Mathematics for Machine Learning

Vazgen Mikayelyan

July 25, 2020

FAST DISCOVERING THE FUTURE

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* $V \cong W$.

Linear Mappings

Definition

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

Theorem

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

Linear Mappings

Definition

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

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.*

Matrix Representation of Linear Mappings

Definition

Let V, W *be vector spaces with corresponding bases* $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ *and* $C = (\mathbf{c}_1, \ldots, \mathbf{c}_m)$*. Moreover, let* $T : V \to W$ *be a linear mapping with*

$$T(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \ldots + \alpha_{mj}\mathbf{c}_m = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i.$$

Matrix Representation of Linear Mappings

Definition

Let V, W *be vector spaces with corresponding bases* $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ *and* $C = (\mathbf{c}_1, \dots, \mathbf{c}_m)$. *Moreover, let* $T : V \to W$ *be a linear mapping with*

$$T(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \ldots + \alpha_{mj}\mathbf{c}_m = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i.$$

Then, we call the $m \times n$*-matrix* A_T *whose elements are given by*

$$A_T(i, j) = \alpha_{ij}$$

the **transformation matrix** *of* T *(with respect to the ordered bases* B *of* V *and* C *of* W*).*

Matrix Representation of Linear Mappings

Definition

Let V, W be vector spaces with corresponding bases $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $C = (\mathbf{c}_1, \ldots, \mathbf{c}_m)$. Moreover, let $T : V \to W$ be a linear mapping with

$$T(\mathbf{b}_j) = \alpha_{1j}\mathbf{c}_1 + \ldots + \alpha_{mj}\mathbf{c}_m = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i.$$

Then, we call the $m \times n$-matrix A_T whose elements are given by

$$A_T(i, j) = \alpha_{ij}$$

the **transformation matrix** *of T (with respect to the ordered bases B of V and C of W).*

Note that

$$A_T = [[T(\mathbf{b}_1)]_C \ldots [T(\mathbf{b}_n)]_C] \text{ and } [T(\mathbf{x})]_C = A_T[\mathbf{x}]_B.$$

Matrix Representation of Linear Mappings

Example

Write the transformation matrix of the linear mapping (homomorphism) $T : V \to W$ *w.r.t. ordered bases* $B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ *of* V *and* $C = (\mathbf{c}_1, \ldots, \mathbf{c}_5)$ *of* W *satisfying*

$$T(\mathbf{b}_1) = \mathbf{c}_1 - \mathbf{c}_3 + 3\mathbf{c}_5$$
$$T(\mathbf{b}_2) = \mathbf{c}_2 + 2\mathbf{c}_4 + 5\mathbf{c}_5$$
$$T(\mathbf{b}_3) = \mathbf{c}_1 - 3\mathbf{c}_2 + \mathbf{c}_3$$

Using the transformation matrix A_T *find the coordinate vector* $[T(\mathbf{x})]_C$*, given that* $[\mathbf{x}]_B = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$.

Matrix Representation of Linear Mappings

Example

Let's consider three linear transformations of a set of vectors in $\mathbb{R}^2$ *with the transformation matrices*

$$A_1 = \begin{bmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix}, \; A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \; A_3 = \frac{1}{2} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix}.$$

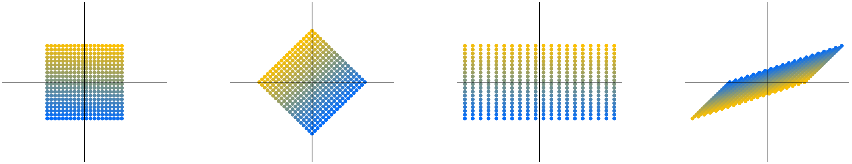
Matrix Representation of Linear Mappings

Example

Let's consider three linear transformations of a set of vectors in $\mathbb{R}^2$ *with the transformation matrices*

$$A_1 = \begin{bmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) \end{bmatrix}, \ A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \ A_3 = \frac{1}{2}\begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix}.$$

How are the vectors in figure a) transformed when applying each of the matrices $A_i, \ i = 1, 2, 3$.



(a) Original data. (b) Rotation by $45°$. (c) Stretch along the horizontal axis. (d) General linear mapping.

