UNIVERSITY
OF OSLO

IN4310 Linear Algebra Review

ali@uio.no

A Bit of Background

Research Agenda: Learning under Resource Constraints in Real World

Associate Professor, Department of Informatics, University of Oslo
Principal Investigator, SFI Visual Intelligence
Norwegian Centre for Knowledge-driven Machine Learning (Integreat)
European Laboratory for Learning and Intelligent Systems (ELLIS) Member
`https://alirk.github.io/`

UNIVERSITY OF OSLO

Norwegian Centre for Knowledge-driven Machine Learning

A Bit of Background

What Is This Course About?

Algorithms, Practice, Theory, Major Issues of Deep Learning

Main Application: Image Data

Not a Pure Programming and Math/Stats Course

Useful Tools for Industry/Academic Career

Matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ a & b & c \end{bmatrix}$$

Notation

$A \in \mathbb{R}^{m \times n}$ matrix with m rows and n columns with real entries

$\mathbf{x} \in \mathbb{R}^n$ n-dimensional column vector

$\mathbf{x}^\top$ the transpose of $\mathbf{x}$ (row vector)

$\mathbf{a}_j \in \mathbb{R}^m$ or $A_{:,j} \in \mathbb{R}^m$ j-th column of A

$$A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$$

$$A = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix}$$

Vector-Vector Products and Matrix-Vector Products

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$

Inner product or dot product $\quad \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i$

Outer product $\quad \mathbf{x}\mathbf{y}^\top = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{bmatrix}$

Writing A by rows $\mathbf{y} = A\mathbf{x} = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b}_1^\top \mathbf{x} \\ \vdots \\ \mathbf{b}_m^\top \mathbf{x} \end{bmatrix}$

Writing A by columns $\mathbf{y} = A\mathbf{x} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} \mathbf{a}_i x_i$

Transpose, Symmetric Matrices, and Trace [1]

Transpose of $A \in \mathbb{R}^{m\times n}$ denoted by $A^\top \in \mathbb{R}^{n\times m}$ $\left(A^\top\right)_{ij} = A_{ji}$

$\left(A^\top\right)^\top = A; \quad (AB)^\top = B^\top A^\top; \quad (A+B)^\top = A^\top + B^\top$

Square matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^\top$

Trace of a square matrix $\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}$

$\mathrm{tr}(A) = \mathrm{tr}(A^\top); \quad \mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B); \quad \mathrm{tr}(AB) = \mathrm{tr}(BA)$

$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$

Norms

Informally a measure of the length of a vector

Euclidean or ℓ_2 norm $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$

$\|\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{x}$

ℓ_1 norm $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$

ℓ_∞ norm $\|\mathbf{x}\|_\infty = \max_i |x_i|$

ℓ_p norm for some $p \geq 1$ $\quad \|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$

Frobenius norm $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\mathrm{tr}(A^\top A)}$

Quadratic Forms and Positive Semidefinite Matrices

Let $A \in \mathbb{R}^{n \times n}$ and $\mathbf{x} \in \mathbb{R}^n$. The scalar $\mathbf{x}^\top A\mathbf{x}$ is quadratic form

$$\mathbf{x}^\top A\mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j$$

$$\mathbf{x}^\top A\mathbf{x} = (\mathbf{x}^\top A\mathbf{x})^\top = \mathbf{x}^\top A^\top \mathbf{x} = \mathbf{x}^\top \left(\tfrac{1}{2}A + \tfrac{1}{2}A^\top\right)\mathbf{x}$$

A symmetric A is positive definite if for all non-zero $\mathbf{x}$, $\mathbf{x}^\top A\mathbf{x} > 0$
A is positive semidefinite if for all non-zero $\mathbf{x}$, $\mathbf{x}^\top A\mathbf{x} \geq 0$ $(A \succeq 0)$
A is negative definite if for all non-zero $\mathbf{x}$, $\mathbf{x}^\top A\mathbf{x} < 0$
A is negative semidefinite if for all non-zero $\mathbf{x}$, $\mathbf{x}^\top A\mathbf{x} \leq 0$ $(A \preceq 0)$
A is indefinite if it is neither PSD nor NSD

Given $A \in \mathbb{R}^{n\times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of A with corresponding non-zero eigenvector $\mathbf{x} \in \mathbb{C}^n$ if $A\mathbf{x} = \lambda\mathbf{x}$

$(\lambda, \mathbf{x})$ is an eigenvalue-eigenvector pair if $(\lambda\mathbf{I} - A)\mathbf{x} = 0,\ \mathbf{x} \neq 0$

$(\lambda\mathbf{I} - A)\mathbf{x} = 0$ has non-zero solution iff $\lambda\mathbf{I} - A$ is singular

$$\det(\lambda\mathbf{I} - A) = 0$$

Trace equals sum of eigenvalues $\mathrm{tr}(A) = \sum_{i=1}^n \mathrm{eig}(A) = \sum_{i=1}^n \lambda_i$

Determinant equals product of eigenvalues $\det(A) = \prod_{i=1}^n \lambda_i$

Matrix Calculus: Gradient

$g : \mathbb{R}^{m \times n} \to \mathbb{R}$ takes an $m \times n$ matrix input and returns a real value

The gradient of g w.r.t. A is a matrix of partial derivatives

$$\nabla_A g(A) = \begin{bmatrix} \frac{\partial g(A)}{\partial A_{11}} & \frac{\partial g(A)}{\partial A_{12}} & \cdots & \frac{\partial g(A)}{\partial A_{1n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g(A)}{\partial A_{m1}} & \frac{\partial g(A)}{\partial A_{m2}} & \cdots & \frac{\partial g(A)}{\partial A_{mn}} \end{bmatrix}$$

$f : \mathbb{R}^n \to \mathbb{R}$ takes an n-dimensional vector input

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Matrix Calculus: Hessian

$f : \mathbb{R}^n \to \mathbb{R}$ takes a vector input and returns a real scalar

The Hessian matrix w.r.t. $\mathbf{x}$ is $n \times n$ matrix of partial derivatives

$$\nabla_{\mathbf{x}}^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}$$

$$\left(\nabla_{\mathbf{x}}^2 f(\mathbf{x})\right)_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

Exercise: Suppose $f(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x}$. Then show

$$\nabla_{\mathbf{x}} \mathbf{x}^\top A \mathbf{x} = 2A\mathbf{x}, \quad \nabla_{\mathbf{x}}^2 \mathbf{x}^\top A \mathbf{x} = 2A$$

Hint: Show $\frac{\partial f(\mathbf{x})}{\partial x_i} = 2\sum_{j=1}^n A_{ij} x_j$ and $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = 2A_{ij}$

References

[1] Z. Kolter and C. Do. Linear algebra review. https://cs229.stanford.edu/section/cs229-linalg.pdf.