Linear algebra: vectors, matrices, dot product

Statistical Natural Language Processing 1

Çağrı Çöltekin

University of Tübingen
Seminar für Sprachwissenschaft

Winter Semester 2025/2026

version: 7763d51 @2025-10-28

Some practical remarks

(recap)

- Course web page: `https://snlp1-2025.github.io` (public)
  `https://github.com/snlp1-2025/snlp1/` (private)
- If you haven't done already, please fill in the questionnaire on Moodle

Today's lecture

- Some concepts from linear algebra
  - Vectors
  - Dot product
  - Matrices

This is only a high-level, informal introduction/refresher.

Linear algebra

Linear algebra is the field of mathematics that studies *vectors* and *matrices*.

- A vector is an ordered sequence of numbers

$$\boldsymbol{v} = (6, 17)$$

- A matrix is a rectangular arrangement of numbers

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

- A well-known application of linear algebra is solving a set of linear equations

$$\begin{array}{rcrcr} 2x_1 & + & x_2 & = & 6 \\ x_1 & + & 4x_2 & = & 17 \end{array} \quad \Longleftrightarrow \quad \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 17 \end{bmatrix}$$

Why study linear algebra?

Consider an application counting words in a document

	the	and	of	to	in	...
	121	106	91	83	43	...

Why study linear algebra?

Consider an application counting words in a document

	the	and	of	to	in	...	
(	121	106	91	83	43	...	)

Why study linear algebra?

Consider an application counting words in multiple documents

	the	and	of	to	in	…
document_1	121	106	91	83	43	…
document_2	142	136	86	91	69	…
document_3	107	94	41	47	33	…
…	…	…	…	…	…	

You should already be seeing vectors and matrices here.

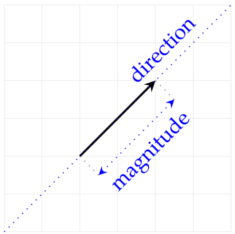
Why study linear algebra?

- Insights from linear algebra are helpful in understanding many NLP methods
- In machine learning, we typically represent input, output, parameters as vectors or matrices (or tensors)
- It makes notation concise and manageable
- In programming, many machine learning libraries make use of vectors and matrices explicitly
- In programming, vector-matrix operations correspond to loops
- 'Vectorized' operations may run much faster on GPUs, and on modern CPUs

Vectors

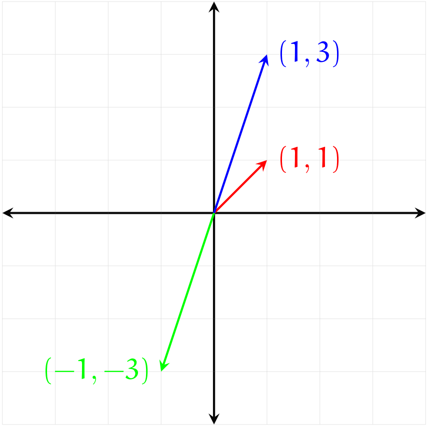
- Vectors are objects with a magnitude and a direction
- We represent vectors with an ordered list of numbers $\boldsymbol{v} = (v_1, v_2, \ldots v_n)$
- The number n (the number of elements or entries of the vector) is its dimension
- We often call an n dimensional vector as n-vector
- The vector of n real numbers is said to be in $\mathbb{R}^n$ ($\boldsymbol{v} \in \mathbb{R}^n$)
- Typical notation for vectors:

$$\boldsymbol{v} = \vec{v} = (v_1, v_2, v_3) = \langle v_1, v_2, v_3 \rangle = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$



Geometric interpretation of vectors

- Geometrically, vectors are represented with arrows from the origin in the Euclidean space
- The endpoint of the vector $\boldsymbol{v} = (v_1, v_2)$ correspond to the Cartesian coordinates defined by v_1, v_2
- These generally make sense for two or three-dimensional spaces
- The intuitions often (!) generalize to higher dimensional spaces



Some special vectors

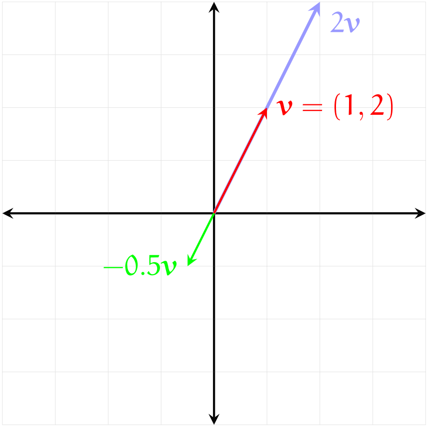
- The zero vector, $\mathbf{0}$, is the vector whose all entries are 0
- The vector of all 1s, $\mathbf{1}$, is also often interesting
- A more interesting set of vectors is *standard unit vectors* (examples below are 4-dimensional standard unit vectors)

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\1\\0\end{bmatrix} \quad \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

- n-dimensional standard unit vectors form the standard *basis* for n-dimensional (vector) space
- In some textbooks, standard unit vectors of two (and three) dimensions are represented by $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$
- In ML they are related to *one-hot representation*: we represent categorical predictors (variables) with n values as n-dimensional standard unit vectors