Image and Kernel

Definition

For $T : V \to W$, we dfine the kernel/null space

$$\ker(T) := T^{-1}(\mathbf{0}_W) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0_W}\}$$

and the image/range

$$\mathrm{Im}(T) := T(V) = \{\mathbf{w} \in W : \exists \mathbf{v} \in V \text{ s.t. } T(\mathbf{v}) = \mathbf{w}\}.$$

Image and Kernel

Definition

For $T : V \to W$, we dfine the kernel/null space

$$\ker(T) := T^{-1}(\mathbf{0}_W) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}_\mathbf{W}\}$$

and the image/range

$$\mathrm{Im}(T) := T(V) = \{\mathbf{w} \in W : \exists \mathbf{v} \in V \textit{ s.t. } T(\mathbf{v}) = \mathbf{w}\}.$$

Proposition

- $T(\mathbf{0}_V) = \mathbf{0}_W$ *and, therefore,* $\mathbf{0}_V \in \ker(T)$.
- $\mathrm{Im}(T) \subset W$ *is a subspace of* W, *and* $\ker(T) \subset V$ *is a subspace of* V.
- T *is injective (one-to-one) if and only if* $\ker(T) = \{0\}$.

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m \times n}$ *be a matrix with columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *then*

- $\text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m \times n}$ *be a matrix with columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *then*

- $\text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with its transformation matrix $A \in \mathbb{R}^{m \times n}$, i.e. $T(\mathbf{x}) = A\mathbf{x}$.

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m \times n}$ *be a matrix with columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *then*

- $\operatorname{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with its transformation matrix $A \in \mathbb{R}^{m \times n}$, i.e. $T(\mathbf{x}) = A\mathbf{x}$.

- If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, then $\operatorname{Im}(T) = \{A\mathbf{x}; \mathbf{x} \in \mathbb{R}^n\}$
$= \{x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n; x_1, \ldots, x_n \in \mathbb{R}\} = \operatorname{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$.

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m \times n}$ *be a matrix with columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *then*

- $\mathrm{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with its transformation matrix $A \in \mathbb{R}^{m \times n}$, i.e. $T(\mathbf{x}) = A\mathbf{x}$.

- If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, then $\mathrm{Im}(T) = \{A\mathbf{x}; \mathbf{x} \in \mathbb{R}^n\}$
  $= \{x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n; x_1, \ldots, x_n \in \mathbb{R}\} = \mathrm{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$.
- The image $\mathrm{Im}(T)$ is a subspace of $\mathbb{R}^m$.

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m \times n}$ *be a matrix with columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *then*

- $\mathrm{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with its transformation matrix $A \in \mathbb{R}^{m \times n}$, i.e. $T(\mathbf{x}) = A\mathbf{x}$.

- If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, then $\mathrm{Im}(T) = \{A\mathbf{x}; \mathbf{x} \in \mathbb{R}^n\}$
  $= \{x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n; x_1, \ldots, x_n \in \mathbb{R}\} = \mathrm{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$.
- The image $\mathrm{Im}(T)$ is a subspace of $\mathbb{R}^m$.
- rank$(A) = \dim(\mathrm{Im}(T))$

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m\times n}$ *be a matrix with columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, *then*

- $\text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with its transformation matrix $A \in \mathbb{R}^{m\times n}$, i.e. $T(\mathbf{x}) = A\mathbf{x}$.

- If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, then $\text{Im}(T) = \{A\mathbf{x}; \mathbf{x} \in \mathbb{R}^n\}$
  $= \{x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n; x_1, \ldots, x_n \in \mathbb{R}\} = \text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$.
- The image $\text{Im}(T)$ is a subspace of $\mathbb{R}^m$.
- rank$(A) = \dim(\text{Im}(T))$
- The kernel $\ker(T)$ is a subspace of $\mathbb{R}^n$.

Image and Kernel

Definition

Let $A \in \mathbb{R}^{m \times n}$ be a matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then

- $\text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n] \subset \mathbb{R}^m$ *is called* **column space**,
- $\{A\mathbf{x} = 0; \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^n$ *is called* **null space**.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping with its transformation matrix $A \in \mathbb{R}^{m \times n}$, i.e. $T(\mathbf{x}) = A\mathbf{x}$.

