Mathematics for Machine Learning

Vazgen Mikayelyan

July 21, 2020

Linear Independence

Linear Independence

Definition

A vector $\mathbf{v} \in V$ *is called linear combination of vectors* $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$*, if there are scalars* $c_1, \ldots, c_k \in \mathbb{R}$ *so that*

$$\mathbf{v} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_k\mathbf{x}_k = \sum_{i=1}^{k} c_i\mathbf{x}_i.$$

Linear Independence

Definition

The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$ *are called* **linearly dependent**, *if there is a non-trivial linear combination, such that*

$$\sum_{i=1}^{k} c_i \mathbf{x}_i = \mathbf{0}$$

with at least one $c_i \neq 0$. *If only the trivial solution exists, i.e.,* $c_1 = \ldots = c_k = 0$, *then the vectors* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ *are* **linearly independent.**

Linear Independence

When vectors are linearly dependent?

- If at least one of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is $\mathbf{0}$ then they are linearly dependent.

Linear Independence

When vectors are linearly dependent?

- If at least one of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is $\mathbf{0}$ then they are linearly dependent.
- The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent if and only if (at least) one of them is a linear combination of the others. In particular, if one vector is a multiple of another vector, i.e., $\mathbf{x}_i = c\mathbf{x}_j, c \in \mathbb{R}$, then the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is linearly dependent.

Linear Independence

How to check linear independence?

A way of checking whether vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent is to write all vectors as columns of a matrix A. Gaussian elimination yields a matrix in (reduced) row echelon form

- The pivot columns indicate the vectors, which are linearly independent.
- The non-pivot columns can be expressed as linear combinations of the pivot columns on their left.

Linear Independence

Example

Are vectors $\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}1\\3\\5\end{bmatrix}$ *linearly dependent?*

Linear Independence

Let $\mathbf{b}_1, \ldots, \mathbf{b}_k$ be linearly independent vectors and

$$\mathbf{x}_j = \sum_{i=1}^{k} c_{ij}\mathbf{b}_i,\ j = 1, \ldots, m.$$

Linear Independence

Let $\mathbf{b}_1, \ldots, \mathbf{b}_k$ be linearly independent vectors and

$$\mathbf{x}_j = \sum_{i=1}^{k} c_{ij}\mathbf{b}_i, \; j = 1, \ldots, m.$$

Equivalently $\mathbf{x}_j = B\mathbf{c}_j, \text{ where } B = [\mathbf{b}_1, \ldots, \mathbf{b}_k], \; \mathbf{c}_j = \begin{bmatrix} c_{1j} \\ \vdots \\ c_{kj} \end{bmatrix}$.

Linear Independence

Let $\mathbf{b}_1, \ldots, \mathbf{b}_k$ be linearly independent vectors and

$$\mathbf{x}_j = \sum_{i=1}^{k} c_{ij} \mathbf{b}_i, \; j = 1, \ldots, m.$$

Equivalently $\mathbf{x}_j = B\mathbf{c}_j, \text{ where } B = [\mathbf{b}_1, \ldots, \mathbf{b}_k], \; \mathbf{c}_j = \begin{bmatrix} c_{1j} \\ \vdots \\ c_{kj} \end{bmatrix}.$

Proposition

The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent **iff** *the columns $\mathbf{c}_1, \ldots, \mathbf{c}_m$ are linearly independent.*

Example

Consider linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \in \mathbb{R}^n$ *Let*

$$\begin{array}{rcrcrcr} \mathbf{x}_1 & = & -\mathbf{b}_1 & + & 2\mathbf{b}_2 & + & 3\mathbf{b}_3 \\ \mathbf{x}_2 & = & \mathbf{b}_1 & + & \mathbf{b}_2 & - & 2\mathbf{b}_3 \\ \mathbf{x}_3 & = & -\mathbf{b}_1 & + & 5\mathbf{b}_2 & + & 4\mathbf{b}_3 \end{array}$$

Are the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$ *linearly independent?*

Linear Independence

Example

Consider linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \in \mathbb{R}^n$ *Let*

$$\begin{aligned} \mathbf{x}_1 &= -\mathbf{b}_1 + 2\mathbf{b}_2 + 3\mathbf{b}_3 \\ \mathbf{x}_2 &= \mathbf{b}_1 + \mathbf{b}_2 - 2\mathbf{b}_3 \\ \mathbf{x}_3 &= -\mathbf{b}_1 + 5\mathbf{b}_2 + 4\mathbf{b}_3 \end{aligned}$$