Multiplying a vector with a scalar

- For a vector $\boldsymbol{v} = (v_1, v_2)$ and a scalar a,
$$a\boldsymbol{v} = (av_1, av_2)$$
- multiplying with a scalar 'scales' the vector
- We can use the notation $a\mathbf{1}$ for a vector whose all entries are a



Vector addition and subtraction

For vectors $\mathbf{v} = (v_1, v_2)$ and $\mathbf{u} = (u_1, u_2)$

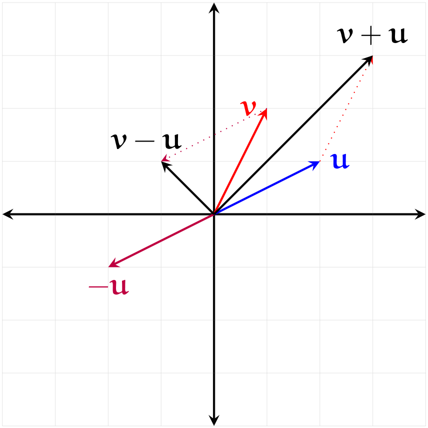
- $\mathbf{v} + \mathbf{u} = (v_1 + u_1, v_2 + u_2)$

$$(1, 2) + (2, 1) = (3, 3)$$

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

$$(1, 2) - (2, 1) = (-1, 1)$$

- For any vector $\mathbf{v}$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$



Properties of vector operations

- Vector addition and scalar multiplication is commutative

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

$$a\mathbf{u} = \mathbf{u}a$$

- Scalar multiplication and vector addition also show the following distributive properties

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

$$(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$$

Linearity and linear functions

- A linear function $f(\cdot)$ (or mapping) follows
  - $f(a\boldsymbol{v}) = af(\boldsymbol{v})$ (homogeneity)
  - $f(\boldsymbol{v} + \boldsymbol{u}) = f(\boldsymbol{v}) + f(\boldsymbol{u})$ (additivity)
  - combined together: $f(a\boldsymbol{v} + b\boldsymbol{u}) = af(\boldsymbol{v}) + bf(\boldsymbol{u})$
- A combination of vectors as in

$$a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \ldots + a_n\boldsymbol{v}_n.$$

is called a linear combination (another vector)

Linearity and linear functions

- A linear function $f(\cdot)$ (or mapping) follows
  - $f(a\mathbf{v}) = af(\mathbf{v})$ (homogeneity)
  - $f(\mathbf{v} + \mathbf{u}) = f(\mathbf{v}) + f(\mathbf{u})$ (additivity)
  - combined together: $f(a\mathbf{v} + b\mathbf{u}) = af(\mathbf{v}) + bf(\mathbf{u})$
- A combination of vectors as in

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n.$$

is called a linear combination (another vector)

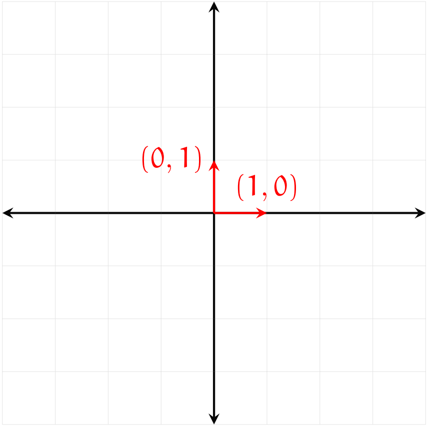
Question: Is $f(x) = ax + b$ linear?

Linear combinations of standard unit vectors

- Any n-vector can be written as a linear combination of standard unit vectors. Example:

$$\begin{bmatrix} 2 \\ -3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- n-dimensional standard unit vectors form a *basis* for $\mathbb{R}^n$
- n-dimensional standard unit vectors form *span* the vector space $\mathbb{R}^n$

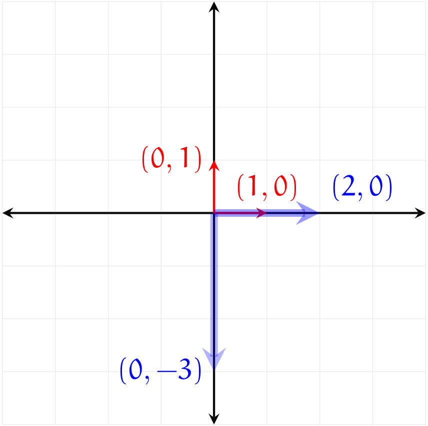


Linear combinations of standard unit vectors

- Any n-vector can be written as a linear combination of standard unit vectors. Example:

$$\begin{bmatrix} 2 \\ -3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- n-dimensional standard unit vectors form a *basis* for $\mathbb{R}^n$
- n-dimensional standard unit vectors form *span* the vector space $\mathbb{R}^n$

