CSE 840: Computational Foundations of Artificial Intelligence	Sep. 20, 2023

Lecture 7

Instructor: Vishnu Boddeti — *Scribe: Siqi Liang*

1 Inner Product and Hilbert Spaces

- metric: measures distances
- norm: measures distances, lengths
- product: measure distances, lengths, angles

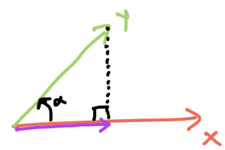

Figure 1: Inner product: $\langle x\ ,\ y\rangle = \|x\| \cdot \|y\| \cdot \cos\alpha$

In ML, we use cosine similarity: $\cos(\theta) = \frac{\langle x\ ,\ y\rangle}{\|x\| \cdot \|y\|}$.

Inner product $\leftrightarrow$ scalar product $\leftrightarrow$ dot product

Definition 1 (inner product) *Consider a vector space V. A mapping $\langle \cdot\ ,\ \cdot\rangle : V \times V \to F$ is called a inner product if*

(P1) (linearity) $\langle x_1 + x_2\ ,\ y\rangle = \langle x_1\ ,\ y\rangle + \langle x_2\ ,\ y\rangle$

(P2) (linearity) $\langle \lambda x\ ,\ y\rangle = \lambda\langle x\ ,\ y\rangle$, $(\lambda \in F)$

(P3) (symmetry) $\langle x\ ,\ y\rangle = \langle y\ ,\ x\rangle$ *(if F is $\mathbb{R}$);* $\langle x\ ,\ y\rangle = \overline{\langle y\ ,\ x\rangle}$ *(if F is $\mathbb{C}$)*

(P4) (positive definite) $\langle x\ ,\ x\rangle \geq 0$

(P5) (positive definite) $\langle x\ ,\ x\rangle = 0 \iff x = 0$

Examples:

- Euclidean inner product on $\mathbb{R}^n$:

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

  where $x = [x_1, \ldots, x_n]^\top$, and $y = [y_1, \ldots, y_n]^\top$
- On $\mathbb{C}$, $\langle x\ ,\ y\rangle = \sum_{i=1}^{n} x_i \bar{y}_i$
- $\mathcal{C}(\{[a, b]\})$: $\langle f\ ,\ g\rangle = \int_a^b f(t)g(t)dt$ is an inner product (but space would not be complete)

Definition 2 *A vector space with a norm is called normed space. If a normed space is complete (all Cauchy sequences converge), then V is called a Banach Space. A vector space with an inner product is called a pre-Hillbert space. If it is additionally complete, then V is called a Hilbert space.*

Given an inner product, we can define a norm space; while given a norm operation, we may not find a proper inner product.

Consider a vector space with an inner product $\langle \cdot\ ,\ \cdot \rangle$. Define $\| \cdot \| : V \to \mathbb{R}$ as $\|x\| := \sqrt{\langle x\ ,\ x \rangle}$. Then $\| \cdot \|$ is a norm on V, the norm is induced by $\langle \cdot\ ,\ \cdot \rangle$.

In general, the other way does not work.

Consider a vector space V with norm $\| \cdot \|$. Then $d : V \times V \to \mathbb{R}$, $d(x, y) := \|x - y\|$ is a metric on V, the metric is induced by the norm. In general, the other direction does not work.

inner product $\Rightarrow$ norm $\Rightarrow$ metric.

inner product $\not\Leftarrow$ norm $\not\Leftarrow$ metric.

2 Orthogonal Basis and Projections

Definition 3 (orthogonal) *Consider a pre-Hilbert-space V. Two vectors $v_1, v_2 \in V$ are called orthogonal if $\langle v_1, v_2 \rangle = 0$.*

Notaion: $v_1 \perp v_2$.

Two sets $V_1, V_2 \subset V$ are called orthogonal if $\forall v_1 \in V_1, v_2 \in V_2 : \langle v_1, v_2 \rangle = 0$.

Vectors are called orthonormal if additionally the two vectors have norm of 1 :

- $\langle v_1, v_2 \rangle = 0$
- $\|v_1\| = 1, \|v_2\| = 1$

A set of vectors $v_1, v_2, \ldots v_n$ is called orthonormal if any two vectors are orthonormal.

