SYMMETRIC MATRICES

Dr. Matthieu R Bloch

Wednesday October 27, 2021

LOGISTICS

Drop date: October 30, 2021

My office hours on Tuesdays

- 8am-9am on BlueJeans (https://bluejeans.com/205357142)
- This was great!

Midterm 2:

- Moved to **Monday November 8, 2021** (gives you weekend to prepare)
- Coverage: everything since Midterm 1 (dont' forget the fundamentals though), emphasis on **regression**

WHAT'S ON THE AGENDA FOR TODAY?

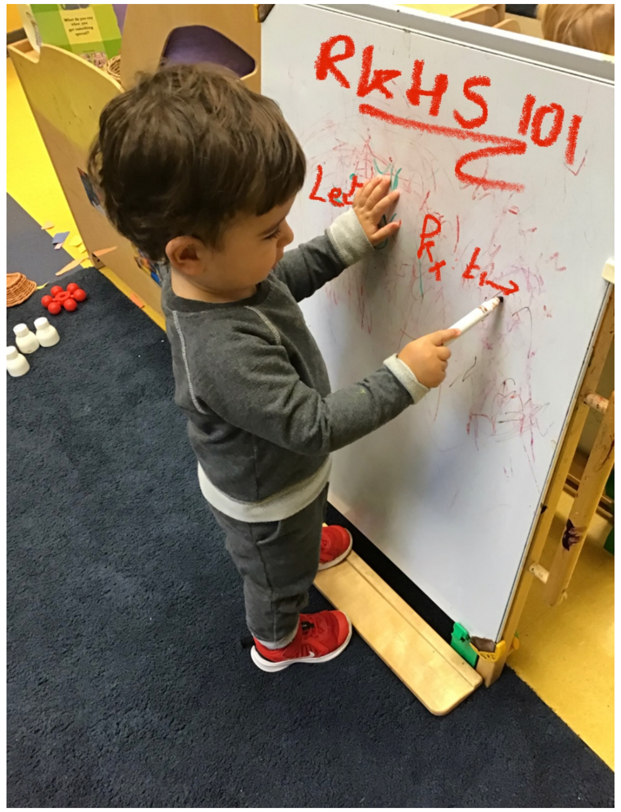
Last time:

- RKHS

Today:

- Final thoughts on RKHS
- Symmetric matrices: more linear algebra (*objective:* further understand least-square problems)

Reading: Romberg, lecture notes 10/11/12



Toddlers can do it!

LAST TIME: ARONSZJAN’S THEOREM

Definition. (Inner product kernel)

An *inner product kernel* is a mapping $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ for which there exists a Hilbert space $\mathcal{H}$ and a mapping $\Phi : \mathbb{R}^d \to \mathcal{H}$ such that

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d \quad k(\mathbf{u}, \mathbf{v}) = \langle \Phi(\mathbf{u}), \Phi(\mathbf{v}) \rangle_{\mathcal{H}}$$

Definition. (Positive semidefinite kernel)

A function $\boxed{k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}}$ is a *positive semidefinite* kernel if

- k is symmetric, i.e., $k(\mathbf{u}, \mathbf{v}) = k(\mathbf{v}, \mathbf{u})$
- for all $\{\mathbf{x}_i\}_{i=1}^N$, the *Gram matrix* $\mathbf{K}$ is positive semidefinite, i.e.,

$$\mathbf{x}^\intercal \mathbf{K} \mathbf{x} \geq 0 \text{ with } \mathbf{K} = [K_{i,j}] \text{ and } \boxed{K_{i,j} \triangleq k(\mathbf{x}_i, \mathbf{x}_j)}$$

Theorem.

A function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is an inner product kernel if and only if k is a positive semidefinite kernel.