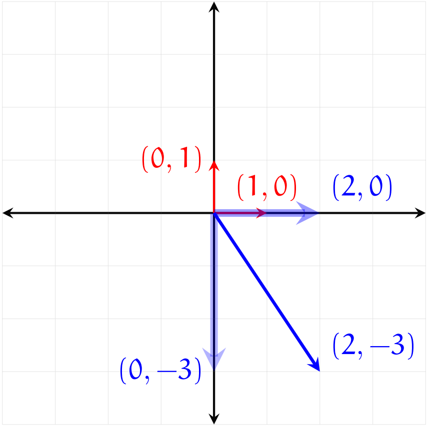


Linear combinations of standard unit vectors

- Any n-vector can be written as a linear combination of standard unit vectors. Example:

$$\begin{bmatrix} 2 \\ -3 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- n-dimensional standard unit vectors form a *basis* for $\mathbb{R}^n$
- n-dimensional standard unit vectors form *span* the vector space $\mathbb{R}^n$



Dot (inner) product

- Dot product is an operation between two vectors with same dimensions

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

Dot (inner) product

- Dot product is an operation between two vectors with same dimensions

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

- Calculate the dot products for the following vectors

$$\begin{bmatrix}4\\3\end{bmatrix} \cdot \begin{bmatrix}3\\4\end{bmatrix} \qquad \begin{bmatrix}4\\-3\end{bmatrix} \cdot \begin{bmatrix}-3\\4\end{bmatrix} \qquad \begin{bmatrix}1\\2\\3\end{bmatrix} \cdot \begin{bmatrix}-2\\-4\\-6\end{bmatrix} \qquad \begin{bmatrix}1\\2\\3\end{bmatrix} \cdot \begin{bmatrix}2\\4\\6\end{bmatrix} \qquad \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \cdot \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

Dot (inner) product

- Dot product is an operation between two vectors with same dimensions

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

- Calculate the dot products for the following vectors

$$\begin{bmatrix} 4 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \begin{bmatrix} 4 \\ -3 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 4 \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ -4 \\ -6 \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

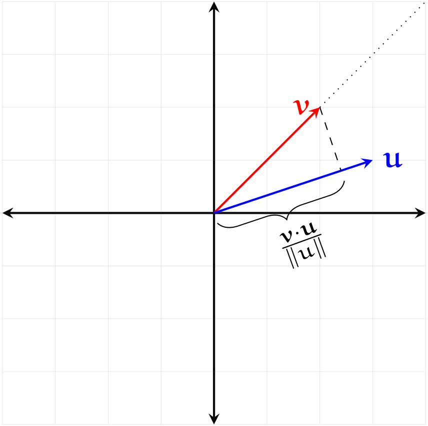
- Note that dot product is larger when the vectors are 'similar'

Properties of dot product

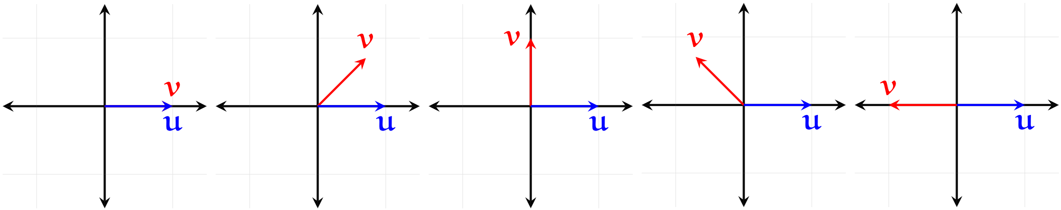
- *Commutativity* $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- *Distributivity with vector addition* $\mathbf{u} \cdot (\mathbf{v} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{u}$
- *Associativity with scalar multiplication* $(a\mathbf{u}) \cdot (b\mathbf{v}) = ab(\mathbf{u} \cdot \mathbf{v})$.
- Note that dot product is not associative, since the result of the dot product is not a vector, but a scalar

Geometric interpretation of the dot product

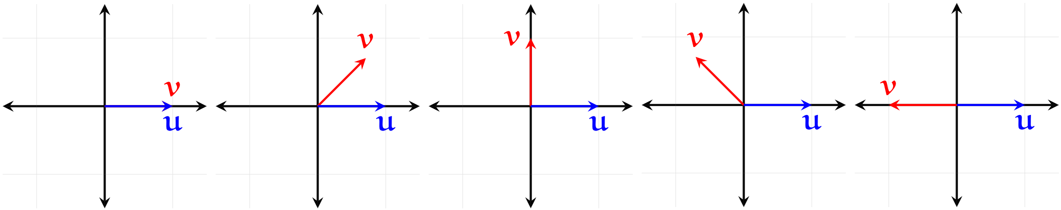
- The dot product of two vectors gives the (orthogonal) projection of one of the vectors to the line defined by the other



