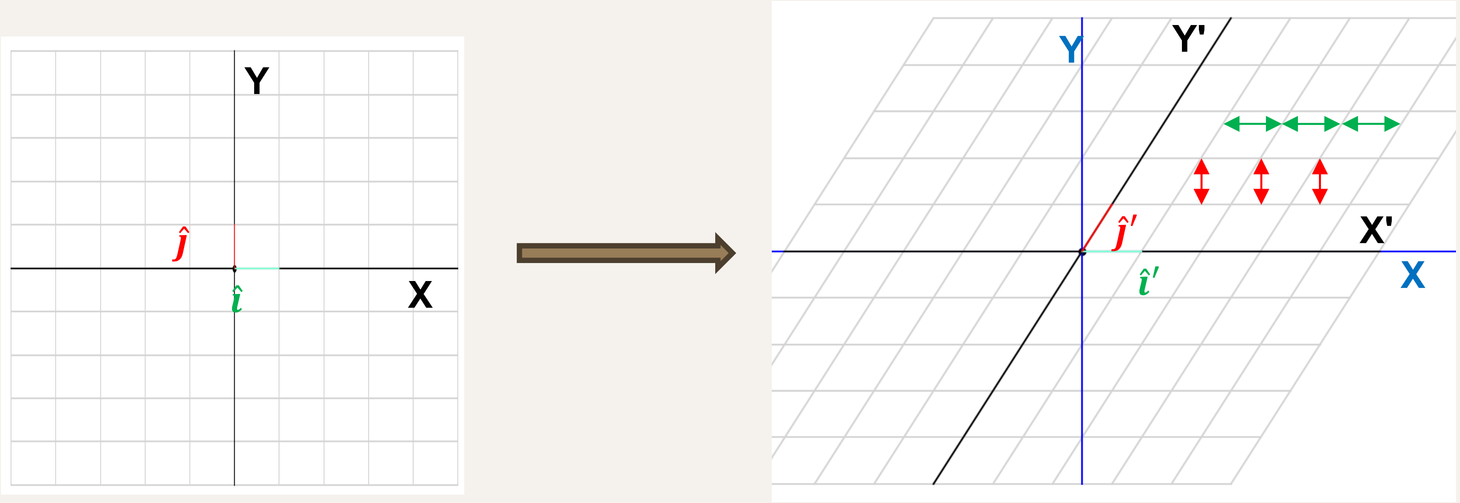
Rotation

- For Transformation $L: V \rightarrow W$, how can we know if it is linear? (e.g., $90°$ c.c.w.)
  - Origin will not be affected;



Shear

- For Transformation $L: V \to W$, how can we know if it is linear?
  - Grids will be evenly spaced;



Matrix Multiplication

- We can say that matrix multiplication is a linear transformation that maps a vector from one vector space to another;

- Assuming we want to transform vector $\boldsymbol{v} \in V$ in a way that we get $\boldsymbol{w} \in W$. Mathematically:

$$M\boldsymbol{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} av_1 + bv_2 \\ cv_1 + dv_2 \end{bmatrix} = \boldsymbol{w}$$

- What if we use basis vectors to multiply a matrix with a vector?
  - Assume $\hat{\imath}$ and $\hat{\jmath}$ are basis vectors of V;
  - Then $\boldsymbol{v}$ can be represented with them;
  - Using this information, what can we deduce about M?

Matrix Multiplication

- $\boldsymbol{v} \in \mathbb{R}^n$ and we would like to get vector $\boldsymbol{w} \in \mathbb{R}^n$ (i.e., input and output vectors are in the same shape);
- To perform this transformation, we will need a Matrix M, which has $n \times n$ shape
- Then, we can say that i^{th} column of M tells where the i^{th} basis vector of V will land in W, where $i \in \{1, 2, \ldots, n\}$.

$$\begin{bmatrix} m_{1,1} & \cdots & m_{1,j} & \cdots & m_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ m_{j,1} & \cdots & m_{j,j} & \cdots & m_{(j,n)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,j} & \cdots & m_{n,n} \end{bmatrix} \begin{bmatrix} v_{1,1} \\ \vdots \\ v_{j,1} \\ \vdots \\ v_{n,1} \end{bmatrix} = \begin{bmatrix} w_{1,1} \\ \vdots \\ w_{j,1} \\ \vdots \\ w_{n,1} \end{bmatrix} = \begin{bmatrix} m_{1,1} \\ \vdots \\ m_{j,1} \\ \vdots \\ m_{n,1} \end{bmatrix} v_{1,1} + \cdots + \begin{bmatrix} m_{1,j} \\ \vdots \\ m_{j,j} \\ \vdots \\ m_{n,j} \end{bmatrix} v_{j,1} + \cdots + \begin{bmatrix} m_{1,n} \\ \vdots \\ m_{j,n} \\ \vdots \\ m_{n,n} \end{bmatrix} v_{n,1}$$