Are the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$ *linearly independent? Let*

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{b}_1 - 2\mathbf{b}_2 + 3\mathbf{b}_3 \\ \mathbf{y}_2 &= \mathbf{b}_1 + \mathbf{b}_2 - 3\mathbf{b}_3 \\ \mathbf{y}_3 &= 2\mathbf{b}_1 + \mathbf{b}_2 - 3\mathbf{b}_3 \end{aligned}$$

Are the vectors $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in \mathbb{R}^n$ *linearly independent?*

Basis and Rank

Generating Set and Basis

Definition

Consider a set of vectors $\mathcal{A}$ in a vector space V. The set of all linear combinations of vectors in $\mathcal{A}$ is called the **span** *of $\mathcal{A}$ and is denoted by* $\operatorname{span}(\mathcal{A})$, *i.e.*

$$\operatorname{span}(\mathcal{A}) = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i;\ k \in \mathbb{N},\ \mathbf{x}_i \in \mathcal{A},\ \lambda_i \in R \right\}$$

If every vector $v \in V$ can be expressed as a linear combination of the vectors of $\mathcal{A}$, then $\mathcal{A}$ is called a **generating set** *of V and we say $\mathcal{A}$ spans the vector space V.*

Generating Set and Basis

Definition

A **basis** $\mathcal{B}$ *of a vector space* V *is a linearly independent subset of* V *that spans* V*. Equivalently,* $\mathcal{B}$ *is a basis of* V *if*

- $\mathcal{B}$ *is linearly independent*
- $\mathcal{B}$ *is a spanning set of* V*:* $\operatorname{span}(\mathcal{B}) = V$.

Generating Set and Basis

Definition

A **basis** $\mathcal{B}$ *of a vector space* V *is a linearly independent subset of* V *that spans* V. *Equivalently,* $\mathcal{B}$ *is a basis of* V *if*

- $\mathcal{B}$ *is linearly independent*
- $\mathcal{B}$ *is a spanning set of* V*:* $\text{span}(\mathcal{B}) = V$.

Definition

A generating set $\mathcal{A}$ *of* V *is called* **minimal** *if there exists no smaller set* $\tilde{\mathcal{A}} \subset \mathcal{A} \subset V$ *that spans* V*.*

Generating Set and Basis

Proposition

The following statements are equivalent:

- $\mathcal{B}$ *is a basis of* V
- $\mathcal{B}$ *is a minimal generating (spanning) set of* V
- $\mathcal{B}$ *is a maximal linearly independent set of vectors in* V *, i.e., adding any other vector to this set will make it linearly dependent.*
- *Every vector* $\mathbf{x} \in V$ *is a unique linear combination of vectors from* $\mathcal{B}$*, i.e.*

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{b}_i = \sum_{i=1}^{k} \psi_i \mathbf{b}_i$$

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

Example

- In $\mathbb{R}^3$, the canonical/standard basis is

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Example

- In $\mathbb{R}^3$, the canonical/standard basis is

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

- Different bases in $\mathbb{R}^3$ are

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}, \mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$$

Example

The set

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

is linearly independent, but not a generating set (and not a basis) of $\mathbb{R}^4$: For instance, the vector $[0, 0, 0, 1]^T$ cannot be obtained by a linear combination of elements in $\mathcal{A}$.

Generating Set and Basis

Proposition

Let V be a vector space with a finite basis. Then every basis of V has the same number of elements.

Generating Set and Basis

Proposition

Let V be a vector space with a finite basis. Then every basis of V has the same number of elements.

Definition

The **dimension** *of V is the number of basis vectors, and it is denoted by* $\dim(V)$.

Generating Set and Basis

Proposition

Let V be a vector space with a finite basis. Then every basis of V has the same number of elements.

Definition

The **dimension** *of V is the number of basis vectors, and it is denoted by* $\dim(V)$.

How to find a basis?

A basis of a subspace $U = \text{span}[\mathbf{x}_1, \ldots, \mathbf{x}_m] \subset \mathbb{R}^n$ can be found by executing the following steps:

1. Write the spanning vectors as columns of a matrix A
2. Determine the row echelon form of A,
3. The spanning vectors associated with the pivot columns are a basis of U.