Dot product with unit vectors

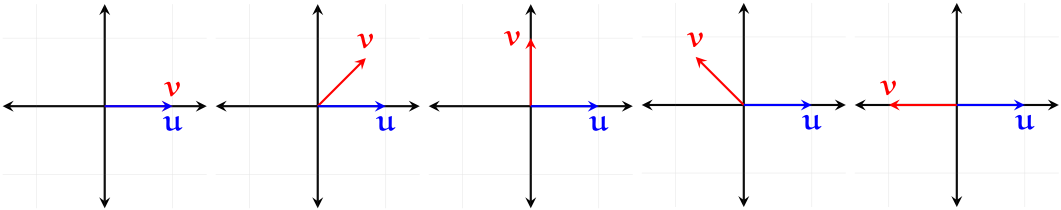


Dot product with unit vectors



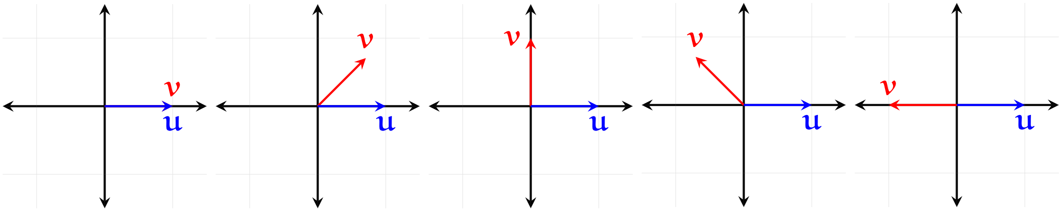
$\mathbf{u} \cdot \mathbf{v} = 1$

Dot product with unit vectors



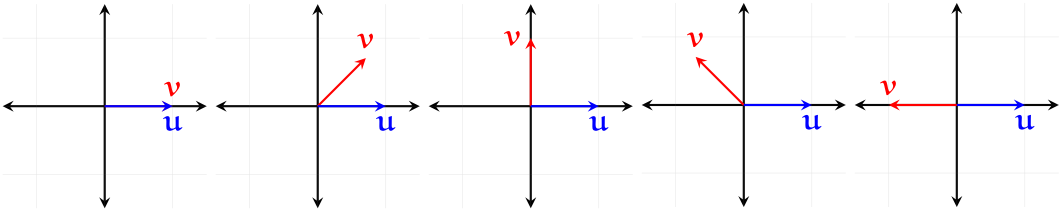
$\mathbf{u} \cdot \mathbf{v} = 1$ $\mathbf{u} \cdot \mathbf{v} = 0.7$

Dot product with unit vectors



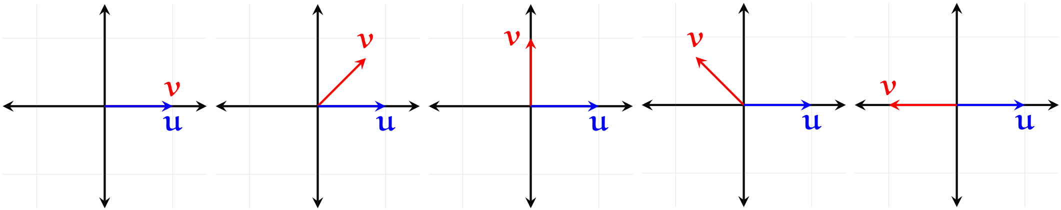
$\mathbf{u} \cdot \mathbf{v} = 1$ $\mathbf{u} \cdot \mathbf{v} = 0.7$ $\mathbf{u} \cdot \mathbf{v} = 0$

Dot product with unit vectors



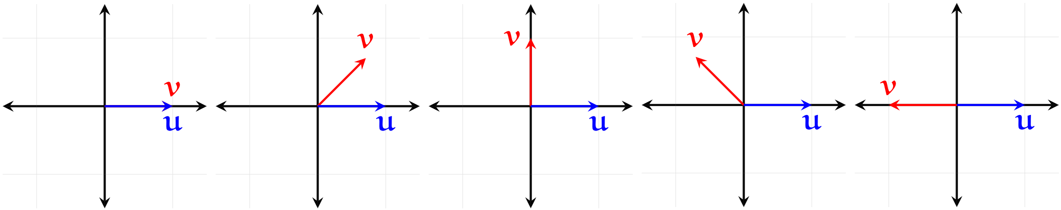
$\mathbf{u} \cdot \mathbf{v} = 1$ $\quad \mathbf{u} \cdot \mathbf{v} = 0.7$ $\quad \mathbf{u} \cdot \mathbf{v} = 0$ $\quad \mathbf{u} \cdot \mathbf{v} = -0.7$

Dot product with unit vectors



$\mathbf{u} \cdot \boldsymbol{v} = 1$ $\quad$ $\mathbf{u} \cdot \boldsymbol{v} = 0.7$ $\quad$ $\mathbf{u} \cdot \boldsymbol{v} = 0$ $\quad$ $\mathbf{u} \cdot \boldsymbol{v} = -0.7$ $\quad$ $\mathbf{u} \cdot \boldsymbol{v} = -1$

Dot product with unit vectors



$\mathbf{u} \cdot \mathbf{v} = 1$ $\mathbf{u} \cdot \mathbf{v} = 0.7$ $\mathbf{u} \cdot \mathbf{v} = 0$ $\mathbf{u} \cdot \mathbf{v} = -0.7$ $\mathbf{u} \cdot \mathbf{v} = -1$

- The dot product is larger if the vectors point to the similar directions
- The dot product is 0 if the vectors are orthogonal

