

Basic Calculus Review

Lorenzo Rosasco

ISML Mod. 2 - Machine Learning

Vector Spaces

Functionals and Operators (Matrices)

Vector Space

- ▶ A **vector space** is a set V with binary operations

$$+ : V \times V \rightarrow V \quad \text{and} \quad \cdot : \mathbb{R} \times V \rightarrow V$$

such that for all $a, b \in \mathbb{R}$ and $v, w, x \in V$:

1. $v + w = w + v$
2. $(v + w) + x = v + (w + x)$
3. There exists $0 \in V$ such that $v + 0 = v$ for all $v \in V$
4. For every $v \in V$ there exists $-v \in V$ such that $v + (-v) = 0$
5. $a(bv) = (ab)v$
6. $1v = v$
7. $(a + b)v = av + bv$
8. $a(v + w) = av + aw$

Vector Space

- ▶ A **vector space** is a set V with binary operations

$$+ : V \times V \rightarrow V \quad \text{and} \quad \cdot : \mathbb{R} \times V \rightarrow V$$

such that for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$:

1. $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$
 2. $(\mathbf{v} + \mathbf{w}) + \mathbf{x} = \mathbf{v} + (\mathbf{w} + \mathbf{x})$
 3. There exists $0 \in V$ such that $\mathbf{v} + 0 = \mathbf{v}$ for all $\mathbf{v} \in V$
 4. For every $\mathbf{v} \in V$ there exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = 0$
 5. $\mathbf{a}(\mathbf{b}\mathbf{v}) = (\mathbf{a}\mathbf{b})\mathbf{v}$
 6. $1\mathbf{v} = \mathbf{v}$
 7. $(\mathbf{a} + \mathbf{b})\mathbf{v} = \mathbf{a}\mathbf{v} + \mathbf{b}\mathbf{v}$
 8. $\mathbf{a}(\mathbf{v} + \mathbf{w}) = \mathbf{a}\mathbf{v} + \mathbf{a}\mathbf{w}$
- ▶ Example: \mathbb{R}^n , space of polynomials, space of functions.

Inner Product

- ▶ An **inner product** is a function $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ such that for all $\alpha, \beta \in \mathbb{R}$ and $v, w, x \in V$:

Inner Product

- An **inner product** is a function $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ such that for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$:
1. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
 2. $\langle \mathbf{a}\mathbf{v} + \mathbf{b}\mathbf{w}, \mathbf{x} \rangle = \mathbf{a}\langle \mathbf{v}, \mathbf{x} \rangle + \mathbf{b}\langle \mathbf{w}, \mathbf{x} \rangle$
 3. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Inner Product

- ▶ An **inner product** is a function $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ such that for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$:
 1. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
 2. $\langle \mathbf{a}\mathbf{v} + \mathbf{b}\mathbf{w}, \mathbf{x} \rangle = \mathbf{a}\langle \mathbf{v}, \mathbf{x} \rangle + \mathbf{b}\langle \mathbf{w}, \mathbf{x} \rangle$
 3. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
- ▶ $\mathbf{v}, \mathbf{w} \in V$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Inner Product

- ▶ An **inner product** is a function $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ such that for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$:
 1. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
 2. $\langle \mathbf{a}\mathbf{v} + \mathbf{b}\mathbf{w}, \mathbf{x} \rangle = \mathbf{a}\langle \mathbf{v}, \mathbf{x} \rangle + \mathbf{b}\langle \mathbf{w}, \mathbf{x} \rangle$
 3. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.
- ▶ $\mathbf{v}, \mathbf{w} \in V$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.
- ▶ Given $W \subseteq V$, we have $V = W \oplus W^\perp$, where $W^\perp = \{ \mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in W \}$.

Inner Product

- ▶ An **inner product** is a function $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ such that for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$:
 1. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
 2. $\langle \mathbf{a}\mathbf{v} + \mathbf{b}\mathbf{w}, \mathbf{x} \rangle = \mathbf{a}\langle \mathbf{v}, \mathbf{x} \rangle + \mathbf{b}\langle \mathbf{w}, \mathbf{x} \rangle$
 3. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = 0$.
- ▶ $\mathbf{v}, \mathbf{w} \in V$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.
- ▶ Given $W \subseteq V$, we have $V = W \oplus W^\perp$, where $W^\perp = \{ \mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in W \}$.
- ▶ Cauchy-Schwarz inequality: $\langle \mathbf{v}, \mathbf{w} \rangle \leq \langle \mathbf{v}, \mathbf{v} \rangle^{1/2} \langle \mathbf{w}, \mathbf{w} \rangle^{1/2}$.

Norm

- ▶ Can define norm from inner product: $\|\mathbf{v}\| = \langle \mathbf{v}, \mathbf{v} \rangle^{1/2}$.

Norm

- ▶ A **norm** is a function $\| \cdot \|: V \rightarrow \mathbb{R}$ such that for all $\mathbf{a} \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in V$:
 1. $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$
 2. $\|\mathbf{a}\mathbf{v}\| = |\mathbf{a}| \|\mathbf{v}\|$
 3. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$
- ▶ Can define norm from inner product: $\|\mathbf{v}\| = \langle \mathbf{v}, \mathbf{v} \rangle^{1/2}$.

Metric

- ▶ Can define metric from norm: $d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$.

Metric

- ▶ A **metric** is a function $d: V \times V \rightarrow \mathbb{R}$ such that for all $v, w, x \in V$:
 1. $d(v, w) \geq 0$, and $d(v, w) = 0$ if and only if $v = w$
 2. $d(v, w) = d(w, v)$
 3. $d(v, w) \leq d(v, x) + d(x, w)$
- ▶ Can define metric from norm: $d(v, w) = \|v - w\|$.

Basis

- ▶ $B = \{v_1, \dots, v_n\}$ is a **basis** of V if every $v \in V$ can be uniquely decomposed as

$$v = a_1 v_1 + \dots + a_n v_n$$

for some $a_1, \dots, a_n \in \mathbb{R}$.

Basis

- ▶ $B = \{v_1, \dots, v_n\}$ is a **basis** of V if every $v \in V$ can be uniquely decomposed as

$$v = a_1 v_1 + \dots + a_n v_n$$

for some $a_1, \dots, a_n \in \mathbb{R}$.

- ▶ An orthonormal basis is a basis that is orthogonal ($\langle v_i, v_j \rangle = 0$ for $i \neq j$) and normalized ($\|v_i\| = 1$).

Vector Spaces

Functionals and Operators (Matrices)

Maps

Next we are going to review basic properties of maps on a Hilbert space.

- ▶ functionals: $\Psi : \mathcal{H} \rightarrow \mathbb{R}$
- ▶ linear operators $A : \mathcal{H} \rightarrow \mathcal{H}$, such that
 $A(\mathbf{a}f + \mathbf{b}g) = \mathbf{a}Af + \mathbf{b}Ag$, with $\mathbf{a}, \mathbf{b} \in \mathbb{R}$ and $f, g \in \mathcal{H}$.

Representation of Continuous Functionals

Let \mathcal{H} be a Hilbert space and $\mathbf{g} \in \mathcal{H}$, then

$$\Psi_{\mathbf{g}}(\mathbf{f}) = \langle \mathbf{f}, \mathbf{g} \rangle, \quad \mathbf{f} \in \mathcal{H}$$

is a continuous linear functional.

Riesz representation theorem

The theorem states that every continuous linear functional Ψ can be written uniquely in the form,

$$\Psi(\mathbf{f}) = \langle \mathbf{f}, \mathbf{g} \rangle$$

for some appropriate element $\mathbf{g} \in \mathcal{H}$.

Matrix

- ▶ Every linear operator $L: \mathbb{R}^m \rightarrow \mathbb{R}^n$ can be represented by an $m \times n$ matrix A .

Matrix

- ▶ Every linear operator $L: \mathbb{R}^m \rightarrow \mathbb{R}^n$ can be represented by an $m \times n$ matrix A .
- ▶ If $A \in \mathbb{R}^{m \times n}$, the transpose of A is $A^T \in \mathbb{R}^{n \times m}$ satisfying

$$\langle Ax, \mathbf{y} \rangle_{\mathbb{R}^m} = (Ax)^T \mathbf{y} = \mathbf{x}^T A^T \mathbf{y} = \langle \mathbf{x}, A^T \mathbf{y} \rangle_{\mathbb{R}^n}$$

for every $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$.

Matrix

- ▶ Every linear operator $L: \mathbb{R}^m \rightarrow \mathbb{R}^n$ can be represented by an $m \times n$ matrix A .
- ▶ If $A \in \mathbb{R}^{m \times n}$, the transpose of A is $A^T \in \mathbb{R}^{n \times m}$ satisfying

$$\langle Ax, \mathbf{y} \rangle_{\mathbb{R}^m} = (Ax)^T \mathbf{y} = \mathbf{x}^T A^T \mathbf{y} = \langle \mathbf{x}, A^T \mathbf{y} \rangle_{\mathbb{R}^n}$$

for every $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$.

- ▶ A is symmetric if $A^T = A$.

Eigenvalues and Eigenvectors

- ▶ Let $A \in \mathbb{R}^{n \times n}$. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector of A with corresponding eigenvalue $\lambda \in \mathbb{R}$ if $A\mathbf{v} = \lambda\mathbf{v}$.

Eigenvalues and Eigenvectors

- ▶ Let $A \in \mathbb{R}^{n \times n}$. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector of A with corresponding eigenvalue $\lambda \in \mathbb{R}$ if $A\mathbf{v} = \lambda\mathbf{v}$.
- ▶ Symmetric matrices have real eigenvalues.

Eigenvalues and Eigenvectors

- ▶ Let $A \in \mathbb{R}^{n \times n}$. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector of A with corresponding eigenvalue $\lambda \in \mathbb{R}$ if $A\mathbf{v} = \lambda\mathbf{v}$.
- ▶ Symmetric matrices have real eigenvalues.
- ▶ **Spectral Theorem:** Let A be a symmetric $n \times n$ matrix. Then there is an orthonormal basis of \mathbb{R}^n consisting of the eigenvectors of A .

Eigenvalues and Eigenvectors

- ▶ Let $A \in \mathbb{R}^{n \times n}$. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector of A with corresponding eigenvalue $\lambda \in \mathbb{R}$ if $A\mathbf{v} = \lambda\mathbf{v}$.
- ▶ Symmetric matrices have real eigenvalues.
- ▶ **Spectral Theorem:** Let A be a symmetric $n \times n$ matrix. Then there is an orthonormal basis of \mathbb{R}^n consisting of the eigenvectors of A .
- ▶ Eigendecomposition: $A = V\Lambda V^T$, or equivalently,

$$A = \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^T.$$

Singular Value Decomposition

- ▶ Every $A \in \mathbb{R}^{m \times n}$ can be written as

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{m \times m}$ is orthogonal, $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal, and $V \in \mathbb{R}^{n \times n}$ is orthogonal.

Singular Value Decomposition

- ▶ Every $A \in \mathbb{R}^{m \times n}$ can be written as

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{m \times m}$ is orthogonal, $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal, and $V \in \mathbb{R}^{n \times n}$ is orthogonal.

- ▶ Singular system:

$$\begin{aligned} Av_i &= \sigma_i u_i & AA^T u_i &= \sigma_i^2 u_i \\ A^T u_i &= \sigma_i v_i & A^T Av_i &= \sigma_i^2 v_i \end{aligned}$$

Matrix Norm

- ▶ The spectral norm of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is

$$\|\mathbf{A}\|_{\text{spec}} = \sigma_{\max}(\mathbf{A}) = \sqrt{\lambda_{\max}(\mathbf{A}\mathbf{A}^{\top})} = \sqrt{\lambda_{\max}(\mathbf{A}^{\top}\mathbf{A})}.$$

Matrix Norm

- ▶ The spectral norm of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is

$$\|\mathbf{A}\|_{\text{spec}} = \sigma_{\max}(\mathbf{A}) = \sqrt{\lambda_{\max}(\mathbf{A}\mathbf{A}^{\top})} = \sqrt{\lambda_{\max}(\mathbf{A}^{\top}\mathbf{A})}.$$

- ▶ The Frobenius norm of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is

$$\|\mathbf{A}\|_{\text{F}} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}.$$

Positive Definite Matrix

A real symmetric matrix $A \in \mathbb{R}^{m \times m}$ is positive definite if

$$\mathbf{x}^t A \mathbf{x} > 0, \quad \forall \mathbf{x} \in \mathbb{R}^m.$$

A positive definite matrix has positive eigenvalues.

Note: for positive semi-definite matrices $>$ is replaced by \geq .