Generating Set and Basis

Example

Determine a basis for the subspace $U \subset \mathbb{R}^5$ *spanned by the vectors*

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 0 \\ 2 \\ -1 \\ -1 \\ -2 \end{bmatrix}.$$

Rank

Proposition

The number of linearly independent columns of a matrix $A \in \mathbb{R}^{m \times n}$ *equals the number of linearly independent rows.*

Rank

Proposition

The number of linearly independent columns of a matrix $A \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows.

Definition

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

Rank

Proposition

The number of linearly independent columns of a matrix $A \in \mathbb{R}^{m \times n}$ *equals the number of linearly independent rows.*

Definition

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

Corollary

If $A \in \mathbb{R}^{m \times n}$ *then* $\operatorname{rank}(A) \leq \min(m, n)$.

Rank

Proposition

The number of linearly independent columns of a matrix $A \in \mathbb{R}^{m \times n}$ *equals the number of linearly independent rows.*

Definition

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

Corollary

If $A \in \mathbb{R}^{m \times n}$ *then* $\mathrm{rank}\,(A) \leq \min\,(m, n)$.

Definition

A matrix $A \in \mathbb{R}^{m \times n}$ *has* **full rank** *if its rank equals the largest possible rank for a matrix of the same dimensions, i.e. rank*$(A) = \min(m, n)$.
A matrix is said to be **rank deficient** *if it does not have full rank.*

Rank

Proposition

- *rank*$(A) =$ *rank*(A^T) *for all* $A \in \mathbb{R}^{m \times n}$

Rank

Proposition

- *rank*$(A) =$ *rank*(A^T) *for all* $A \in \mathbb{R}^{m \times n}$
- *The columns of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $U \subset \mathbb{R}^m$ *with* $\dim(U) =$ *rank*(A). *Later, we will call this subspace the image or range.*

Rank

Proposition

- *rank*$(A) =$ *rank*(A^T) *for all* $A \in \mathbb{R}^{m \times n}$
- *The columns of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $U \subset \mathbb{R}^m$ *with* $\dim(U) =$ *rank*(A). *Later, we will call this subspace the image or range.*
- *The rows of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $W \subset \mathbb{R}^n$ *with* $\dim(W) =$ *rank*(A).

Rank

Proposition

- *rank*$(A) =$ *rank*(A^T) *for all* $A \in \mathbb{R}^{m\times n}$
- *The columns of* $A \in \mathbb{R}^{m\times n}$ *span a subspace* $U \subset \mathbb{R}^m$ *with* $\dim(U) =$ *rank*(A). *Later, we will call this subspace the image or range.*
- *The rows of* $A \in \mathbb{R}^{m\times n}$ *span a subspace* $W \subset \mathbb{R}^n$ *with* $\dim(W) =$ *rank*(A).
- *For all* $A \in \mathbb{R}^{n\times n}$ *holds:* A *is regular (invertible) if and only if* A *has full rank, i.e. rank*$(A) = n$.

Rank

Proposition

- *rank*$(A) =$ *rank*(A^T) *for all* $A \in \mathbb{R}^{m \times n}$
- *The columns of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $U \subset \mathbb{R}^m$ *with* $\dim(U) =$ *rank*(A). *Later, we will call this subspace the image or range.*
- *The rows of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $W \subset \mathbb{R}^n$ *with* $\dim(W) =$ *rank*(A).
- *For all* $A \in \mathbb{R}^{n \times n}$ *holds:* A *is regular (invertible) if and only if* A *has full rank, i.e.* *rank*$(A) = n$.
- *For all* $A \in \mathbb{R}^{m \times n}$ *and all* $\mathbf{b} \in \mathbb{R}^m$ *it holds that the SLE* $A\mathbf{x} = \mathbf{b}$ *has a solution if and only if* *rank*$(A) =$ *rank*$(A|\mathbf{b})$, *where* $A|\mathbf{b}$ *denotes the augmented matrix.*