Vector norms

- The *norm* of a vector is an indication of its size (magnitude)
- The norm of a vector is the distance from its tail to its tip
- Norms are related to distance measures
- Vector norms are particularly important for understanding some machine learning techniques

L2 norm

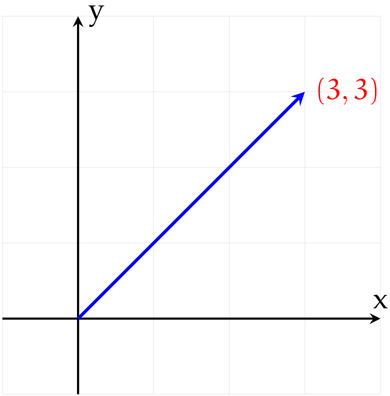
- Euclidean norm, or L2 (or L_2) norm is the most commonly used norm
- For $\boldsymbol{v} = (v_1, v_2, \ldots v_n)$,

$$\|v\|_2 = \sqrt{v_1^2 + v_2^2 + \ldots v_n^2}$$
$$= \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$$

- For example,

$$\|(3,3)\|_2 = \sqrt{3^2 + 3^2} = \sqrt{18}$$

- L2 norm is the default, we often skip the subscript $\|\boldsymbol{v}\|$

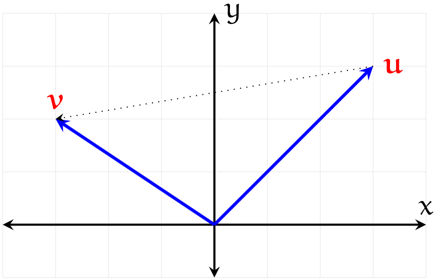


Euclidean distance

- Euclidean distance between two vectors is the L2 norm of their difference

$$D(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(-6)^2 + (-1)^2}$$

- Euclidean distance is a metric
  - symmetric $\|\mathbf{v} - \mathbf{u}\| = \|\mathbf{u} - \mathbf{v}\|$
  - non-negative
  - and obeys the triangle inequality $D(\mathbf{u}, \mathbf{v}) \leqslant D(\mathbf{u}, \mathbf{w}) + D(\mathbf{w}, \mathbf{v})$ for any $\mathbf{w}$



Cauchy–Schwarz inequality

$$|\mathbf{u} \cdot \boldsymbol{v}| \leqslant \|u\|\|v\|$$

- In words: the product of the norms of two vectors is greater than or equal to absolute value of their dot product
- Inequality is important as it links norms and the dot product, it is useful proofs of many intersting relations
- Equality holds only when two vectors are aligned

Chebyshev's inequality

Given a vector $\boldsymbol{v} = (v_1, v_2, \ldots v_n)$ with k entries with the property $|x_i| \geqslant a$,

$$\|v\|^2 \geqslant ka^2$$

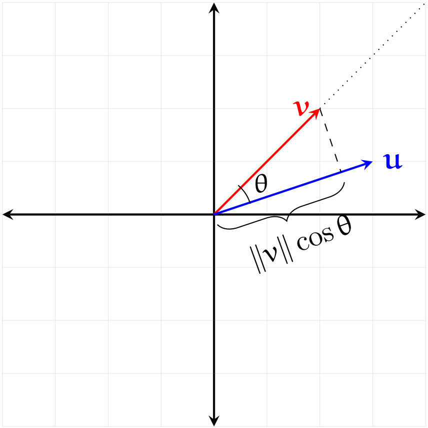
- This puts a limit on the number of large (divergent) values in a vector in terms of its norm
- A trivial consequence: no entry in a vector can be larger than its norm
- The inequality also has important consequences in statistics (we will return to it later)

Cosine similarity

- The cosine of the angle between two vectors

$$\cos\theta = \frac{\boldsymbol{v} \cdot \boldsymbol{u}}{\|\boldsymbol{v}\| \cdot \|\boldsymbol{u}\|}$$

  is called *cosine similarity*
- Unlike dot product, the cosine similarity is not sensitive to the magnitudes of the vectors
- The cosine similarity is bounded in range $[-1, +1]$



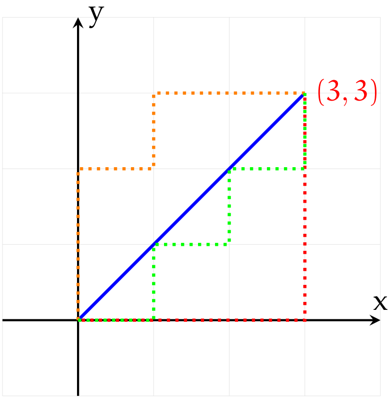
L1 norm

- Another norm we will often encounter is the L1 norm

$$\|v\|_1 = |v_1| + |v_2|$$

$$\|(3,3)\|_1 = |3| + |3| = 6$$

- L1 norm is related to Manhattan distance



L_P norm

In general, L_P norm, is defined as

$$\|\boldsymbol{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}$$

L_P norm

In general, L_P norm, is defined as

$$\|\boldsymbol{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}}$$

We will only work with than L1 and L2 norms, but you may also see L_0 and L_∞ norms in related literature

Matrices

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots & a_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,m} \end{bmatrix}$$

