Machine Learning, STOR 565

Overview of Matrix and Linear Algebra

Andrew Nobel

August, 2021

Vectors, Addition and Scalar Multiplication

Vectors $\mathbf{u} \in \mathbb{R}^d$ represented in column form by convention

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix} = (u_1, \ldots, u_d)^t$$

Vector addition and scalar multiplication: for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ and $a \in \mathbb{R}$

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ \vdots \\ u_d + v_d \end{bmatrix} \quad \text{and} \quad a\mathbf{u} = \begin{bmatrix} au_1 \\ \vdots \\ au_d \end{bmatrix}$$

Inner Product

Definition: The *inner product* of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ is given by

$$\langle \mathbf{u}, \mathbf{v} \rangle \;=\; \mathbf{u}^t \mathbf{v} \;=\; \sum_{i=1}^{d} u_i \, v_i$$

Basic Properties: Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^d$ and $a, b \in \mathbb{R}$

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle a\mathbf{u}, b\mathbf{v} \rangle = ab \, \langle \mathbf{u}, \mathbf{v} \rangle$
3. $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$

Euclidean Norm

Definition: The *Euclidean norm* of a vector $\mathbf{u} \in \mathbb{R}^d$ is

$$||\mathbf{u}|| = \langle \mathbf{v}, \mathbf{u} \rangle^{1/2} = (u_1^2 + \cdots + u_d^2)^{1/2}$$

Basic Properties

1. $||\mathbf{u}|| \geq 0$ with equality if and only if $x = 0$
2. For $a \in \mathbb{R}$, $||a\,\mathbf{u}|| = |a|\,||\mathbf{u}||$
3. $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + ||\mathbf{v}||^2$
4. $|\langle \mathbf{u}, \mathbf{v} \rangle| = |\mathbf{u}^t \mathbf{v}| \;\leq\; ||\mathbf{u}||\,||\mathbf{v}||$ (Cauchy-Schwarz inequality)
5. $||\mathbf{u} + \mathbf{v}|| \;\leq\; ||\mathbf{u}|| + ||\mathbf{v}||$ (triangle inequality)
6. $|\,||\mathbf{u}|| - ||\mathbf{v}||\,| \;\leq\; ||\mathbf{u} - \mathbf{v}||$ (reverse triangle inequality)

Orthogonality and Projections

Definition: Vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal, written $\mathbf{u} \perp \mathbf{v}$, if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$

Defn: Let V be a subspace of $\mathbb{R}^n$. The *projection* of $\mathbf{u} \in \mathbb{R}^n$ onto V is the vector $\mathbf{w} \in V$ closest to $\mathbf{u}$. Formally,

$$\text{proj}_V(\mathbf{u}) = \underset{\mathbf{w} \in V}{\operatorname{argmin}} \, ||\mathbf{u} - \mathbf{w}||$$

Fact: Let $V = \{\alpha \mathbf{v} : \alpha \in \mathbb{R}\}$ be the 1-d subspace generated by $\mathbf{v} \in \mathbb{R}^n$

1. $\text{proj}_V(\mathbf{u}) = \langle \mathbf{u}, \mathbf{v} \rangle \, \mathbf{v} / ||\mathbf{v}||^2$
2. $(\mathbf{u} - \text{proj}_V(\mathbf{u})) \perp \mathbf{v}$

Matrix Basics

Notation: $\mathbb{R}^{m \times n}$ denotes set of $m \times n$ matrices $\mathbf{A}$ with real entries

$$\mathbf{A} = \{a_{ij} : 1 \le i \le m, 1 \le j \le n\}$$

- *Transpose* of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is $\mathbf{A}^t \in \mathbb{R}^{n \times m}$ defined by $a^t_{ij} = a_{ji}$
- $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *symmetric* if $\mathbf{A}^t = \mathbf{A}$
- If $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$ then product $\mathbf{C} = \mathbf{A}\,\mathbf{B} \in \mathbb{R}^{m \times p}$ has entries

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

- In general, $\mathbf{A}\,\mathbf{B} \neq \mathbf{B}\,\mathbf{A}$
- $(\mathbf{A}\,\mathbf{B})^t = \mathbf{B}^t \mathbf{A}^t$

Determinants and Inverses

The determinant of an $n \times n$ matrix $\mathbf{A}$ is denoted by $\det(\mathbf{A})$.

- $\det(\mathbf{A}\,\mathbf{B}) = \det(\mathbf{A})\,\det(\mathbf{B})$
- $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$
- $\det(\mathbf{A}^t) = \det(\mathbf{A})$

The *inverse* of $\mathbf{A}$ is the unique matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}^{-1}\,\mathbf{A} \;=\; \mathbf{A}\,\mathbf{A}^{-1} \;=\; \mathbf{I}$$

- $\mathbf{A}^{-1}$ exists iff $\det(\mathbf{A}) \neq 0$.
- If $\mathbf{A}$ and $\mathbf{B}$ invertible, then $(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}$

Eigenvalues and Eigenvectors

Each matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has n (possibly complex) *eigenvalues* $\lambda_1, \dots, \lambda_n$

For each eigenvalue λ_i there is a corresponding *eigenvector* $\mathbf{v}_i \neq 0$ s.t.

$$\mathbf{A}\,\mathbf{v}_i \;=\; \lambda_i\,\mathbf{v}_i$$

- $\lambda_1, \dots, \lambda_n$ are the roots of the polynomial $p(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$
- $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$
- Eigenvalues can be repeated
- If $\mathbf{A}$ is symmetric then all of its eigenvalues are real

Orthogonal Matrices

Vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are *orthonormal* if $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbb{I}(i = j)$ for $1 \leq i, j \leq n$

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *orthogonal* if $\mathbf{A}^t \mathbf{A} = \mathbf{I}$. If $\mathbf{A}$ is orthogonal then

- $\mathbf{A}^{-1} = \mathbf{A}^t$
- $\mathbf{A}\,\mathbf{A}^t = \mathbf{I}$
- the rows and columns of $\mathbf{A}$ are orthonormal
- the eigenvalues $\lambda_i(\mathbf{A}) \in \{+1, -1\}$
- $\det(\mathbf{A}) \in \{+1, -1\}$

Quadratic Forms

Each symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has an associated *quadratic form* $q_A : \mathbb{R}^n \to \mathbb{R}$ defined by

$$q_A(\mathbf{u}) = \mathbf{u}^t \mathbf{A}\, \mathbf{u} = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i\, a_{ij}\, u_j$$

- $\mathbf{A}$ is *non-negative definite* ($\mathbf{A} \geq 0$) if $\mathbf{u}^t \mathbf{A}\, \mathbf{u} \geq 0$ for every $\mathbf{u}$
- $\mathbf{A}$ is *positive definite* ($\mathbf{A} > 0$) if $\mathbf{u}^t \mathbf{A}\, \mathbf{u} > 0$ for every $\mathbf{u} \neq \mathbf{0}$

Fact: Let $\mathbf{A}$ $n \times n$ be symmetric.

- $\mathbf{A} \geq 0$ iff all its eigenvalues are non-negative
- $\mathbf{A} > 0$ iff all its eigenvalues are positive

Trace of a Matrix

Definition: The *trace* of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements

$$\text{tr}(A) \;=\; \sum_{i=1}^{n} a_{ii}$$

- $\text{tr}(\mathbf{A}) =$ sum of eigenvalues of $\mathbf{A}$
- $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^t)$
- If $\mathbf{B}$ is $n \times n$ then $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$

Frobenius Norm

Definition: The *Frobenius norm* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is

$$||\mathbf{A}|| = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

Basic Properties

- $||\mathbf{A}||^2 = \text{tr}(\mathbf{A}^t \mathbf{A})$
- $||\mathbf{A}|| = 0$ if and only if $\mathbf{A} = 0$
- $||b\mathbf{A}|| = |b| \, ||\mathbf{A}||$
- $||\mathbf{A} + \mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}||$
- $||\mathbf{A}\mathbf{B}|| \leq ||\mathbf{A}|| \, ||\mathbf{B}||$

Rank of a Matrix

Definition: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be an m x n matrix

- row-space of $\mathbf{A}$ = span of the rows of $\mathbf{A}$ (subspace of $\mathbb{R}^n$)
- col-space of $\mathbf{A}$ = span of the cols of $\mathbf{A}$ (subspace of $\mathbb{R}^m$)
- row-rank$(\mathbf{A}) :=$ dim of the row-space of $\mathbf{A}$ (at most n)
- col-rank$(\mathbf{A}) :=$ dim of the col-space of $\mathbf{A}$ (at most m)

Fact: row-rank$(\mathbf{A}) =$ col-rank$(\mathbf{A})$

Definition: The *rank* of $\mathbf{A}$ is the common value of the row and column ranks

Basic Properties of the Rank

- If $\mathbf{A} \in \mathbb{R}^{m \times n}$ then $\text{rank}(\mathbf{A}) \leq \min\{m, n\}$
- $\text{rank}(\mathbf{A}\,\mathbf{B}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$
- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$
- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^t) = \text{rank}(\mathbf{A}^t\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^t)$
- $\mathbf{A} \in \mathbb{R}^{n \times n}$ has at most $\text{rank}(\mathbf{A})$ non-zero eigenvalues
- $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible iff $\text{rank}(\mathbf{A}) = n$, that is, $\mathbf{A}$ is of full rank

Outer Products

Definition: The *outer product* $\mathbf{u}\mathbf{v}^t$ of vectors $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ is an $m \times n$ matrix with entries

$$(\mathbf{u}\mathbf{v}^t)_{ij} \ = u_i v_j$$

- If $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ then rank$(\mathbf{u}\mathbf{v}^t) = 1$
- $||\mathbf{u}\mathbf{v}^t|| = ||\mathbf{u}|| \, ||\mathbf{v}||$
- If $m = n$ then tr$(\mathbf{u}\mathbf{v}^t) = \langle \mathbf{u}, \mathbf{v} \rangle$

The Spectral Theorem

Spectral Theorem: If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.

Spectral Decomposition: If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric then $\mathbf{A}$ can be expressed as

$$\mathbf{A} \; = \; \mathbf{\Gamma} \, \mathbf{D} \, \mathbf{\Gamma}^t$$

where $\mathbf{\Gamma} \in \mathbb{R}^{n \times n}$ is orthogonal and $\mathbf{D} = \text{diag}(\lambda_1(A), \ldots, \lambda_n(A))$.

Corollary: If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric then for $k \geq 1$ we have

$$\mathbf{A}^k \; = \; \mathbf{\Gamma} \, \mathbf{D}^k \, \mathbf{\Gamma}^t$$

Courant Fischer Theorem

Thm: Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric with eigenvalues $\lambda_1(\mathbf{A}) \geq \cdots \geq \lambda_n(A)$.

$$\lambda_1(\mathbf{A}) = \max_{\mathbf{v} \neq 0} \frac{\mathbf{v}^t \mathbf{A} \mathbf{v}}{\mathbf{v}^t \mathbf{v}} = \max_{\mathbf{v}:||\mathbf{v}||=1} \mathbf{v}^t \mathbf{A} \mathbf{v}$$

$$\lambda_n(\mathbf{A}) = \min_{\mathbf{v} \neq 0} \frac{\mathbf{v}^t \mathbf{A} \mathbf{v}}{\mathbf{v}^t \mathbf{v}} = \min_{\mathbf{v}:||\mathbf{v}||=1} \mathbf{v}^t \mathbf{A} \mathbf{v}$$

$$\lambda_i(\mathbf{A}) = \max_{V:\dim(V)=i} \min_{\mathbf{v} \in V, ||\mathbf{v}||=1} \mathbf{v}^t \mathbf{A} \mathbf{v}$$