Rank

Proposition

- *rank*$(A) =$ *rank*(A^T) *for all* $A \in \mathbb{R}^{m \times n}$
- *The columns of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $U \subset \mathbb{R}^m$ *with* $\dim(U) =$ *rank*(A). *Later, we will call this subspace the image or range.*
- *The rows of* $A \in \mathbb{R}^{m \times n}$ *span a subspace* $W \subset \mathbb{R}^n$ *with* $\dim(W) =$ *rank*(A).
- *For all* $A \in \mathbb{R}^{n \times n}$ *holds:* A *is regular (invertible) if and only if* A *has full rank, i.e. rank*$(A) = n$.
- *For all* $A \in \mathbb{R}^{m \times n}$ *and all* $\mathbf{b} \in \mathbb{R}^m$ *it holds that the SLE* $A\mathbf{x} = \mathbf{b}$ *has a solution if and only if rank*$(A) =$ *rank*$(A|\mathbf{b})$, *where* $A|\mathbf{b}$ *denotes the augmented matrix.*
- *For* $A \in \mathbb{R}^{m \times n}$ *the dimension of the subspace of solutions for* $A\mathbf{x} = \mathbf{0}$ *is* $n -$ *rank*(A). *Later, we will call this subspace the kernel or the null space.*

Rank

Example

Find the rank of the following matrices

$$A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 6 \\ 0 & 3 & 6 & 9 \end{bmatrix}, \; B = \begin{bmatrix} 1 & 5 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \; C = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & 3 \\ 3 & 7 & 9 \end{bmatrix}.$$

Linear Mappings

Linear Mappings

Definition

For vector spaces V, W, a mapping $T : V \to W$ is called a linear mapping (or vector space homomorphism/ linear transformation) if

$$T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y}),\ \forall x, y \in V, a, b \in \mathbb{R}$$

or equivalently,

- $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$
- $T(a\mathbf{x}) = aT(\mathbf{x})$

Linear Mappings

Definition

Consider a transformation $T : V \to W$, *where V, W can be arbitrary sets. Then* T *is called*

- *injective, if for any* $\mathbf{x}, \mathbf{y} \in V$ *it follows that* $T(\mathbf{x}) \neq T(\mathbf{y})$ *if and only if* $\mathbf{x} \neq \mathbf{y}$.
- *surjective, if* $T(V) = W$.
- *bijective if it is injective and surjective.*

Linear Mappings

Definition

Consider a transformation $T : V \to W$, where V, W can be arbitrary sets. Then T is called

- *injective, if for any $\mathbf{x}, \mathbf{y} \in V$ it follows that $T(\mathbf{x}) \neq T(\mathbf{y})$ if and only if $\mathbf{x} \neq \mathbf{y}$.*
- *surjective, if $T(V) = W$.*
- *bijective if it is injective and surjective.*

Definition

Let V, W be vector spaces.

- *$T : V \to W$ is called* **isomorphism**, *if it is linear and bijective.*
- *$T : V \to V$ is called* **endomorphism**, *if T is linear.*
- *$T : V \to V$ is called* **automorphism**, *if T is linear and bijective.*
- *We define $id_V : V \to V$, $id_v(\mathbf{x}) = \mathbf{x}$ as the* **identity mapping** *in V.*

Linear Mappings

Definition

Vector spaces V and W are called isomorphic if there exists an isomorphism $T : V \to W$ and we write $V \cong W$.

Definition

Vector spaces V *and* W *are called isomorphic if there exists an isomorphism* $T : V \to W$ *and we write* $V \cong W$.

Theorem

Finite-dimensional vector spaces V *and* W *are isomorphic if and only if* $\dim(V) = \dim(W)$.

Definition

Vector spaces V and W are called isomorphic if there exists an isomorphism $T : V \to W$ and we write $V \cong W$.

Theorem

Finite-dimensional vector spaces V and W are isomorphic if and only if $\dim(V) = \dim(W)$.

Example

Show that $R^{m\times n} \cong R^{mn}$.

Matrix Representation of Linear Mappings

Consider a vector space V with an (unordered) basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$. Let $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ be an ordered basis, and $\mathbf{B} = [\mathbf{b_1}, \ldots, \mathbf{b_n}]$ be a matrix whose columns are the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$.

Matrix Representation of Linear Mappings

Consider a vector space V with an (unordered) basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$. Let $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ be an ordered basis, and $\mathbf{B} = [\mathbf{b_1}, \ldots, \mathbf{b_n}]$ be a matrix whose columns are the vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$.

Definition

For any $\mathbf{x} \in V$ there is a unique representation

$$\mathbf{x} = \alpha_1 \mathbf{b}_1 + \ldots + \alpha_n \mathbf{b}_n.$$

Then $\alpha_1, \ldots, \alpha_n$ are the coordinates of $\mathbf{x}$ with respect to B, and the vector

$$[\mathbf{x}]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n$$

is the **coordinate vector** *of $\mathbf{x}$ with respect to the ordered basis B.*