- We can think of matrices as collection of row or column vectors
- A matrix with n rows and m columns is in $\mathbb{R}^{n \times m}$
- Some operations in linear algebra also generalize to more than 2-D objects
- A *tensor* can be thought of a generalization of vectors and matrices to multiple dimensions

Matrices

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \ldots & a_{1,m} \\ a_{2,1} & a_{2,2} & a_{2,3} & \ldots & a_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \ldots & a_{n,m} \end{bmatrix}$$

- We can think of matrices as collection of row or column vectors
- A matrix with n rows and m columns is in $\mathbb{R}^{n \times m}$
- Some operations in linear algebra also generalize to more than 2-D objects
- A *tensor* can be thought of a generalization of vectors and matrices to multiple dimensions

Multiplying a matrix with a scalar

Similar to vectors, each element is multiplied by the scalar.

$$2\begin{bmatrix}2 & 1\\1 & 4\end{bmatrix} = \begin{bmatrix}2\times 2 & 2\times 1\\2\times 1 & 2\times 4\end{bmatrix} = \begin{bmatrix}4 & 2\\2 & 8\end{bmatrix}$$

Matrix addition and subtraction

Each element is added to (or subtracted from) the corresponding element

$$\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$$

Note:

- Matrix addition and subtraction are defined on matrices of the same dimensions

Transpose of a matrix

Transpose of a $n \times m$ matrix is an $m \times n$ matrix whose rows are the columns of the original matrix.
Transpose of a matrix $\mathbf{A}$ is denoted with $\mathbf{A}^T$.

$$\text{If } \mathbf{A} = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}, \mathbf{A}^T = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}.$$

Some special matrices

Identity matrix

- A square matrix in which all the elements of the principal diagonal are one and all other elements are zero is called *identity matrix* ($\mathbf{I}$)

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Multiplying a vector and matrix with the identity matrix has no affect

Some special matrices

Diagonal matrices

- Diagonal matrices are similar to $\mathbf{I}$. All non-diagonal entries are 0s, but non-zero entries can only be in the main diagonal
- Example:

$$\begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Some special matrices

Upper/lower triangular matrices

- Triangular matrices are useful in many applications
- An upper triangular matrix have all 0s below main diagonal. Example:

$$\begin{bmatrix} 1 & 4 & -2 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

- An lower triangular matrix have all 0s above main diagonal. Example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \\ 8 & 7 & 1 \end{bmatrix}$$

Symmetric matrices

- Symmetric matrices arise in many applications, including in ML/NLP (e.g., similarity or distance matrices)
- A symmetric matrix $\mathbf{A}$ satisfies $a_{ij} = a_{ji}$, or $\mathbf{A} = \mathbf{A}^T$
- Example:

$$\begin{bmatrix} 1 & 4 & -2 \\ 4 & 3 & 0 \\ -2 & 0 & 5 \end{bmatrix}$$

- Symmetric matrices have some interesting properties (that we will return later)

Matrix–vector multiplication

- An $n \times m$ matrix can be multiplied with a m-vector to yield a n-vector

Matrix–vector multiplication

- An $n \times m$ matrix can be multiplied with a m-vector to yield a n-vector
- Example

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \times 0 + 1 \times 1 + 0 \times 0 \\ 1 \times 0 + 0 \times 1 + 1 \times 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Matrix–vector multiplication

- An $n \times m$ matrix can be multiplied with a m-vector to yield a n-vector
- Example

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \times 0 + 1 \times 1 + 0 \times 0 \\ 1 \times 0 + 0 \times 1 + 1 \times 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- One view of this operation: each entry in the resulting vector is a dot product (of rows of the matrix and the vector)

Matrix–vector multiplication

- An $n \times m$ matrix can be multiplied with a m-vector to yield a n-vector
- Example

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \times 0 + 1 \times 1 + 0 \times 0 \\ 1 \times 0 + 0 \times 1 + 1 \times 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- One view of this operation: each entry in the resulting vector is a dot product (of rows of the matrix and the vector)
- Another: the result is a linear combination of the columns of the matrix (with the entries in the vector as coefficients)

$$0 \times \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0 \times \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Matrix multiplication as linear transformation

- Multiplying a vector with a matrix transforms the vector
- The result is another vector (possibly in a different vector space)
- Many operations on vectors can be expressed with multiplying with a matrix (linear transformations)

Transformation examples

identity

- Identity transformation maps a vector to itself
- For example:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Transformation examples

permutation

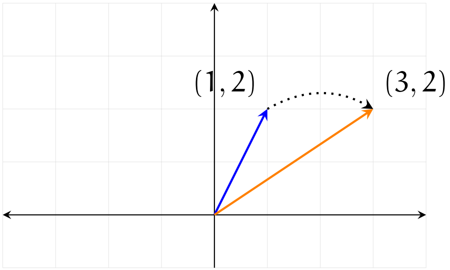
- Another simple transformation is to permute (re-arrange) the elements (rows) of the vector
- For example:

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_4 \\ x_2 \\ x_3 \\ x_1 \end{bmatrix}$$

