Linear Algebra

Lecture slides for Chapter 2 of *Deep Learning*

Ian Goodfellow

2016-06-24

About this chapter

- Not a comprehensive survey of all of linear algebra
- Focused on the subset most relevant to deep learning
- Larger subset: e.g., ***Linear Algebra*** **by Georgi Shilov**

Scalars

- A scalar is a single number
- Integers, real numbers, rational numbers, etc.
- We denote it with italic font:

$$a, n, x$$

(Goodfellow 2016)

Vectors

- A vector is a 1-D array of numbers:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \quad (2.1)$$

- Can be real, binary, integer, etc.

- Example notation for type and size:

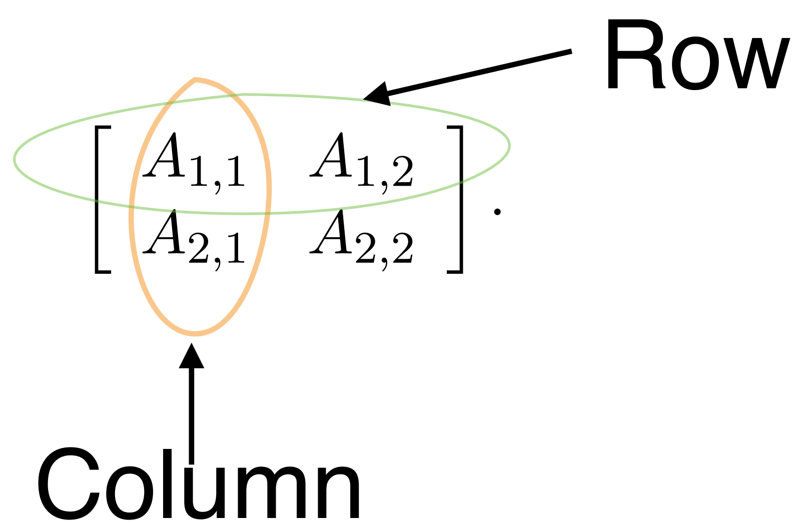
$$\mathbb{R}^n$$

(Goodfellow 2016)

Matrices

- A matrix is a 2-D array of numbers:

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}. \quad (2.2)$$



- Example notation for type and shape:

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

(Goodfellow 2016)

Tensors

- A tensor is an array of numbers, that may have
  - zero dimensions, and be a scalar
  - one dimension, and be a vector
  - two dimensions, and be a matrix
  - or more dimensions.

(Goodfellow 2016)

Matrix Transpose

$$(\boldsymbol{A}^\top)_{i,j} = A_{j,i}. \tag{2.3}$$

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

Figure 2.1: The transpose of the matrix can be thought of as a mirror image across the main diagonal.

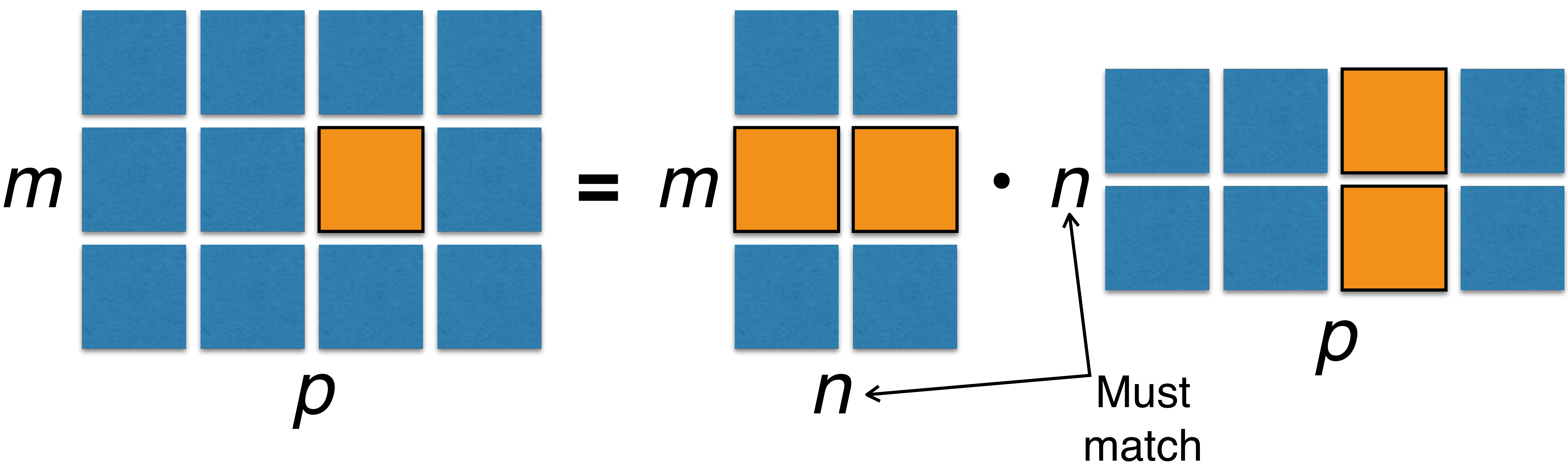
$$(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top. \tag{2.9}$$