Composition

- Assume we would like to apply several transformations to $\boldsymbol{v} \in \mathbb{R}^n$ to get $\boldsymbol{u} \in \mathbb{R}^n$;
  - Initially, we would like to apply rotation with some arbitrary degree: M_1
$$M_1\boldsymbol{v} = \boldsymbol{w}$$
  - Then, we would like to apply shear with M_2;
$$M_2\boldsymbol{w} = \boldsymbol{u}$$
  - Composing these transformations:
$$M_2M_1\boldsymbol{v} = M\boldsymbol{v} = \boldsymbol{u}$$
- Then, by multiplying two or more matrices, we compose several linear transformations!
- Assume we would like to multiply $M_{m\times n}$ with $B_{n\times p}$ to get $C_{m\times p}$, then $c_{i,j}$:
$$c_{i,j} = \sum_{k=1}^{n} m_{i,k}b_{k,p} \,, i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, p\}$$

Properties

- Two matrices can be added, if and only if they are in the same shape (i.e., number of rows and columns are same);
- A and B can be multiplied in the order of AB, if and only if number of rows of A is same with the number of columns of B;
- If A, B and C satisfy the first one, then $A+B+C = A+C+B=B+C+A=...$;
- Two generic cases for matrix multiplication:
  - Associativity:

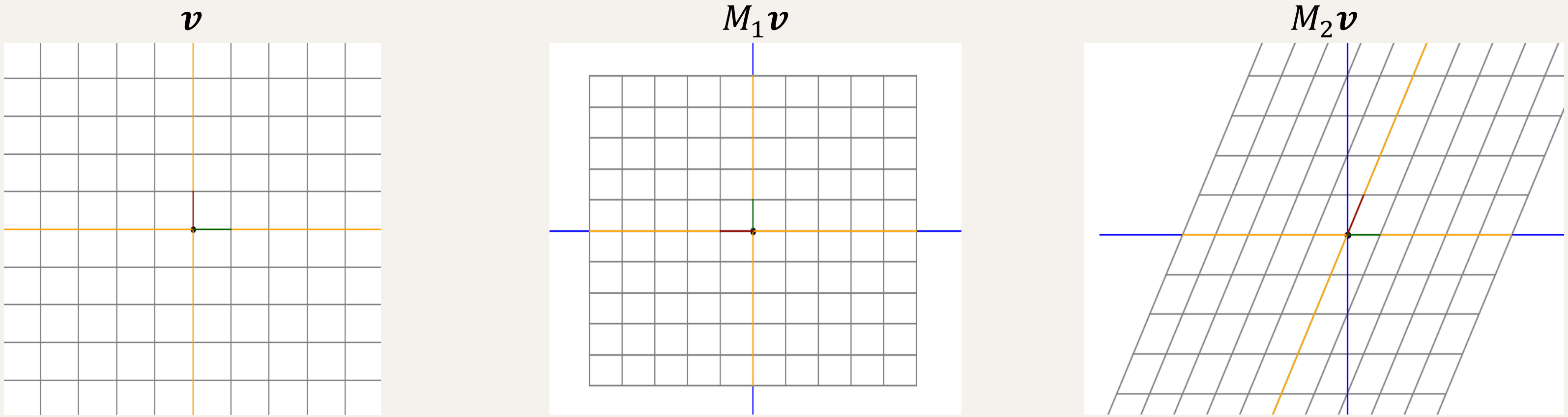
$$A(BC) = (AB)C$$

  - Not commutative:

$$AB \neq BA$$

Not Commutative

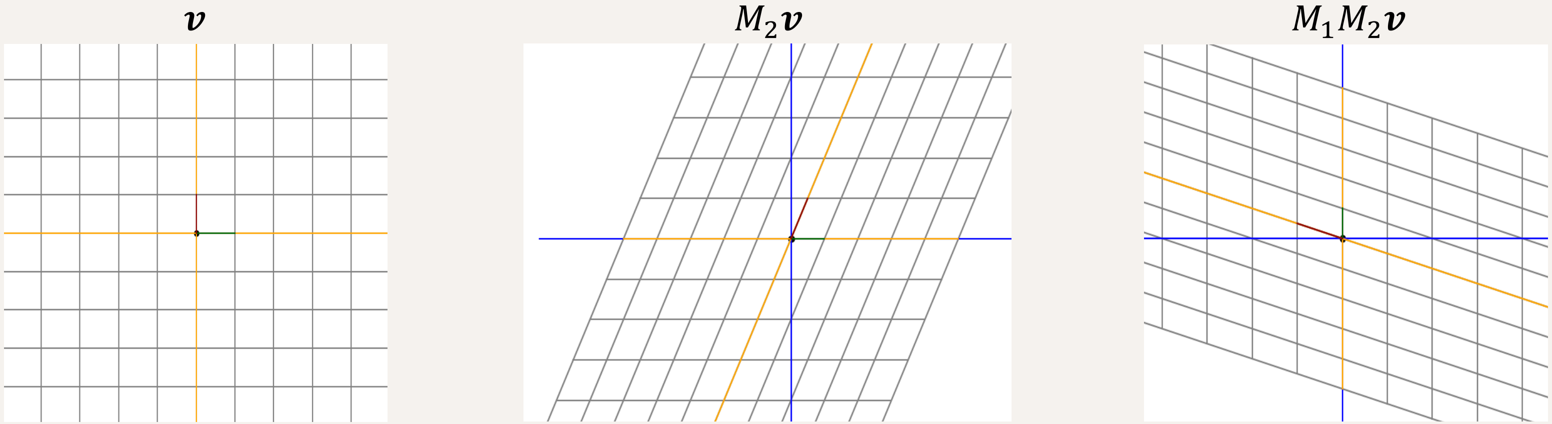
- Let's say M_1 (b) and M_2 (c) stand for rotation and shear, respectively;



(a) (b) (c)

Not Commutative

- Initially we apply M_2 (b) then M_1 (c) (i.e., first rotate then shear)



(a) (b) (c)

Determinant

Meaning?

In 2-D

In 3-D

In n-D

Significance

- Suppose we have 2 vectors in V: $\boldsymbol{v}, \boldsymbol{w} \in V$
- When we combine them, we can get a figure F, with area of S
- We apply L to map these vectors from V to W with the matrix M:

$$L: V \rightarrow W$$

  - We get new vectors let's say: $\boldsymbol{v}', \boldsymbol{w}' \in W$
  - We create new figure F' with S':

$$\det(M) = \frac{S'}{S}$$

- Significant usage: Singularity Detection

In 2-D

- How to compute it?
- We have already a matrix in 2-Dimensional Space:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

- Determinant of this matrix will be:

$$\det(M) = ad - bc$$

- Then we can say: The newly generated figure's area will be modified as much as $\det(M)$

In 3-D: Laplacian Expansion Steps

- We have a Matrix $A_{3\times3}$:
- Now we will visualize how to perform this process;

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$

$$(-1)^{1+1} \; a_{1,1} \times \begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix}$$

$$+ \quad (-1)^{1+2} \; a_{1,2} \times \begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix}$$

$$+ \quad (-1)^{1+3} \; a_{1,3} \times \begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}$$

Laplacian Expansion: Algorithm

- Assuming, we have a Matrix $A_{3\times3}$:
  - $a_{i,j}$ will be an element stays in the i^{th} row and j^{th} column;
  - Choose a row **or** column (e.g., the first row, where $i = 1$)
  - Entries of the chosen row (or column) will be used as multipliers:

$$(-1)^{i+j}a_{i,j}$$

  - Now, we need to have Minors for corresponding i and j:
    - ❖ Minor $M_{i,j}$ is the submatrix of A, which is made by removing i^{th} row and j^{th} column from the matrix A
  - Then we can iterate through each entry in the chosen row (or column) and sum them up:

$$\det(A) = \sum_{j=1}^{n}(-1)^{i+j}a_{i,j}M_{i,j} \text{ for } i^{th} \text{ row}$$
$$\det(A) = \sum_{i=1}^{n}(-1)^{i+j}a_{i,j}M_{i,j} \text{ for } j^{th} \text{ column}$$

- You can expand it into n-D;