Matrices transform vectors

stretch along the x axis

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



Transformation examples

rotation

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

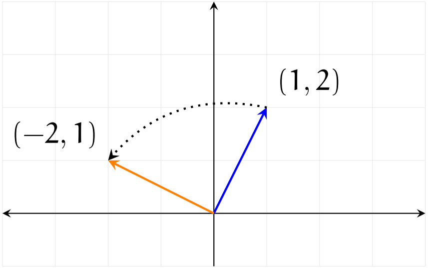
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

Transformation examples

rotation

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$



Transformations by rectangular matrices

- Multiplying a vector with (compatible) rectangular matrix results in a vector with different dimensionality
- Example $\mathbb{R}^3 \to \mathbb{R}^2$

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- Example $\mathbb{R}^3 \to \mathbb{R}^4$

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}$$

Dot product as matrix multiplication

In machine learning (and many other disciplines), we treat an n-vector as an $n \times 1$ matrix.

Then, the *dot product* of two vectors is

$$\mathbf{u}^\mathsf{T}\boldsymbol{v}$$

For example, $\mathbf{u} = (2, 2)$ and $\boldsymbol{v} = (2, -2)$,

$$\begin{bmatrix} 2 & 2 \end{bmatrix} \times \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

Dot product as matrix multiplication

In machine learning (and many other disciplines), we treat an n-vector as an $n \times 1$ matrix.

Then, the *dot product* of two vectors is

$$\mathbf{u}^{\mathsf{T}}\boldsymbol{v}$$

For example, $\mathbf{u} = (2, 2)$ and $\boldsymbol{v} = (2, -2)$,

$$\begin{bmatrix} 2 & 2 \end{bmatrix} \times \begin{bmatrix} 2 \\ -2 \end{bmatrix} = 2 \times 2 + 2 \times -2 = 4 - 4 = 0$$

- This is a 1×1 matrix, but matrices and vectors with single entries are often treated as scalars

Question: What is the transformation performed by dot product?

Outer product

The *outer product* of two column vectors is defined as

$$\boldsymbol{v}\boldsymbol{u}^{\mathsf{T}}$$

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

Outer product

The *outer product* of two column vectors is defined as

$$\boldsymbol{v}\boldsymbol{u}^{\mathsf{T}}$$

$$\begin{bmatrix}1\\2\end{bmatrix} \times \begin{bmatrix}1 & 2 & 3\end{bmatrix} = \begin{bmatrix}1 & 2 & 3\\2 & 4 & 6\end{bmatrix}$$

Note:

- The result is a matrix
- The vectors do not have to be the same length

Matrix multiplication

- if $\mathbf{A}$ is a $n \times k$ matrix, and $\mathbf{B}$ is a $k \times m$ matrix, their product $\mathbf{C}$ is a $n \times m$ matrix
- Elements of C, $c_{i,j}$, are defined as

$$c_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}$$

- Note: $c_{i,j}$ is the dot product of the i^{th} row of $\mathbf{A}$ and the j^{th} column of $\mathbf{B}$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + \dots a_{1k}b_{k1}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication
(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1m} \\ b_{21} & b_{22} & \ldots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \ldots & b_{km} \end{pmatrix}$$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} + \ldots a_{1k}b_{k2}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \ldots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{1m} = a_{11}b_{1m} + a_{12}b_{2m} + \dots a_{1k}b_{km}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{21} = a_{21}b_{11} + a_{22}b_{21} + \dots a_{2k}b_{k1}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{22} = a_{21}b_{12} + a_{22}b_{22} + \dots a_{2k}b_{k2}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{2m} = a_{21}b_{1m} + a_{22}b_{2m} + \dots a_{2k}b_{km}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{n1} = a_{n1}b_{11} + a_{n2}b_{22} + \dots a_{nk}b_{k1}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{n2} = a_{n1}b_{12} + a_{n2}b_{22} + \dots a_{nk}b_{k2}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{nm} = a_{n1}b_{1m} + a_{n2}b_{2m} + \dots a_{nk}b_{km}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication

(demonstration)

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{km} \end{pmatrix}$$

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots a_{ik}b_{kj}$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}$$

Matrix multiplication example

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} =$$

Matrix multiplication example

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 1 & 2 \end{bmatrix}$$

Complexity of matrix multiplication

- How many scalar multiplications do we perform when multiplying an $n \times m$ matrix with a $m \times k$ matrix?

Complexity of matrix multiplication

- How many scalar multiplications do we perform when multiplying an $n \times m$ matrix with a $m \times k$ matrix?
  - We do $n \times k$ dot products
  - Each dot product multiplies m-dimensional vectors
  - we have $n \times m \times k$ multiplications
- In general, the complexity of matrix multiplication is $O(n^3)$
- Faster algorithms exist ($O(n^{2.81})$ or even $O(n^{2.37})$)

Properties of matrix multiplication

- Associativity

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

- Distributivity

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

- Multiplication by identity matrix

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}$$

- Matrix multiplication is not commutative $\mathbf{AB} \neq \mathbf{BA}$ (in general)
- Matrix multiplication and transpose

$$(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$$

Question

We have three matrices:

- $\mathbf{A}$: a 10×2 matrix
- $\mathbf{B}$: a 2×5 matrix
- $\mathbf{C}$: a 5×10 matrix

- What is the dimensionality of $\mathbf{ABC}$

Question

We have three matrices:

- $\mathbf{A}$: a 10×2 matrix
- $\mathbf{B}$: a 2×5 matrix
- $\mathbf{C}$: a 5×10 matrix

- What is the dimensionality of $\mathbf{ABC}$
- Does it matter if we perform the multiplication as

Question

We have three matrices:

- $\mathbf{A}$: a 10×2 matrix
- $\mathbf{B}$: a 2×5 matrix
- $\mathbf{C}$: a 5×10 matrix

- What is the dimensionality of $\mathbf{ABC}$
- Does it matter if we perform the multiplication as
  - $(\mathbf{AB})\mathbf{C}$, or

Question

We have three matrices:

- $\mathbf{A}$: a 10×2 matrix
- $\mathbf{B}$: a 2×5 matrix
- $\mathbf{C}$: a 5×10 matrix

- What is the dimensionality of $\mathbf{ABC}$
- Does it matter if we perform the multiplication as
  - $(\mathbf{AB})\mathbf{C}$, or
  - $\mathbf{A}(\mathbf{BC})$

Alternative ways to think about matrix multiplication

If we have $\mathbf{AB} = \mathbf{C}$,

- Column vectors of $\mathbf{C}$, $\mathbf{c}_j = \mathbf{A}\mathbf{b}_j$
- Row vectors of $\mathbf{C}$, $\mathbf{c}_i^T = \mathbf{a}_i^T\mathbf{B}$
- $\mathbf{C}$ is also the sum of outer product of columns of $\mathbf{A}$ and rows of $\mathbf{B}$

$$C = \sum \mathbf{a}_i \mathbf{b}_i^T$$

Matrix-vector representation of a set of linear equations

The set of linear equations

$$\begin{aligned} 2x_1 &+ x_2 &= 6 \\ x_1 &+ 4x_2 &= 17 \end{aligned}$$

can be written as:

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}}_{\boldsymbol{W}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} 6 \\ 17 \end{bmatrix}}_{\boldsymbol{b}}$$

One can solve the above equation using row reduction, or *Gaussian elimination* (we will not cover it today).

Inverse of a matrix

(a first introduction)

- Inverse of a square matrix $\boldsymbol{W}$, when it exists, is denoted $\boldsymbol{W}^{-1}$, and defined as

$$\boldsymbol{W}\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}\boldsymbol{W} = \mathbf{I}$$

- Can you calculate the inverse of the following matrix?

$$\begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$$

Inverse of a matrix

(a first introduction)

- Inverse of a square matrix $\boldsymbol{W}$, when it exists, is denoted $\boldsymbol{W}^{-1}$, and defined as

$$\boldsymbol{W}\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}\boldsymbol{W} = \mathbf{I}$$

- Can you calculate the inverse of the following matrix?

$$\begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Inverse of a matrix

(a first introduction)

- Inverse of a square matrix $\boldsymbol{W}$, when it exists, is denoted $\boldsymbol{W}^{-1}$, and defined as

$$\boldsymbol{W}\boldsymbol{W}^{-1} = \boldsymbol{W}^{-1}\boldsymbol{W} = \mathbf{I}$$

- Can you calculate the inverse of the following matrix?

$$\begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Once we invert the matrix, the solution to $W\mathbf{x} = \mathbf{b}$ is simply $\mathbf{x} = W^{-1}\mathbf{b}$
- We'll learn later how to calculate the inverse systematically

Summary & next week

- Vectors, matrices
- Operations between vectors and matrices
- Dot product

Next:

- Solving systems of linear equations

Further reading

- A classic reference book in the field is Strang (2009)
- Shifrin and Adams (2011) and Farin and Hansford (2014) are textbooks with a more practical/graphical orientation.
- Cherney, Denton, and Waldron (2013) and Beezer (2014) are two textbooks that are freely available.
- For more alternatives, see
  `http://www.openculture.com/free-math-textbooks`
- You may also find the MIT video lectures on introductory linear algebra at
  `https://www.youtube.com/playlist?list=PL49CF3715CB9EF31D`

Further reading (cont.)

- Beezer, Robert A. (2014). *A First Course in Linear Algebra*. version 3.40. Congruent Press. ISBN: 9780984417551. URL: `http://linear.ups.edu/`.
- Cherney, David, Tom Denton, and Andrew Waldron (2013). *Linear algebra*. math.ucdavis.edu. URL: `https://www.math.ucdavis.edu/~linear/`.
- Farin, Gerald E. and Dianne Hansford (2014). *Practical linear algebra: a geometry toolbox*. Third edition. CRC Press. ISBN: 978-1-4665-7958-3.
- Shifrin, Theodore and Malcolm R Adams (2011). *Linear Algebra. A Geometric Approach*. 2nd. W. H. Freeman. ISBN: 978-1-4292-1521-3.
- Strang, Gilbert (2009). *Introduction to Linear Algebra, Fourth Edition*. 4th ed. Wellesley Cambridge Press. ISBN: 9780980232714.