(Goodfellow 2016)

Matrix (Dot) Product

$$C = AB. \tag{2.4}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}. \tag{2.5}$$



(Goodfellow 2016)

Identity Matrix

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

Figure 2.2: *Example identity matrix*: This is $\boldsymbol{I}_3$.

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}. \tag{2.20}$$

(Goodfellow 2016)

Systems of Equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.11}$$

expands to

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1 \tag{2.12}$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2 \tag{2.13}$$

$$\dots \tag{2.14}$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m \tag{2.15}$$

(Goodfellow 2016)

Solving Systems of Equations

- A linear system of equations can have:
  - No solution
  - Many solutions
  - Exactly one solution: this means multiplication by the matrix is an invertible function

(Goodfellow 2016)

Matrix Inversion

- Matrix inverse:

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n. \tag{2.21}$$

- Solving a system using an inverse:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.22}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{2.23}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{2.24}$$

- Numerically unstable, but useful for abstract analysis

(Goodfellow 2016)

Invertibility

- Matrix can’t be inverted if…
  - More rows than columns
  - More columns than rows
  - Redundant rows/columns (“linearly dependent”, “low rank”)

(Goodfellow 2016)

Norms

- Functions that measure how "large" a vector is
- Similar to a distance between zero and the point represented by the vector
  - $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
  - $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$ (the *triangle inequality*)
  - $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

(Goodfellow 2016)

Norms

- L^p norm

$$||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

- Most popular norm: L2 norm, *p*=2
- L1 norm, *p*=1: $$||\boldsymbol{x}||_1 = \sum_i |x_i|. \quad (2.31)$$
- Max norm, infinite *p:* $$||\boldsymbol{x}||_\infty = \max_i |x_i|. \quad (2.32)$$

(Goodfellow 2016)

Special Matrices and Vectors

- Unit vector:

$$||\boldsymbol{x}||_2 = 1. \quad (2.36)$$

- Symmetric Matrix:

$$\boldsymbol{A} = \boldsymbol{A}^\top. \quad (2.35)$$

- Orthogonal matrix:

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}. \quad (2.37)$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$

(Goodfellow 2016)

Trace

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i \boldsymbol{A}_{i,i}. \tag{2.48}$$

$$\mathrm{Tr}(\boldsymbol{ABC}) = \mathrm{Tr}(\boldsymbol{CAB}) = \mathrm{Tr}(\boldsymbol{BCA}) \tag{2.51}$$

(Goodfellow 2016)

Learning linear algebra

- Do a lot of practice problems
- Start out with lots of summation signs and indexing into individual entries
- Eventually you will be able to mostly use matrix and vector product notation quickly and easily

Linear Algebra - Part II

Projection, Eigendecomposition, SVD

Punit Shah

(Adapted from Sargur Srihari's slides)

