Lecture 2: Hilbert Space, Matrix Decomposition

Karl Stratos

October 3, 2018

Overview

- Hilbert Space

  Reproducing Kernel Hilbert Space (RKHS)

- Matrix

  Eigendecomposition
  Singular Value Decomposition (SVD)

Hilbert Space

Hilbert space $(H, \langle \cdot, \cdot \rangle)$ is an inner product space whose norm $||u|| := \sqrt{\langle u, u \rangle}$ induces a complete metric space.

$$\lim_{\substack{n \to \infty \\ m \to \infty}} ||u_n - u_m|| = 0 \quad \Longrightarrow \quad \lim_{n \to \infty} ||u_n - u|| = 0 \text{ for some } u \in V$$

(Nontrivial) It can be verified that

- $\mathbb{R}^d$ is a Hilbert space under $\langle u, v \rangle := u \cdot v$.
- l^2 is a Hilbert space under $\langle u, v \rangle := \sum_{i=1}^{\infty} u_i v_i$.
- $L_w^2([a, b])$ is a Hilbert space under $\langle f, g \rangle := \int_a^b f(x)g(x)w(x)dx$ (Lebesgue, not Riemann)

We will always assume infinite dimension in the context of a Hilbert space.

New Definition of an Orthonormal Basis for Hilbert Space

An **orthonormal basis** of a Hilbert space $(H, \langle \cdot, \cdot \rangle)$ is a countably infinite set of orthonormal vectors $u_1, u_2, \ldots \in H$ such that every $u \in H$ can be uniquely written as

$$u = \sum_{i=1}^{\infty} \alpha_i u_i := \lim_{n \to \infty} \sum_{i=1}^{n} \alpha_i u_i$$

Convergence Result

If $u_1, u_2, \ldots \in H$ are orthonormal, then $\sum_{i=1}^{\infty} \alpha_i u_i$ converges iff $\alpha = (\alpha_1, \alpha_2, \ldots) \in l^2$.

Proof. If $\alpha \in l^2$, letting $s_n = \sum_{i=1}^{n} \alpha_i u_i$,

$$\left|\left|s_n - s_m\right|\right|^2 = \left|\left|\sum_{i=m+1}^{n} \alpha_i u_i\right|\right|^2 = \sum_{i=m+1}^{n} \left|\alpha_i\right|^2 \xrightarrow{n,m\to\infty} 0$$

thus s_n converges since H is complete.

$$\infty > \left|\left|u\right|\right|^2 = \left|\left|\lim_{n\to\infty} \sum_{i=1}^{n} \alpha_i u_i\right|\right|^2 = \lim_{n\to\infty} \left|\left|\sum_{i=1}^{n} \alpha_i u_i\right|\right|^2 = \lim_{n\to\infty} \sum_{i=1}^{n} \left|\alpha_i\right|^2$$

by the continuity of the norm $||\cdot||$, thus $\alpha \in l^2$.

New Definition of Linear Combination for Hilbert Space

An l^2 **linear combination** of orthonormal vectors $u_1, u_2, \ldots \in H$ is

$$u = \sum_{i=1}^{\infty} \alpha_i u_i \qquad\qquad (\alpha_1, \alpha_2, \ldots) \in l^2$$

which converges by the previous slide.

Checkable facts:

- (Coefficient formula) $\alpha_i = \langle u_i, u \rangle$.
- (Inner product formula) If $u = \sum_{i=1}^{\infty} \alpha_i u_i$ and $v = \sum_{i=1}^{\infty} \beta_i u_i$, then $\langle u, v \rangle = \sum_{i=1}^{\infty} \alpha_i \beta_i$.
- (Bessel's inequality) For any $x \in H$,

$$\sum_{i=1}^{\infty} \left| \langle x, u_i \rangle \right|^2 \leq \left|\left| x \right|\right|^2$$

Orthogonal Projection in Hilbert Space

- An l^2 **span** of orthonormal vectors $u_1, u_2, \ldots \in H$ is the set of all l^2 linear combinations:

$$S = \mathrm{span}\left(\{u_1, u_2, \ldots\}\right) = \left\{ \sum_{i=1}^{\infty} \alpha_i u_i : \; (\alpha_1, \alpha_2, \ldots) \in l^2 \right\}$$

- **Claim.** For any $u \in H$, the (unique) projection $u_S \in S$ of u onto S (i.e., $\langle u - u_S, v \rangle = 0$ for all $v \in S$) is given by

$$u_S = \sum_{i=1}^{\infty} \langle u, u_i \rangle \, u_i$$

  (Proof: show existence by Bessel's inequality, and use $\langle u, u_i \rangle = \langle u_S, u_i \rangle$ for all i.)

- **Hilbert projection theorem.**

$$u_S = \arg\min_{v \in S} ||u - v||$$

Characterization of Orthonormal Basis in Hilbert Space

Theorem. Orthonormal vectors $u_1, u_2, \ldots \in H$ are an orthonormal basis

1. Iff the set of their finite linear combinations is dense in H
2. Iff $0 = \langle x, u_1 \rangle = \langle x, u_2 \rangle = \cdots$ implies $x = 0$

Examples:

- $\{e_1, e_2, \ldots \ldots\}$ is an orthonormal basis for l^2 (use 2).
- The normalized **Fourier basis**

$$f_n(x) = \frac{1}{\sqrt{2\pi}} \exp(inx) \qquad\qquad \forall n \in \mathbb{Z}$$

  is an orthonormal basis for $L^2([-\pi, \pi])$ (use 1).

More Orthonormal Bases for L^2

- The normalized Legendre polynomials form an orthonormal basis for $L^2([-1,1])$ (use 1 and Weierstrass approximation theorem).

- The normalized **Hermite polynomials**

$$H_n(x) := \sqrt{n!} \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k}{k!} \frac{x^{n-2k}}{(n-2k)!2^k} \qquad \forall n = 0, 1, \ldots$$

  is an orthonormal basis for

$$L^2_{\mathcal{N}(0,1)}(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{R} : \ \mathbf{E}_{x \sim \mathcal{N}(0,1)} \left[\left| f(x) \right|^2 \right] < \infty \right\}$$

  Impication: give me any function $f : \mathbb{R} \to \mathbb{R}$ square-integrable under Gaussian measure, and I can write it as

$$f(x) = \sum_{n=0}^{\infty} \langle f, H_n(x) \rangle \, H_n(x)$$

Overview

- Hilbert Space

  Reproducing Kernel Hilbert Space (RKHS)

- Matrix

  Eigendecomposition
  Singular Value Decomposition (SVD)

Positive Definite Kernel

- For any nonempty set $\mathcal{X}$, a **positive definite (p.d.) kernel on** $\mathcal{X}$ is a symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that for any finite subset $x_1 \ldots x_n \in \mathcal{X}$,

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0 \qquad\qquad \forall c_1 \ldots c_n \in \mathbb{R}$$

- Any function $\phi : \mathcal{X} \to H$ induces a p.d. kernel on H defined by $k(x, y) = \langle \phi(x), \phi(y) \rangle$ since

$$\sum_{i,j=1}^{n} c_i c_j \langle \phi(x_i), \phi(y_j) \rangle = \left\langle \sum_{i=1}^{n} c_i \phi(x_i), \sum_{j=1}^{n} c_j \phi(x_j) \right\rangle \geq 0$$

Reproducing Kernel Hilbert Space (RKHS)

- RKHS is a Hilbert space of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ equipped with a (symmetric) **reproducing kernel** $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that
  - For each $x \in \mathcal{X}$, the function $k(\cdot, x) : \mathcal{X} \rightarrow \mathbb{R}$ is itself a member of H (canonical feature map), and
  - For each $f \in H$, we have the reproducing property $\langle f, k(\cdot, x) \rangle = f(x)$.
- In particular, it induces a "Hilbert space embedding" of $x \in \mathcal{X}$ by $\phi(x) := k(\cdot, x) \in H$ which satisfies

$$k(x, y) = \langle \phi(x), \phi(y) \rangle$$

- **Moore-Aronszajn.** Every p.d. kernel k is associated with a unique RKHS H.

Overview

- Hilbert Space

  Reproducing Kernel Hilbert Space (RKHS)

- Matrix

  Eigendecomposition
  Singular Value Decomposition (SVD)

Matrix

- A **linear transformation** $f : V \to W$ from vector space V to W is any function such that

$$f\left(\sum_i \alpha_i u_i\right) = \sum_i \alpha_i f\left(u_i\right)$$

- A matrix $A \in \mathbb{R}^{m \times n}$ is a linear transformation $u \mapsto Au$.
- (Exercise) Any linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be represented by $u \mapsto Au$ with a matrix $A \in \mathbb{R}^{m \times n}$. Thus

$$\mathbb{R}^{n \times m} = \{\text{all linear transformations from } \mathbb{R}^n \text{ to } \mathbb{R}^m\}$$

Vector Space of Matrices

- $(\mathbb{R}^{n \times m}, \langle \cdot, \cdot \rangle_F)$ is an inner product space where

$$\langle A, B \rangle_F := \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} A_{i,j} B_{i,j}}$$

- Induces the **Frobenius norm**

$$||A||_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} A_{i,j}^2} = \sqrt{\sum_{i=1}^{n} [A^\top A]_{i,i}} = \sqrt{\text{tr}\left(A^\top A\right)}$$

Matrix Subspaces

Any $A \in \mathbb{R}^{m \times n}$ is associated with the subspaces

$$\text{range}(A) := \{Au : \ u \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$$
$$\text{null}(A) := \{u \in \mathbb{R}^n : \ Au = 0\} \subseteq \mathbb{R}^n$$

with dimensions

$$\text{rank}(A) := \dim(\text{range}(A))$$
$$\text{nullity}(A) := \dim(\text{null}(A))$$

Rank-nullity theorem.

$$\text{rank}(A) + \text{nullity}(A) = n$$

To see why, monitor these quantities as you add columns to A.

A Spectacular Fact

$$\dim\left(\text{range}\left(A^\top\right)\right) = \dim\left(\text{range}\left(A\right)\right)$$

"Proof". Gaussian elimination $(E_T \dots E_1)A$ preserves $\text{range}\left(A^\top\right)$ and thus $\dim\left(\text{range}\left(A^\top\right)\right)$. It also preserves $\text{null}\left(A\right)$ and thus $\dim\left(\text{range}\left(A\right)\right)$ by the rank-nullity theorem. But it outputs

$$\begin{bmatrix} a & * & * & * & * & * & * & * & * \\ 0 & 0 & b & * & * & * & * & * & * \\ 0 & 0 & 0 & c & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & d & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & e \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Orthogonal Projection

If the columns of $U = [u_1 \dots u_m] \in \mathbb{R}^{n \times m}$ are orthonormal (i.e., $U^\top U = I_m$), the projection of $v \in \mathbb{R}^m$ onto $\mathrm{range}\,(U)$ is given by

$$UU^\top v = \sum_{i=1}^{n} (v^\top u_i) u_i$$

where $UU^\top \in \mathbb{R}^{n \times n}$ is a projection operator.

When $n = m$, $UU^\top$ is the projection onto $\mathbb{R}^n$ and thus I_n. In this case,

$$U^\top U = UU^\top = I_n$$

Such a matrix $U \in \mathbb{R}^{n \times n}$ is called an **orthogonal matrix**.

Overview

- Hilbert Space

  Reproducing Kernel Hilbert Space (RKHS)

- Matrix

  Eigendecomposition
  Singular Value Decomposition (SVD)

Definition

An **eigenvalue** λ of $A \in \mathbb{R}^{n \times n}$ is a scalar such that

$$Av = \lambda v$$

for some nonzero vector $v \in \mathbb{R}^m$. Any such $v \neq 0$ is an **eigenvector** associated with λ.

Existence of (Complex-Valued) Eigenvalues

λ is an eigenvalue of $A \in \mathbb{R}^{n \times n}$ iff there is some $v \neq 0$ such that

$$\begin{aligned}
(A - \lambda I_n)v = 0 \quad &\iff \quad \text{nullity}\,(A - \lambda I_n) > 0 \\
&\iff \quad \text{rank}\,(A - \lambda I_n) < n \\
&\iff \quad A - \lambda I_n \text{ is not invertible} \\
&\iff \quad \det(A - \lambda I_n) = 0 \\
&\iff \quad p_n(\lambda) = 0
\end{aligned}$$

By the fundamental theorem of algebra, any polynomial of degree n has n (possibly complex-valued) roots, counted with duplicates.

Relationship Between Eigenvalues, Trace, and Determinant

For $A \in \mathbb{R}^{2\times 2}$, the eigenvalues λ_1, λ_2 are the roots of

$$\det(A - \lambda I_n) = \det\left(\begin{bmatrix} a-\lambda & b \\ c-\lambda & d \end{bmatrix}\right) = \lambda^2 - \underbrace{(a+d)}_{\mathrm{tr}(A)}\lambda + \underbrace{(ad+bc)}_{\det(A)} = 0$$

On the other hand,

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2 = 0$$

In general,

$$\mathrm{tr}\,(A) = \sum_{i=1}^{n} \lambda_i$$

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

Spectral Theorem

Statement. If $A \in \mathbb{R}^{n \times n}$ is **symmetric**, we can find n real-valued sorted eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and corresponding orthonormal eigenvectors $v_1 \ldots v_n \in \mathbb{R}^n$.

Implication. Organizing $V = [v_1 \ldots v_n] \in \mathbb{R}^{n \times n}$ and $\Lambda = \text{diag}((\lambda_1 \ldots \lambda_n))$, we can write

$$A = V \Lambda V^\top$$

v_1 is called the top eigenvector; $v_1 \ldots v_k$ are called the top k eigenvectors.

Variational Characterization

Claim 1.

$$v_1 \in \operatorname*{arg\,max}_{v \in \mathbb{R}^n :\ ||v||=1} v^\top A v$$

Claim 2. For $i = 1 \ldots k$,

$$v_i \in \operatorname*{arg\,max}_{\substack{v \in \mathbb{R}^n :\ ||v||=1 \\ \langle v, v_j \rangle = 0\ \forall j < i}} v^\top A v$$

Overview

- Hilbert Space

  Reproducing Kernel Hilbert Space (RKHS)

- Matrix

  Eigendecomposition
  Singular Value Decomposition (SVD)

Definition

Given any $A \in \mathbb{R}^{n \times m}$, let

- $u_1 \ldots u_n \in \mathbb{R}^n$: orthonormal eigenvectors of $AA^\top \in \mathbb{R}^{n \times n}$ corresponding to top eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$
- $v_1 \ldots v_m \in \mathbb{R}^m$: orthonormal eigenvectors of $A^\top A \in \mathbb{R}^{m \times m}$ corresponding to top eigenvalues $\lambda'_1 \geq \ldots \geq \lambda'_m$

Fact. For $i = 1 \ldots \min(m, n)$,

$$\lambda_i = \lambda'_i \geq 0 \qquad A^\top u_i = \sqrt{\lambda_i} v_i \qquad A v_i = \sqrt{\lambda_i} u_i$$

We call u_i and v_i the **left** and **right singular vector** of A corresponding to i-th largest **singular value** $\sigma_i := \sqrt{\lambda_i} \geq 0$.

SVD

Any $A \in \mathbb{R}^{n \times m}$ (assume $n \geq m$) can be written as

$$A = U \Sigma V^\top = \sum_{i=1}^{m} \sigma_i u_i v_i^\top$$

where

$$\begin{aligned} U &\in \mathbb{R}^{n \times n} & U &= [u_1 \dots u_n] \\ V &\in \mathbb{R}^{m \times m} & V &= [v_1 \dots v_m] \\ \Sigma &\in \mathbb{R}^{n \times m} & \Sigma_{i,i} &= \sigma_i \end{aligned}$$

Furthermore, if $\mathrm{rank}\,(A) = r \leq m$, then $\sigma_i = 0$ for $i > r$ and

$$A = U_r \Sigma_r V_r^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$$

Generally $U_k \Sigma_k V_k^\top$ is called a rank-k SVD.

Variational Characterization

Claim 1. For $i = 1 \ldots k$,

$$(u_i, v_i) \in \operatorname*{arg\,max}_{\substack{(u,v) \in \mathbb{R}^n \times \mathbb{R}^m: \\ ||u|| = ||v|| = 1 \\ u^\top u_j = v^\top v_j = 0 \ \forall j < i}} u^\top A v$$

Claim 2. Letting $V_k := [v_1 \ldots v_k] \in \mathbb{R}^{m \times k}$,

$$V_k \in \operatorname*{arg\,max}_{W \in \mathbb{R}^{m \times k}: \ W^\top W = I_k} \underbrace{\mathrm{tr}\left(W^\top A^\top A W\right)}_{||AW||_F^2}$$

Application: Spectral Norm

$$||A||_2 := \max_{w \in \mathbb{R}^n:\ ||w||=1} ||Aw||$$

Frobenius/spectral norm of $A \in \mathbb{R}^{n \times m}$ $(n \geq m)$ in singular values:

$$||A||_F = \sqrt{\mathrm{tr}\left(A^\top A\right)} = \sqrt{\sum_{i=1}^{m} \lambda_i(A^\top A)} = \sqrt{\sum_{i=1}^{m} \sigma_i(A)^2}$$

$$||A||_2 = \sqrt{\max_{w \in \mathbb{R}^m:\ ||w||=1} w^\top A^\top A w} = \sqrt{\lambda_1(A^\top A)} = \sigma_1(A)$$

Application: Orthonormal Bases for Matrix Subspaces

For any $A \in \mathbb{R}^{n \times m}$ with rank $r \leq \min(n, m)$, if $U_r \in \mathbb{R}^{n \times r}$ and $U_{n-r} \in \mathbb{R}^{n \times (n-r)}$ are singular vectors corresponding to nonzero and zero singular values (likewise for V_r and V_{m-r}),

$$\begin{aligned}
\mathrm{range}\left(A\right) &= \mathrm{range}\left(U_r\right) \\
\mathrm{null}\left(A\right) &= \mathrm{range}\left(U_{n-r}\right) \\
\mathrm{range}\left(A^\top\right) &= \mathrm{range}\left(V_r\right) \\
\mathrm{null}\left(A^\top\right) &= \mathrm{range}\left(V_{m-r}\right)
\end{aligned}$$

Application: Best Low-Rank Approximation

For any $A \in \mathbb{R}^{n \times m}$ with rank-k SVD $U_k \Sigma_k V_k$,

$$U_k \Sigma_k V_k \in \operatorname*{arg\,min}_{Z \in \mathbb{R}^{n \times m}:\ \mathrm{rank}(Z) \leq k} ||A - Z||$$

where $||\cdot||$ is Frobenius or Spectral (or any orthogonally invariant norm).

Application: Pseudoinverse

Pseudoinverse of a matrix $A \in \mathbb{R}^{n \times m}$ is the unique matrix $A^+ \in \mathbb{R}^{m \times n}$ such that

1. $AA^+ \in \mathbb{R}^{n \times n}$ is the orthogonal projection onto $\mathrm{range}\,(A)$, and
2. $A^+A \in \mathbb{R}^{m \times m}$ is the orthogonal projection onto $\mathrm{range}\left(A^\top\right)$.

Proposition

Let $A \in \mathbb{R}^{m \times n}$ with $r := \mathrm{rank}\,(A) \leq \min\{m, n\}$. Let $A = U\Sigma V^\top$ denote a rank-r SVD of A. Then

$$A^+ = V\Sigma^{-1}U^\top$$

Application: Best-Fit Subspace

Problem. Given N data points in $\mathbb{R}^d$, identify a k-dimensional subspace such that their projection onto the subspace is closest to the original points.

Solution.

$$U_k \in \operatorname*{arg\,min}_{W \in \mathbb{R}^{d \times k} : W^\top W = I_k} \left|\left| X - WW^\top X \right|\right|_F$$

The projected N points are given by $Y = WW^\top X$.

Relationship with Eigendecomposition I

- If $A \in \mathbb{R}^{n \times n}$ is symmetric with eigendecomposition $A = \overline{V}\text{diag}(\lambda_1 \ldots \lambda_n)\overline{V}^\top$ and SVD $A = U\text{diag}(\sigma_1 \ldots \sigma_n)V^\top$, then for some permutation over columns π

$$\Sigma \overset{\pi}{=} |\Lambda| \qquad\qquad U = V \overset{\pi}{=} \overline{V}$$

- Corollary: SVD and eigendecomposition coincide on symmetric positive semi-definite (i.e., only has nonnegative eigenvalues) matrices.

Relationship with Eigendecomposition II

Let $A \in \mathbb{R}^{n \times m}$ (assume $n \geq m$) with an SVD $A = [U_1 U_2][\Sigma_m; 0_{(n-m)\times m}]V^\top$ where $\Sigma_m = \text{diag}(\sigma_1 \ldots \sigma_m)$. Define an $(n+m) \times (n+m)$ symmetric matrix with eigendecomposition with $W, \Lambda \in \mathbb{R}^{(n+m)\times(n+m)}$

$$\widetilde{A} := \begin{bmatrix} 0_{n\times n} & A \\ A^\top & 0_{m\times m} \end{bmatrix} = W \Lambda W^\top$$

Then up to different signs on U_1, U_2, V (column-wise),

$$W = \begin{bmatrix} U_1/\sqrt{2} & U_2 & -\underline{U_1}/\sqrt{2} \\ V/\sqrt{2} & 0_{m\times(n-m)} & \underline{V}/\sqrt{2} \end{bmatrix}$$
$$\Lambda = \text{diag}\left(\Sigma_m, 0_{(n-m)\times(n-m)}, -\underline{\Sigma}_m, \right)$$

where $\underline{M}$ indicates matrix M with reverse column ordering. In particular, the ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{n+m}$ of $\widetilde{A}$ are

$$\sigma_1 \geq \ldots \geq \sigma_m \geq \underbrace{0 \geq \cdots \geq 0}_{n-m} \geq -\sigma_m \geq \ldots \geq -\sigma_1$$