Applied Deep Learning

Dr. Philippe Blaettchen
Bayes Business School (formerly Cass)

www.bayes.city.ac.uk

Learning objectives of today

Goals: Understand the key concepts of linear algebra relevant to deep learning

- Basic definitions
- Typical operations on vectors and matrices

How will we do this?

- Not a comprehensive review, but focusing on the most relevant concepts for understanding deep learning
- Small repetition in class, as well as an introduction on how to implement concepts in Python

Linear algebra – definitions

Scalars

- Scalar: a single number
- Integers (-1,0,1,2,…), real numbers (0.319375, 1.17, π), rational numbers $\left(\frac{integer}{integer}\right)$

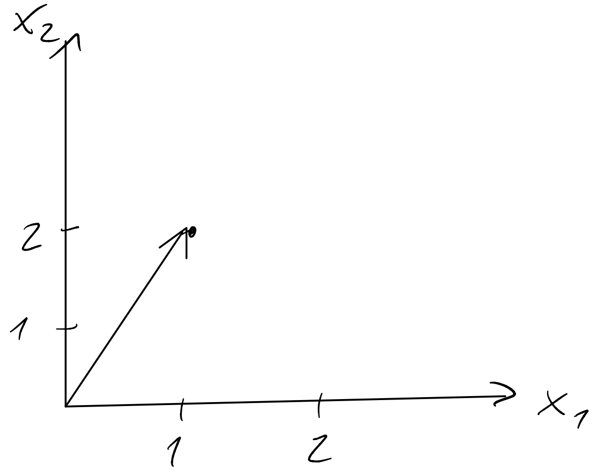
$\alpha, t \qquad \alpha \in \mathbb{R}, \ t \in \mathbb{Z}$

Vectors

- A one-dimensional array of numbers:

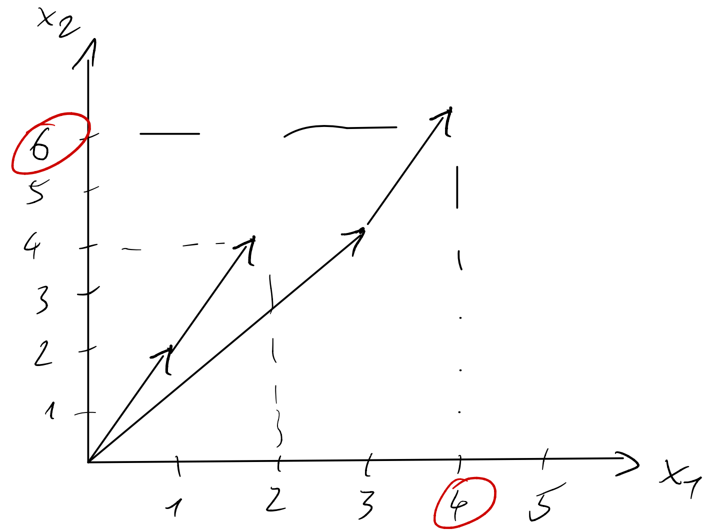
$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$



- Entries can be real numbers, binary, integer, …

  → we usually denote a vector with the type of its entries and its size, e.g., $\mathbf{x} \in \mathbb{R}^n$



- A one-dimensional array of numbers:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Entries can be real numbers, binary, integer, …

  → we usually denote a vector with the type of its entries and its size, e.g., $\mathbf{x} \in \mathbb{R}^n$

- We can perform elementary algebra on vectors:

$x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $y = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ $x + y = \begin{pmatrix} 1+3 \\ 2+4 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$

  - $\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$ (check dimensions!)