- If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, then $\text{Im}(T) = \{A\mathbf{x}; \mathbf{x} \in \mathbb{R}^n\}$ $= \{x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n; x_1, \ldots, x_n \in \mathbb{R}\} = \text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$.
- The image $\text{Im}(T)$ is a subspace of $\mathbb{R}^m$.
- rank$(A) = \dim(\text{Im}(T))$
- The kernel $\ker(T)$ is a subspace of $\mathbb{R}^n$.
- The kernel $\ker(T)$ is the null space of A.

Image and Kernel

Example

Find a basis for image and kernel of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ *given by*

$$T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_2 + 3x_3 \end{bmatrix}$$

Image and Kernel

Example

Find a basis for image and kernel of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ *given by*

$$T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ x_2 + 3x_3 \end{bmatrix}$$

Rank-Nullity Theorem

For vector spaces U, V and a linear mapping $T : U \to V$ it holds that

$$\dim(\ker(T)) + \dim(\mathrm{Im}(T)) = \dim(U),$$

equivalently for a matrix $A \in \mathbb{R}^{n \times m}$

$$\mathrm{nullity}\,(A) + \mathrm{rank}(A) = m,$$

where $\mathrm{nullity}\,(A)$ is the dimension of the null space $\mathrm{null}\,(A)$.

Norm

Definition

Given a vector space V , a **norm** *on* V *is a function* $\|\cdot\|: V \to [0,+)$ *with the following properties: For all* $\mathbf{x}, \mathbf{y} \in V$*, and all* $\lambda \in \mathbb{R}$

- $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ *(triangle inequality)*
- $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ *(absolutely homogeneous)*
- $\|\mathbf{x}\| \geq 0$ *and* $\|\mathbf{x}\| = 0$ *iff* $\mathbf{x} = 0$ *(positive definite)*

Norm

Norm

Example

The **Manhattan norm** *on* $\mathbb{R}^n$ *is defined for* $\mathbf{x} \in \mathbb{R}^n$ *as*

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

The Manhattan norm is also called ℓ_1 **norm**.

Example

The **Manhattan norm** *on* $\mathbb{R}^n$ *is defined for* $\mathbf{x} \in \mathbb{R}^n$ *as*

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

The Manhattan norm is also called ℓ_1 **norm**.

Example

The length of a vector $\mathbf{x} \in \mathbb{R}^n$ *is given by*

$$\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}^T \mathbf{x}}.$$

This norm is called the **Euclidean norm.** *The Euclidean norm is also called* ℓ_2 **norm**.

Inner Products

Inner Products

Definition

A function $f : V \times V \to \mathbb{R}$ *is called* **bilinear** *if*

$$f(a\mathbf{x} + b\mathbf{y}, \mathbf{z}) = af(\mathbf{x}, \mathbf{z}) + bf(\mathbf{y}, \mathbf{z})$$

$$f(\mathbf{x}, a\mathbf{y} + b\mathbf{z}) = af(\mathbf{x}, \mathbf{y}) + bf(\mathbf{x}, \mathbf{z})$$

Inner Products

Definition

A function $f : V \times V \to \mathbb{R}$ *is called* **bilinear** *if*

$$f(a\mathbf{x} + b\mathbf{y}, \mathbf{z}) = af(\mathbf{x}, \mathbf{z}) + bf(\mathbf{y}, \mathbf{z})$$

$$f(\mathbf{x}, a\mathbf{y} + b\mathbf{z}) = af(\mathbf{x}, \mathbf{y}) + bf(\mathbf{x}, \mathbf{z})$$

Definition

A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ *is called* **inner product on** V**,** *if*

- $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ *(symmetric)*
- $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ *and* $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ *iff* $\mathbf{x} = 0$ *(positive definite)*
- $\langle \cdot, \cdot \rangle$ *is a bilinear function*

Inner Products

Example

The scalar product/dot product in $\mathbb{R}^n$, *which is given by*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

The dot product is an inner product.
$\mathbb{R}^n$ *with the dot product is called* **Euclidean vector space.**

Inner Products

Example

The scalar product/dot product in $\mathbb{R}^n$, *which is given by*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

The dot product is an inner product.
$\mathbb{R}^n$ *with the dot product is called* **Euclidean vector space.**

Example

If we define for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$

$$\langle \mathbf{x}, \mathbf{y} \rangle := x_1 y_1 - (x_2 y_1 + x_1 y_2) + 2 x_2 y_2$$

then $\langle \cdot, \cdot \rangle$ *is an inner product but different from the dot product.*