For a set $S \subseteq Y$ we define its orthogonal complement $S^\perp$ as follows:

$$S^\perp := \{v \in V \mid v \perp s, \forall s \in S\}$$

3 Orthogonal Projections

Definition 4 (projection) *$A \in \mathcal{L}(V)$ is called a projection if $A^2 = A$.*

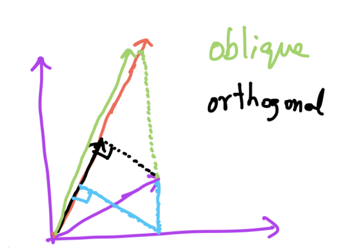

Figure 2: Oblique projection

Definition 5 *Let U be a finite-dim subspace of a pre-Hilbert-space H. Then there exists a linear projection $P_U : H \to U$, and $\ker(P_U) = U^{\perp}$. P_U is then called the orthogonal projection of H on U.*

Construction: Let $v_1, \dots v_n$ be an orthogonal basis of U. Define $P_U : V \to U$ by

$$P_u(\omega) = \sum_{i=1}^{n} \frac{\langle \omega, v_i \rangle}{\|v_i\|} v_i.$$

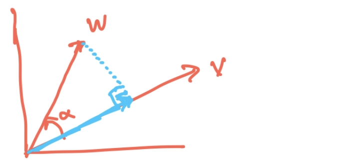

Figure 3: Orthogonal projection

Remark 6 *In an orthonormal basis $u_1 \dots u_n$ the representation of a vector v is given by*

$$v = \sum_{i=1}^{n} \langle v, u_i \rangle \, u_i.$$

Gram-Schmidt orthogonalization: It is a procedure that takes any basis $v_1 \dots v_n$ of a finite-dim vector space and transforms it into another basis $u_1 \dots u_n$ that is orthonormal.

Intrition: iterative procedure

Step 1. $u_1 = \frac{v_1}{\|v_1\|}$, $U_1 = \text{span}\{u_1\}$

Step k. Assume we already have $u_1, u_2, \dots, u_{k-1}$

- Project v_k on U_{k-1} and keep "the rest"

$$\tilde{u}_k = v_k - P_{v_{k-1}}(v_k)$$

- Renormalize:

$$u_k = \frac{\tilde{u}_k}{\|\tilde{u}_k\|}$$

In practice use Householder reflections for a numerically stable orthogonalization.

4 Orthogonal Matrices

Definition 7 *Let $Q \in \mathbb{R}^{n \times n}$ be a matrix with orthonormal column vectors (w.r.t. Euclidean inner product). Then Q is called an orthogonal matrix.*

Definition 8 *If $Q \in \mathbb{C}^{n \times n}$ and the columns are orthonormal (writ. the standard inner product on $\mathbb{C}$), then it is called unitary.*

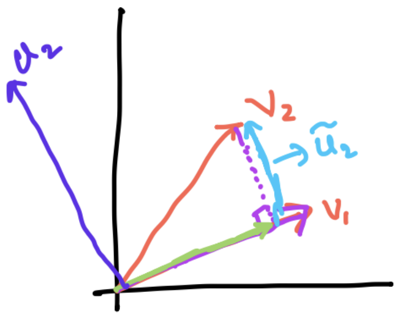

Figure 4: Gram-Schmidt orthogonalization

Examples:

- Identity: $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
- Reflection: $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, reflection about x-axis
- Permutation: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
- Rotation: $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
- Rotation in $\mathbb{R}^3$: rotate about one of the axes:

$$R_{\theta,1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

Properties of orthogonal matrix Q:

- columns are orthogonal $\Leftrightarrow$ rows are orthognd
- Q is always invertible, and $Q^{-1} = Q^\top$
- Q realizes an isometry : $\forall v \in V,\ \|Qv\| = \|v\| \longrightarrow$ keeps lengths intact
- Q preserves angles: $\langle Qu, Qv \rangle = \langle u, v \rangle\ \forall u, v \in V$
- $|\det Q| = 1$

The respective properties also hold for unitary matrices U. ($U^{-1} = \bar{U}^\top$)

Theorem 9 *Let $S \in \mathcal{L}(v)$ for a real vector spare V. Then the following are equivalent:*

- *S is an isometry: $\|Sv\| = \|v\|, v \in V$.*
- *There exists an orthonormal basis of V such that the matrix of S has the following form:*

$$M = \begin{pmatrix} \square & & & 0 \\ & \square & & \\ & & \square & \\ 0 & & & \square \end{pmatrix}$$

where each of the little block
 - *either a 1×1 matrix (one real number) with a value ± 1*

– *or a* 2×2 *rotation matrix*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$