  - $\alpha \boldsymbol{x} = \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$

$\alpha = 2$, $x = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $\alpha x = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$

Matrices

- A two-dimensional array of numbers

Row

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}.$$

Column

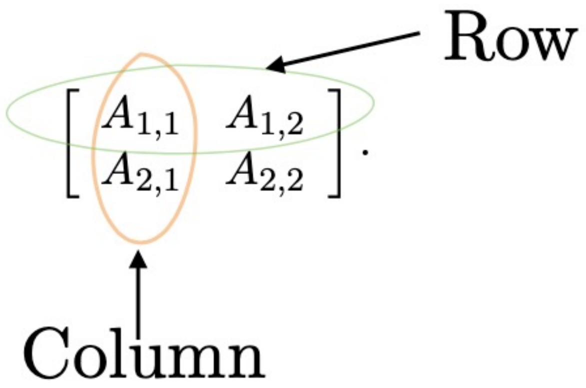
$n = 2$
$m = 2$

- We again denote it with its type and shape: $A \in \mathbb{R}^{n \times m}$, where n is the number of rows and m is the number of columns

Source: Goodfellow

Matrices

- A two-dimensional array of numbers



- We again denote it with its type and shape: $A \in \mathbb{R}^{n \times m}$, where n is the number of rows and m is the number of columns

$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \; B = \begin{pmatrix} 3 & 0 \\ 1 & 5 \end{pmatrix}$

- We can perform elementary algebra on ~~vectors~~ matrices:
  - $\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} = \begin{bmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} \end{bmatrix}$ (check dimensions!)

$A + B = \begin{pmatrix} 1+3 & 2+0 \\ 3+1 & 4+5 \end{pmatrix} = \begin{pmatrix} 4 & 2 \\ 4 & 9 \end{pmatrix}$

$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 1 \\ 1 & 5 & 1 \end{pmatrix}$ ↯

  - $\alpha \boldsymbol{A} = \alpha \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = \begin{bmatrix} \alpha a_{1,1} & \alpha a_{1,2} \\ \alpha a_{2,1} & \alpha a_{2,2} \end{bmatrix}$

Source: Goodfellow

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

A typical use of matrices in deep learning

$$X = \begin{pmatrix} 0 & \cdots & 0 \\ & & \\ & & 0 \end{pmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \boxed{a_{nm}} \end{bmatrix} \longleftrightarrow$$

Pixel's intensity value

Source: Ivanovic

Tensors

- Any array of numbers
- Could have zero dimensions (scalar), one dimension (vector), two dimensions (matrix)
- Could also have three or more dimensions

In Python

- We generally use Python to work with vectors, matrices, and sometimes tensors
- Later, we'll also see the TensorFlow-specific implementation of tensors

Linear algebra – typical operations

Trace of a matrix

- Summing up all the entries on the diagonal of the matrix:

$$A = \begin{pmatrix} 1 & 2 & 5 \\ 4 & 3 & 0 \\ 2 & 1 & 0 \end{pmatrix}$$

$$Tr(A) = 1 + 3 + 0 = 4$$

$$Tr(\boldsymbol{A}) = \sum_{i=1}^{n} \boldsymbol{A}_{i,i}$$

- Note that
  - $Tr(\boldsymbol{ABC}) = Tr(\boldsymbol{CAB}) = Tr(\boldsymbol{BCA})$
  - $Tr(\boldsymbol{A} + \boldsymbol{B}) = Tr(\boldsymbol{A}) + Tr(\boldsymbol{B})$

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \quad A + B = \begin{pmatrix} 2 & 2 \\ 5 & 5 \end{pmatrix}$$

$$Tr(A) = 5 \quad Tr(B) = 2 \quad Tr(A+B) = 7$$

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Matrix transpose

- Essentially a mirror image across the main diagonal

A, T

A'

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow A^{\top} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

$$(A^{\top})_{i,j} = A_{j,i}.$$

- Note that:
  - $(A^T)^T = A$
  - $(AB)^T = B^T A^T$
  - $(A + B)^T = A^T + B^T$

$A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \quad A^T = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$

- We call a matrix symmetric if $A = A^T$

Source: Goodfellow

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Inner product of vectors (also known as dot product)

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Say $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, then:

$$\boldsymbol{x}^T \boldsymbol{y} = [x_1\ x_2\ \dots\ x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