Properties

- All matrices in this section are in the same shape: $n \times n$
- Determinant of matrix is zero when it has linear dependent row (or column)
- Assume $AB = C$, then $\det(C) = \det(A)\det(B)$
- Changing two rows of a matrix, will not change the value of the determinant but sign
- Changing two columns of a matrix will affect as row exchange does
- Assuming i^{th} column (or row) can be simplified with multiplier λ.
  - After simplification we get B from the A;
  - Then, $\det(A) = \lambda \det(B)$

Special Matrices

Symmetric Matrix

Diagonal Matrix

Identity Matrix

Inverse Matrix

Symmetric Matrix

- The matrix M is given as below:

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} \\ m_{2,1} & m_{2,2} & m_{2,3} \\ m_{3,1} & m_{3,1} & m_{3,3} \end{bmatrix}$$

- Diagonal elements of the matrix are entries with row index and column index are equal:

$$m_{i,j}, \qquad i = j$$

- Symmetric matrix have such entries that:

$$m_{i,j} = m_{j,i}$$

- Transpose of the Symmetric matrix M is also M:

$$M^T = M$$

Diagonal Matrix

- The matrix M is given as below:

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} \\ m_{2,1} & m_{2,2} & m_{2,3} \\ m_{3,1} & m_{3,1} & m_{3,3} \end{bmatrix}$$

- What if non-diagonal elements are zero:

$$M = \begin{bmatrix} m_{1,1} & 0 & 0 \\ 0 & m_{2,2} & 0 \\ 0 & 0 & m_{3,3} \end{bmatrix}$$

- Voila! M is a diagonal matrix;

- Diagonal Matrix is used for scaling each dimensions of the vector;

Identity Matrix

- Diagonal Matrix is given as below, which we obtained:

$$M = \begin{bmatrix} m_{1,1} & 0 & 0 \\ 0 & m_{2,2} & 0 \\ 0 & 0 & m_{3,3} \end{bmatrix}$$

- When diagonal elements are equal to 1, we call the resulting matrix as an Identity Matrix;

$$M = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- We can call it, “doing nothing matrix”; Determinant?

Inverse Matrix

- Now assume we have a matrix M, which maps vector $\boldsymbol{v} \in V$ into W, where $W, V \in \mathbb{R}^n$;
- Let's say, the resulting vector $\boldsymbol{w} \in W$ is gotten by:
$$\boldsymbol{w} = M\boldsymbol{v}$$
- Now let's apply such transformation $\boldsymbol{v}$ that does not do anything (Hint: Identity):
$$\boldsymbol{v} = I\boldsymbol{v}$$
- On the other hand, let's apply such transformation (with K) that maps w back into $\boldsymbol{v} \in V$:
$$\boldsymbol{v} = K\boldsymbol{w} = KM\boldsymbol{v} = I\boldsymbol{v} \Rightarrow KM = I$$
- K is inverse matrix M. An inverse matrix of any matrix M is denoted with M^{-1}

Conclusion

Summary

Takeaways

References

Summary

- Linear Transformation matrix M specifies new positions of the basis vectors;
- While number of rows identifies the source dimension, number of columns represents dimensionality in destination;
- Determinants carry significant information for matrices;
- Diagonal matrix simply scale the vector features in each dimension;
- Identity matrix does not do anything specific, but useful to make matrix of vectors;
- Every Identity matrix is a Diagonal matrix, but not vice-versa;

Takeaways

- Inverse matrix simply does “undoing” the transformation;

- Matrix inverse can be taken, when:
- It is a square matrix;
- Determinant is not zero (will be clearer in 2 days)

- Understanding Linear Transformation, is a key for further steps in AI models;

- Solving bunch of equations or using a single matrix?

- Digest the last 2 classes for the next one.

References

- Further information to read:
  - Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.
  - Chapter 2, Sections: 3, 7

- Further videos:
  - **Inverse matrices, column space, null space:**
    - ✓ https://www.youtube.com/watch?v=uQhTuRlWMxw&ab_channel=3Blue1Brown
  - **Computational videos (some but not limited to):**
    - ✓ https://www.youtube.com/watch?v=pgqyULjZgbU&list=PLAFEC355DFEADC30C&index=6&ab_channel=patrickJMT
    - ✓ https://www.youtube.com/watch?v=iMQRo0tHORw&list=PLAFEC355DFEADC30C&index=9&ab_channel=patrickJMT
    - ✓ https://www.youtube.com/watch?v=Ey62H_oaqoE&list=PLAFEC355DFEADC30C&index=11&ab_channel=patrickJMT

The End

Mahammad Namazov