Brief Review from Part 1

- Symmetric Matrix:

$$\mathbf{A} = \mathbf{A}^T$$

- Orthogonal Matrix:

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I} \quad \text{and} \quad \mathbf{A}^{-1} = \mathbf{A}^T$$

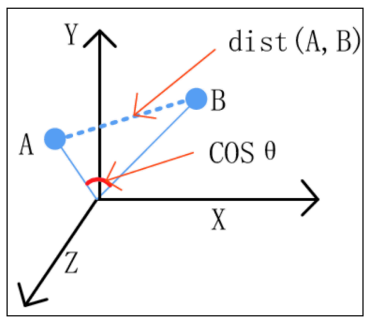
- L2 Norm:

$$||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$$

Angle Between Vectors

- Dot product of two vectors can be written in terms of their L2 norms and the angle θ between them.

$$\mathbf{a}^T\mathbf{b} \;=\; ||\mathbf{a}||_2||\mathbf{b}||_2\cos(\theta)$$



Cosine Similarity

- Cosine between two vectors is a measure of their similarity:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{||\mathbf{a}|| \; ||\mathbf{b}||}$$

- **Orthogonal Vectors**: Two vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal to each other if $\mathbf{a} \cdot \mathbf{b} = 0$.

Vector Projection

- Given two vectors $\mathbf{a}$ and $\mathbf{b}$, let $\hat{\mathbf{b}} = \frac{\mathbf{b}}{||\mathbf{b}||}$ be the unit vector in the direction of $\mathbf{b}$.
- Then $\mathbf{a}_1 = a_1\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{a}$ onto a straight line parallel to $\mathbf{b}$, where

$$a_1 = ||\mathbf{a}|| \cos(\theta) = \mathbf{a} \cdot \hat{\mathbf{b}} = \mathbf{a} \cdot \frac{\mathbf{b}}{||\mathbf{b}||}$$

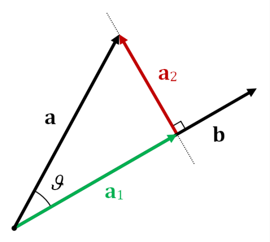


Image taken from wikipedia.

Diagonal Matrix

- Diagonal matrix has mostly zeros with non-zero entries only in the diagonal, e.g. identity matrix.
- A square diagonal matrix with diagonal elements given by entries of vector $\mathbf{v}$ is denoted:

$$\text{diag}(\mathbf{v})$$

- Multiplying vector $\mathbf{x}$ by a diagonal matrix is efficient:

$$\text{diag}(\mathbf{v})\mathbf{x} \;=\; \mathbf{v} \odot \mathbf{x}$$

- Inverting a square diagonal matrix is efficient:

$$\text{diag}(\mathbf{v})^{-1} \;=\; \text{diag}\Big([\frac{1}{v_1}, \dots, \frac{1}{v_n}]^T\Big)$$

Zero Determinant

If $\det(\mathbf{A}) = 0$, then:

- $\mathbf{A}$ is linearly dependent.
- $\mathbf{Ax} = \mathbf{b}$ has no solution or infinitely many solutions.
- $\mathbf{Ax} = \mathbf{0}$ has a non-zero solution.

Matrix Decomposition

- We can decompose an integer into its prime factors, e.g. $12 = 2 \times 2 \times 3$.
- Similarly, matrices can be decomposed into factors to learn universal properties:

$$\mathbf{A} = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$$

Eigenvectors

- An eigenvector of a square matrix $\mathbf{A}$ is a nonzero vector $\mathbf{v}$ such that multiplication by $\mathbf{A}$ only changes the scale of $\mathbf{v}$.

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

- The scalar λ is known as the **eigenvalue**.

- If $\mathbf{v}$ is an eigenvector of $\mathbf{A}$, so is any rescaled vector $s\mathbf{v}$. Moreover, $s\mathbf{v}$ still has the same eigenvalue. Thus, we constrain the eigenvector to be of unit length:

$$||\mathbf{v}|| = 1$$