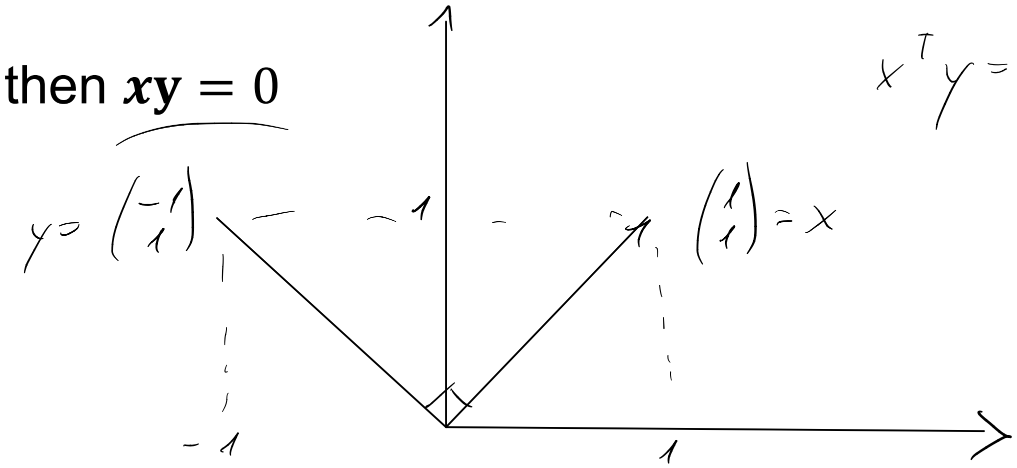
$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad y = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

$$x^T y = (1\ 2\ 3) \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = 2 + 0 + 3 = 5$$

- The inner product is a scalar!

$$x^T y = (1\ 1) \begin{pmatrix} -1 \\ 1 \end{pmatrix} = -1 + 1 = 0$$

- Note: if $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal, then $\boldsymbol{x}\boldsymbol{y} = 0$



Outer product of vectors

- Say $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$ then:

$$\boldsymbol{x}\,\boldsymbol{y}^T = \underset{n}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}} \overset{m}{\begin{bmatrix} y_1 & y_2 & \dots & y_m \end{bmatrix}} = \begin{array}{cc} \overset{m}{\longrightarrow} & \\ \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_m \\ x_2y_1 & x_2y_2 & \dots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_my_2 & \dots & x_ny_m \end{bmatrix} & \Big\downarrow n \end{array}$$

- The outer product is a matrix!

Matrix-vector product

$A_{(1),(:)}$ $a_{1,:}$

- Say $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ then:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{1,1}y_1 + a_{1,2}y_2 + \cdots a_{1,m}y_m \\ a_{2,1}y_1 + a_{2,2}y_2 + \cdots a_{2,m}y_m \\ \vdots \\ a_{n,1}y_1 + a_{n,2}y_2 + \cdots a_{n,m}y_m \end{bmatrix}$$

$$= \begin{bmatrix} A_{1,:}\,\boldsymbol{y} \\ A_{2,:}\,\boldsymbol{y} \\ \vdots \\ A_{n,:}\,\boldsymbol{y} \end{bmatrix}$$

$$= A_{:,1}y_1 + A_{:,2}y_2 + \cdots + A_{:,m}y_m$$

$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$Ay = \begin{pmatrix} 5 \\ 11 \end{pmatrix}$

Matrix multiplication

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}.$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}.$$