Sketch of proof. ① Let k be an inner product kernel (e.g., the kernel of an RKHS)

Then k is symmetric (by definition)

Let $\{x_i\}_{i=1}^n \in \mathbb{R}^d$, form the matrix $K = [k(x_i, x_j)]_{i,j}$

Let $w \in \mathbb{R}^n$

$$w^T K w = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \underbrace{k(x_i, x_j)}_{\langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}}}$$

$$= \left\langle \sum_{i=1}^n w_i \Phi(x_i), \sum_{j=1}^n w_j \Phi(x_j) \right\rangle_{\mathcal{H}}$$

$$= \left\| \sum_{i=1}^n w_i \Phi(x_i) \right\|_{\mathcal{H}}^2 \geq 0$$

② Let k be a SDP kernel

this is a function

Define $\Phi : \mathbb{R}^d \longrightarrow \mathcal{H} : \underline{x} \longmapsto k_{\underline{x}} \triangleq k(\underline{x}, \cdot)$

this is the variable

Let $\mathcal{H}_0 = \operatorname{Span}\{\Phi(\underline{x}) : \underline{x} \in \mathbb{R}^d\}$

Let $f, g \in \mathcal{H}_0$; then $f = \sum_{i=1}^{k} \alpha_i \Phi(x_i)$ and $g = \sum_{j=1}^{\ell} \beta_j \Phi(x_j)$

Define $\langle \widetilde{f, g} \rangle \triangleq \sum_{i=1}^{k} \sum_{j=1}^{\ell} \alpha_i \beta_j k(x_i, x_j)$

Note $\langle \widetilde{f, g} \rangle = \langle \widetilde{g, f} \rangle$ so the candidate inner product is symmetric; it is also bilinear (by definition)

Fact. a symmetric bilinear form satisfies Cauchy-Schwartz : $\langle \widetilde{f, g} \rangle \leq \sqrt{\langle \widetilde{f, f} \rangle} \sqrt{\langle \widetilde{g, g} \rangle}$

b/c $\langle \widetilde{f, f} \rangle = \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) = \alpha^T K \alpha \geqslant 0$

Note $\forall f \in \mathcal{H}_0 \ \forall x \in \mathbb{R}^d \ \langle \widetilde{f, k_x} \rangle = \langle \widetilde{\sum_{i=1}^{n} \alpha_i k_{x_i}, k_x} \rangle = \sum_{i=1}^{n} \alpha_i k(x_i, x) = f(x)$ (reproducing pro

Finally $|f(x)| = |\langle \widetilde{f, k_x} \rangle| \leq \sqrt{\langle \widetilde{f, f} \rangle} \underbrace{\sqrt{\langle \widetilde{k_x, k_x} \rangle}}_{= [k(x, x)]^{1/2}}$

Hence $\langle \widetilde{f, f} \rangle = 0 \Rightarrow f = 0$

Hence $\breve{\langle \cdot, \cdot \rangle}$ is a legitimate inner product.

$\mathcal{H}_0$ + $\langle \cdot ; \cdot \rangle$ is a pre-Hilbert space.

We need to extend $\mathcal{H}_0$ to make it complete

<u>Solution</u> : $\mathcal{H} = \{$ set of functions that are pointwise Cauchy limits of Cauchy sequences in $\mathcal{H}_0 \}$ (see notes)

EXAMPLES OF RKHS

Example. Regression using linear and quadratic functions in \(\bbR^d\)

Introduce a mapping $\Psi : \mathbb{R}^d \longrightarrow \mathbb{R}^n : \underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \longmapsto \Psi(\underline{x}) = \begin{pmatrix} 1 \\ \sqrt{2}\,x_1 \\ \vdots \\ \sqrt{2}\,x_d \\ x_1^2 \\ \vdots \\ x_d^2 \\ \sqrt{2}\,x_1 x_2 \\ \vdots \\ \sqrt{2}\,x_{d-1} x_d \end{pmatrix} \in \mathbb{R}^n$ $\quad n = 1 + d + d + \frac{d(d-1)}{2} \sim d^2$

$$\forall x, y \in \mathbb{R}^d \quad \langle \Psi(x), \Psi(y) \rangle_{\mathbb{R}^n} = 1 + 2\sum_{i=1}^{d} x_i y_i + \sum_{i=1}^{d} x_i^2 y_i^2 + 2 \sum_{i} \sum_{j \neq i} x_i x_j y_i y_j = \left(1 + \sum_{i=1}^{d} x_i y_i\right)^2 = \underbrace{\left(1 + \langle x, y \rangle_{\mathbb{R}^d}\right)^2}_{\text{kernel}}$$

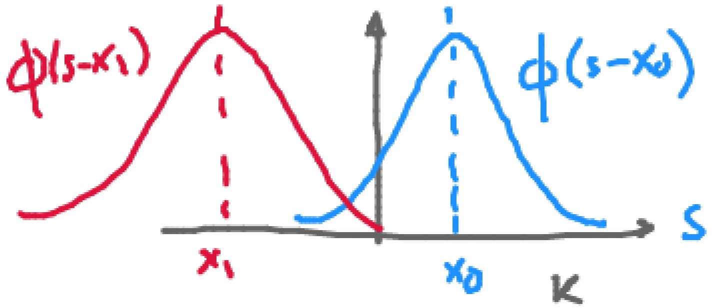
$k(x,y) \overset{\Delta}{=} (1 + \langle x, y \rangle$

EXAMPLES OF RKHS

Example. Regression using linear and quadratic functions in \(\bbR^d\)

Example. Regression using Radial Basis Functions

Consider the mapping $\Phi : \mathbb{R}^d \to \mathcal{F} : \underline{x} \longmapsto \phi(\underline{s}-\underline{x})$ w/ $\phi(\underline{s}) = \left(\frac{1}{\sigma\sqrt{\pi}}\right)^{d/2} \exp\left(-\frac{\|s\|_2^2}{2\sigma^2}\right)$



$x_k \in \mathcal{S}$ bounded in $\mathbb{R}^d$

Consider $\mathcal{F} \triangleq \{$all functions of the form $\sum_{k=1}^{K} \beta_k \phi(\underline{s}-x_k)$ for $K \in \mathbb{N}$, $\{\beta_k\}$, $\{\underline{x}_k\}\}$

Let $f, g \in \mathcal{F}$, i.e., $f(s) = \sum_{k=1}^{K} \beta_k \phi(\underline{s}-\underline{x}_k)$ and $g(s) = \sum_{l=1}^{L} \alpha_l \phi(\underline{s}-y_l)$

Since $\mathcal{F} \subset L_2(\mathbb{R})$, we use $\langle f, g\rangle = \sum_k \sum_l \beta_k \alpha_l \underbrace{\langle \phi(\underline{s}-x_k), \phi(\underline{s}-y_l)\rangle}_{K_{kl}} = \alpha^T K \beta$ $\quad \alpha = \begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_L\end{pmatrix}$ $\beta = \begin{pmatrix}\beta_1\\ \vdots\\ \beta_K\end{pmatrix}$

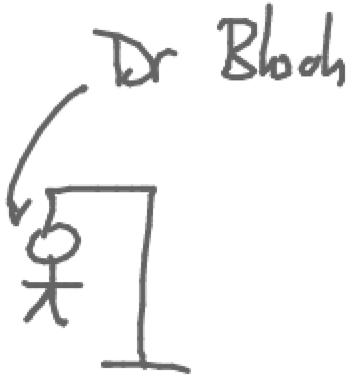
$$K_{kl} \overset{\Delta}{=} \langle \phi(\underline{s}-\underline{x}_k), \phi(\underline{s}-\underline{y}_l)\rangle = \int_{\underline{s}\in\mathbb{R}^d} \phi(\underline{s}-x_k)\,\phi(\underline{s}-\underline{x}_l)\, d\underline{s}$$

$$= \left(\frac{1}{\sigma\sqrt{\pi}}\right)^d \int_{\underline{s}\in\mathbb{R}^d} \underbrace{\exp\left(-\frac{\|\underline{s}-x_k\|_2^2}{2\sigma^2}\right)\exp\left(-\frac{\|\underline{s}-x_l\|_2^2}{2\sigma^2}\right)}_{\exp\left(-\frac{1}{2\sigma^2}\left(\|\underline{s}-x_k\|^2+\|s-x_l\|_2^2\right)\right)} d\underline{s} = \exp\left(-\frac{\|x_k-x_l\|_2^2}{4\sigma^2}\right)$$

Note $\|\underline{s}-x_k\|_2^2 + \|\underline{s}-\underline{x}_l\|_2^2 = \underline{s}^T\underline{s} - 2s^Tx_k + x_k^Tx_k + s^Ts - 2s^Tx_l + x_l^Tx_l$

$$= 2s^Ts - 4s^T\left(\frac{x_k+x_l}{2}\right) + 2\left(\frac{x_k+x_l}{2}\right)^T\left(\frac{x_k+x_l}{2}\right) - x_k^Tx_k - x_l^Tx_l - 4x_kx_l^T$$

$$= \underbrace{2\|\underline{s}-(x_k+x_l)\|_2^2} + \|x_k-x_l\|_2^2$$



EXAMPLES OF RKHS

> **Example.** Regression using linear and quadratic functions in $\mathbb{R}^d$

> **Example.** Regression using Radial Basis Functions

Examples of kernels

- Homogeneous polynomial kernel: $k(\mathbf{u}, \mathbf{v}) = (\mathbf{u}^\intercal \mathbf{v})^m$ with $m \in \mathbb{N}^*$
- Inhomogenous polynomial kernel: $k(\mathbf{u}, \mathbf{v}) = (\mathbf{u}^\intercal \mathbf{v} + c)^m$ with $c > 0, m \in \mathbb{N}^*$
- Radial basis function (RBF) kernel: $k(\mathbf{u}, \mathbf{v}) = \exp\left(-\frac{\|\mathbf{u}-\mathbf{v}\|^2}{2\sigma^2}\right)$ with $\sigma^2 > 0$

SYSTEMS OF SYMMETRIC EQUATIONS

Least square problems involved the normal equations $\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\theta} = \mathbf{X}^\intercal\mathbf{y}$

This is a system of symmetric equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ with $\mathbf{A}^\intercal = \mathbf{A}$

- Ultimately we will talk about the non-symmetric/non square case

> **Definition.**
>
> A real-valued matrix $\mathbf{A}$ is symmetric if $\mathbf{A}^\intercal = \mathbf{A}$ A complex-valued matrix $\mathbf{A}$ is Hermitian if $\mathbf{A}^\dagger = \mathbf{A}$ (also written $\mathbf{A}^H = \mathbf{A}$)

> **Definition.**
>
> Given a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, if a vector $\mathbf{v} \in \mathbb{C}^n$ satisfies $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{C}$, then λ is an *eigenvalue* associated to the *eigenvector* $\mathbf{v}$.

If λ is an eigenvalue, there are infinitely many eigenvectors associated to it

> **Definition.**
>
> Given a matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, if a vector $\mathbf{v} \in \mathbb{C}^n$ satisfies $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{C}$, then λ is an *eigenvalue* associated to the *eigenvector* $\mathbf{v}$.

CHANGE OF BASIS

Consider the canonical basis $\{e_i\}_{i=1}^n$ for $\mathbb{R}^n$; every vector can be viewed as a vector of coefficients $\{\alpha_i\}_{i=1}^n$,

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i e_i = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix}^\intercal$$

How do we find the representation of $\mathbf{x}$ in another basis $\{v_i\}_{i=1}^n$? Write $e_i = \sum_{j=1}^n \beta_{ij} v_j$

Regroup the coefficients

$$\mathbf{x} = \cdots + \left(\sum_{i=1}^{n} \beta_{ij} \alpha_i \right) v_j + \cdots$$

In matrix form

$$\mathbf{x}_{\text{new}} = \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{n1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{1n} & \beta_{2n} & \cdots & \beta_{nn} \end{bmatrix} \mathbf{x}$$

SIMILARITY

A change of basis matrix $\mathbf{P}$ is full rank (basis vectors are linearly independent)

Any full rank matrix $\mathbf{P}$ can be viewed as a change of basis

$\mathbf{P}^{-1}$ takes you back to the original basis

Warning: the columns of $\mathbf{P}$ describe the *old* coordinates as a function of the *new* ones

> **Definition.**
>
> If $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ then $\mathbf{B}$ is similar to $\mathbf{A}$ if there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$

Intuition: similar matrices are the same up to a change of basis

> **Definition.**
>
> $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ with $\mathbf{D}$ diagonal

Not all matrices are diagonalizable!

SPECTRAL THEOREM

Lemma (Existence of eigenvector) Every complex matrix $\mathbf{A}$ has at least one complex eigenvector and every real symmetrix matrix has real eigenvalues and at least one real eigenvector.

Lemma (Schur triangularization lemma) Every matrix $\mathbf{A}^{\in}\mathbb{R}^{n\times n}$ is unitarily similar to an upper triangular matrix, i.e.,

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^{\dagger}$$

with $\boldsymbol{\Delta}$ upper triangular and $\mathbf{V}^{\dagger} = \mathbf{V}^{-1}$

Theorem (Spectral theorem) Every hermitian matrix is unitarily similar to a real-valued diagonal matrix.