Notes: Mathematics of Machine Learning

1807730

October 2022

1 Linear Algebra

Definition 1.1 (An algebra) A set of objects and a set of rules to manipulate these objects is known as an algebra.

Definition 1.2 (A vector) Vectors are objects which when added together or multiplied by a scalar, they produce an object of the same kind.

Examples of vectors include geometric vectors, polynomials and audio signals. Linear Algebra is the study of vectors and the set of rules to manipulate them.

1.1 Matrices

Definition 1.3 (Matrix) With $m, n \in \mathbb{N}$ a real-valued (m, n) matrix $\mathbf{A}$ is an $m.n$ tuple of elements $a_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n$ which is ordered according to a rectangular scheme consisting of m rows and n columns:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1}$$

1.1.1 Matrix Operations

The sum of two matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times n}$ is defined as the element-wise sum i.e.

$$\mathbf{A} + \mathbf{B} := \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{1n} \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{2}$$

For matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times k}$, the elements c_{ij} of the product $\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{m \times k}$ are computed as:

$$\sum_{l=1}^{n} a_{il} b_{lj}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \tag{3}$$

For matrix multiplication $c_{ij} \neq a_{ij} b_{ij}$. If it was, the element-wise operation is known as the *Hadamard Product*. Matrix multiplication is not commutative i.e. $\mathbf{AB} \neq \mathbf{BA}$

1.1.2 Types of Matrices

Definition 1.4 (Identity matrix) In $\mathbb{R}^{n\times n}$, we define the identity matrix as an $n \times x$-matrix containing 1 on the diagonal and 0 everywhere else:

$$I_n := \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \tag{4}$$

1.1.3 Properties of Matrices

Associativity:

$$\forall \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}, \mathbf{C} \in \mathbb{R}^{p\times q} : (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \tag{5}$$

Distributivity:

$$\forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{n\times p} : (\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \tag{6}$$

$$\mathbf{A}(\mathbf{C}+\mathbf{D}) = \mathbf{AC} + \mathbf{AD} \tag{7}$$

Multiplication with the identity matrix:

$$\forall \mathbf{A} \in \mathbb{R}^{m\times n} : \mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A} \tag{8}$$

Note that $\mathbf{I}_m \neq \mathbf{I}_n$ for $m \neq n$.

1.1.4 Inverse and Transpose of a Matrix

Definition 1.5 (Inverse) Consider a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$. Let the matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ have the property that $\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}$. $\mathbf{B}$ is called inverse of $\mathbf{A}$ and denoted by $\mathbf{A}^{-1}$.

Not every matrix $\mathbf{A}$ possesses an inverse $\mathbf{A}^{-1}$. If this inverse does exist, $\mathbf{A}$ is called *regular/invertible/nonsingular*, otherwise *singular/noninvertible*. When the inverse does exist, it is unique.

Definition 1.6 (Transpose) *For $A \in \mathbb{R}^{m\times n}$ the matrix $B \in \mathbb{R}^{n\times m}$ with $b_{ij} = a_{ji}$ is called the transpose of $\mathbf{A}$. We write $\mathbf{B} = \mathbf{A}^\top$.*

The following are properties of inverses and transposes:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A} \tag{9}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{10}$$

$$(\mathbf{A}+\mathbf{B})^{-1} \neq \mathbf{A}^{-1} + \mathbf{B}^{-1} \tag{11}$$

$$(\mathbf{A}^\top)^\top = \mathbf{A} \tag{12}$$

$$(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top \tag{13}$$

$$(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top \tag{14}$$

Definition 1.7 (Symmetric Matrix) *A matrix* $A \in \mathbb{R}^{n \times n}$ *is symmetric if* $\boldsymbol{A} = \boldsymbol{A}^\top$

1.1.5 Multiplication by a Scalar

Let $A \in \mathbb{R}^{m \times n}$ and $\lambda \in \mathbb{R}$. Then $\lambda\mathbf{A} = \mathbf{K}$, $K_{ij} = \lambda a_{ij}$. Practically, λ scales each element of $\mathbf{A}$. For $\lambda, \psi \in \mathbb{R}$, the following holds:

Associativity:

$$(\lambda\psi)\mathbf{C} = \lambda(\psi\mathbf{C}), \quad \mathbf{C} \in \mathbb{R}^{m \times n} \tag{15}$$
$$\lambda(\mathbf{BC}) = (\lambda\mathbf{B})\mathbf{C} = \mathbf{B}(\lambda\mathbf{C}) = (\mathbf{BC})\lambda, \quad \mathbf{B} \in \mathbb{R}^{m \times n}, \mathbf{C} \in \mathbb{R}^{n \times k} \tag{16}$$
$$(\lambda\mathbf{C})^\top = \mathbf{C}^\top\lambda^\top = \mathbf{C}^\top\lambda = \lambda\mathbf{C}^\top \text{ since } \lambda = \lambda^\top \ \forall \ \lambda \in \mathbb{R} \tag{17}$$

Distributivity:

$$(\lambda + \psi)\mathbf{C} = \lambda\mathbf{C} + \psi\mathbf{C}, \quad \mathbf{C} \in \mathbb{R}^{m \times n} \tag{18}$$
$$\lambda(\mathbf{B} + \mathbf{C}) = \lambda\mathbf{B} + \lambda\mathbf{C}, \quad \mathbf{C}, \mathbf{B} \in \mathbb{R}^{m \times n} \tag{19}$$

1.2 Vector Spaces

1.2.1 Groups

Definition 1.8 (Group) Consider a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ defined on $\mathcal{G}$. Then $\mathcal{G} := (\mathcal{G}, \otimes)$ is called a group if the following hold:

1. Closure of $\mathcal{G}$ under $\otimes : \forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. Associativity: $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. Neutral element: $\exists e \in \mathcal{G} \ \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. Inverse element: $\forall x \in \mathcal{G} \ \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$, where e is the neutral element. This is often denoted as x^{-1} however this is defined with respect to $\otimes$ and is not $\frac{1}{x}$.

If additionally $\forall x, y \in \mathcal{G} : x \otimes y = y \otimes x$, then $\mathcal{G} = (\mathcal{G}, \otimes)$ is an *Abelian group* (commutative).

Definition 1.9 (General Linear Group) The set of regular (invertible) matrices $A \in \mathbb{R}^{n \times n}$ is a group with respect to matrix multiplication and is called *general linear group* $GL(n, \mathbb{R})$. However, since matrix multiplication is not commutative, the group is not Abelian.

1.2.2 Vector Spaces

We will consider sets that in addition to an inner operation $+$ also contain an outer operation $\cdot$. The multiplication of a vector $x \in \mathcal{G}$ by a scalar $\lambda \in \mathbb{R}$ would be considered the outer operation.

Definition 1.10 (Vector Space) A real-valued *vector space* $V = (V, +, \cdot)$ is a set $\mathcal{V}$ with two operations:

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V} \quad \text{(Inner Operation)} \tag{20}$$
$$\cdot : \mathbb{E} \times \mathcal{V} \to \mathcal{V} \quad \text{(Outer Operation)} \tag{21}$$

where:

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
   (a) $\forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$
   (b) $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : (\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \mathbf{x}) = (\lambda \psi) \cdot \mathbf{x}$
4. Neutral element with respect to the outer operation: $\forall \mathbf{x} \in \mathcal{V} : 1 \cdot \mathbf{x} = \mathbf{x}$

The elements $\mathbf{x} \in \mathcal{V}$ are called vectors. The neutral element of $(\mathcal{V}, +)$ is the zero vector $\mathbf{0} = [0, \ldots, 0]^\top$, and the inner operation $+$ is called vector addition. The elements $\lambda \in \mathbb{R}$ are called scalars and the outer operation $\cdot$ is a multiplication by scalars.

1.2.3 Vector Subspaces

Definition 1.11 (Vector Subspaces) Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}, \mathcal{U} \neq \emptyset$. Then $\mathcal{U} = (\mathcal{U}, +, \cdot)$ is called vector subspace of V (or linear subspace) if U is a vector space with the vector space operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathcal{R} \times \mathcal{U}$. We write $U \subseteq V$ to denote a subspace U of V.

If $\mathcal{U} \subseteq \mathcal{V}$ and V is a vector space, then U will inherit many properties from V. This includes the Abelian group properties, the distributivity, the associativity and the neutral element. To determine whether $(\mathcal{U}, +, \cdot)$ is a subspace of V we still do need to show:

1. $\mathcal{U} \neq \emptyset$, in particular: $\mathbf{0} \in \mathcal{U}$
2. Closure of U:
   (a) With respect to the outer operation: $\forall \lambda \in \mathbb{R} \forall \mathbf{x} \in \mathcal{U} : \lambda \mathbf{x} \in \mathcal{U}$
   (b) With respect to the inner operation: $\forall \mathbf{x}, \mathbf{y} \in \mathcal{U} : \mathbf{x} + \mathbf{y} \in$

1.3 Linear Independence

Definition 1.12 (Linear Combination) Consider a vector space V and a finite number of vectors $\boldsymbol{x}_i, \ldots, \boldsymbol{x}_k \in V$. Then, every $\boldsymbol{v} \in V$ of the form

$$\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \in V \tag{22}$$

with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a *linear combination* of the vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$.

Definition 1.13 (Linear (In)dependence) Let us consider a vector space V with $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k \in V$. If there is a non-trivial combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ are *linearly dependent*.

1.4 Basis and Rank

1.4.1 Generating Set and Basis

Definition 1.14 (Generating Set and Span) Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \subseteq V$. If every vector $\boldsymbol{v} \in V$ can be expressed as a linear combination of vectors of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, $\mathcal{A}$ is called the generating set of V. The set of all linear combinations of vectors in $\mathcal{A}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans the vector space V, we write $V = span[\mathcal{A}]$.

Definition 1.15 (Basis) Consider a vector space $V = (V, +, \cdot)$ and $\mathcal{A} \subseteq \mathcal{V}$. A generating set $\mathcal{A}$ of V is called *minimal* if there exists no smaller set $\tilde{A} \subseteq \mathcal{A} \subseteq \mathcal{V}$ that spans V. Every linearly independent generating set of V is minimal and is called a *basis* or V.

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{B} \subseteq \mathcal{V}$, $\mathcal{B} \neq \emptyset$. Then, the following statements are equivalent:

- $\mathcal{B}$ is a basis of $\mathcal{V}$.
- $\mathcal{B}$ is the minimal generating set.
- $\mathcal{B}$ is a maximal linearly independent set of vectors in V, i.e., adding other vectors to this set will make it linearly dependent.
- Every vector $\boldsymbol{x} \in V$ is a linear combination of vectors from $\mathcal{B}$, and every linear combination is unique, i.e., with

  $$\boldsymbol{x} = \sum_{i=1}^{k} \lambda_i \boldsymbol{b}_i = \sum_{i=1}^{k} \psi_i \boldsymbol{b}_i \tag{23}$$

  and $\lambda_i, \psi_i \in \mathbb{R}, \boldsymbol{b}_i \in \mathcal{B}$ it follows that $\lambda_i = \psi_i, i = 1, \ldots, k$.

The *dimension* of V is the number of basis vectors of V, and we write $dim(V)$. if $U \subseteq V$ is a subspace of V, then $dim(U) \leq dim(V)$ and $dim(U) = dim(V)$ if and only if $U = V$. Intuitively, the dimension of a vector space can be thought of as the number of independent directions in this vector space.

Definition 1.16 (Rank) The number of linearly independent columns of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows and is called the *rank* of $\mathbf{A}$ and is denoted by $rk(\mathbf{A})$.