$$\begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{pmatrix} = B$$

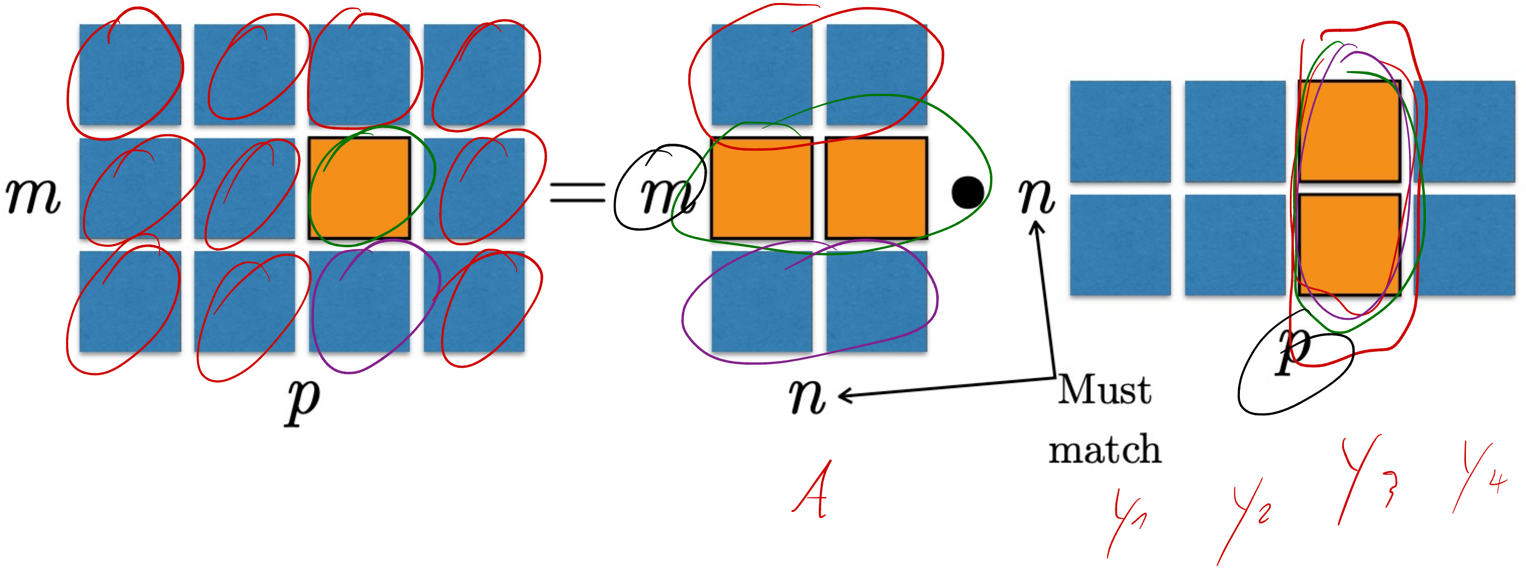
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \qquad \begin{pmatrix} 5 & 11 & \dots \\ 11 & 25 & \dots \\ \dots & & \end{pmatrix} = C$$

rows from A

columns from B

$C \in \mathbb{R}^{m \times p}$

$\mathbb{R}^{3 \times 4}$



Source: Goodfellow

A few important properties of matrix multiplication

- Associative: $(\mathbf{AB})\mathbf{C} = \boldsymbol{A}(\boldsymbol{BC})$
- Distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \boldsymbol{AB} + \boldsymbol{AC}$
- Generally, not commutative: $\boldsymbol{AB} \neq \boldsymbol{BA}$
  (also, just because $\boldsymbol{AB}$ exists, it doesn't mean that $\boldsymbol{BA}$ exists)

$A \in \mathbb{R}^{n \times m}$

$B \in \mathbb{R}^{m \times p}$

$n \neq p$

$\Rightarrow AB \in \mathbb{R}^{n \times p}$

$\cancel{BA \in \mathbb{R}^{m}}$

See you in class!

Sources

- Bhaskhar, 2021, Linear Algebra: https://cs229.stanford.edu/notes2021fall/section1notes-linear-algebra-review.pdf
- Collins, 2012, Intensity Surfaces and Gradients: http://www.cse.psu.edu/~rtc12/CSE486/lecture02_6pp.pdf
- Goodfellow, Bengio, Courville, 2016, The Deep Learning Book: http://www.deeplearningbook.org
- Kolter, Do, & Ma, 2020, Linear Algebra Review and Reference: https://cs229.stanford.edu/summer2020/cs229-linalg.pdf
- Ivanovic, 2017, Python Introduction and Linear Algebra Review: https://web.stanford.edu/class/cs231a/section/section1.pdf

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON