REPRODUCING KERNEL HILBERT SPACES

Dr. Matthieu R Bloch

Monday, October 18, 2021

LOGISTICS

Grades upcoming

- Midterm 1 99% graded - Grades announced after curving (don’t panic)
- Assignment 2 solution underway
- Assignment 3 graded
- Assignment 4 grading started
- Drop date: check!

More office hours

- Tuesday October 19, 2021 8am-9am on BlueJeans (https://bluejeans.com/205357142)
- Come prepared!

Midterm 2: scheduled for Wednesday November 3, 2021

- 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: Representer theorem

- Some infinite dimensional regression problems have surprising solutions!
- We can compute the solution as (finite) linear combination of feature vectors
- (We have to wrap up the proof)

Today:

- Reproducing Kernel Hilbert Spaces
- Justifies the kind of Hilbert spaces where regularized regression can be solved

Reading: Romberg, lecture notes 10

LEAST-SQUARES IN INFINITE DIMENSION HILBERT SPACES

In $\mathbb{R}^d$, the problem $\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \|\mathbf{y}-\mathbf{X}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_2^2$ has a solution

$$\boldsymbol{\theta}^* = \mathbf{X}^\intercal\boldsymbol{\alpha} \text{ with } \boldsymbol{\alpha} = (\mathbf{X}\mathbf{X}^\intercal + \lambda\mathbf{I})^{-1}\mathbf{y}$$

$\mathbf{X}\mathbf{X}^\intercal \in \mathbb{R}^{n\times n}$ is dimension independent! We will be able to extend this to infinite dimensional Hilbert spaces!

Let $\mathcal{F}$ be a Hilbert space and let $f\in\mathcal{F}$ be the function we are trying to estimate

- We will estimate $f\in\mathcal{F}$ using noisy observations $\langle f, x_i\rangle$ with $\{x_i\}_{i=1}^n$ elements of $\mathcal{F}$
- This is the equivalent of saying $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ in finite dimension

Proposition (Representer theorem)

$$\min_{f\in\mathcal{F}} \sum_{i=1}^n \left|y_i - \langle f, x_i\rangle_{\mathcal{H}}\right|^2 + \lambda\|f\|_{\mathcal{H}}$$

has solution

$$f = \sum_{i=1}^n \alpha_i x_i \text{ with } \boldsymbol{\alpha} = (\mathbf{K}+\lambda\mathbf{I})^{-1}\mathbf{y} \qquad \mathbf{K} = [\langle x_i, x_j\rangle]_{1\leq i,j\leq n}$$

Proof: last time we proved that wlog we can assume $f \in \mathrm{Span}(\{x_i\}_{i=1}^n)$

let $\hat{f}$ be our minimizer, we can write $\hat{f} = \sum_{i=1}^n \hat{\alpha}_i x_i \quad \{\hat{\alpha}_i\} \in \mathbb{R} \quad \{x_i\}_{i=1}^n \in \mathcal{H}$

Note: let $u, v \in \mathrm{Span}(\{x_i\}_{i=1}^n)$ then $u = \sum_{i=1}^n c_i x_i$ ($c_i \in \mathbb{R}$) $\quad v = \sum_{i=1}^n d_i x_i$ ($d_i \in \mathbb{R}$)

$$\langle u, v \rangle_{\mathcal{H}} = \sum_{i=1}^n \sum_{j=1}^n c_i d_j \underbrace{\langle x_i, x_j \rangle}_{G_{ij} \text{ component of Gram matrix for } \{x_i\}_{i=1}^n} = d^T G c \quad \text{w/ } d \triangleq (d_1 \dots d_n)^T \text{ and } c \triangleq (c_1 \dots c_n)^T$$

$$\min_{f \in \mathcal{H}} \sum_{i=1}^n |y_i - \langle f, x_i \rangle|^2 + \lambda \|f\|_{\mathcal{H}}^2 = \min_{\alpha \in \mathbb{R}^n} \sum_{i=1}^n \Big| y_i - \underbrace{\sum_{j=1}^n \alpha_j \langle x_j, x_i \rangle}_{[G\alpha]_i} \Big|^2 + \lambda \alpha^T G \alpha \quad (*)$$

$\alpha = (\alpha_1 \cdots \alpha_n)^T$

note that
$$G\alpha = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_n \rangle \\ & & \vdots & \\ \langle x_n, x_1 \rangle & \cdots & \cdots & \langle x_n, x_n \rangle \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n \alpha_j \langle x_1, x_j \rangle = \sum_{j=1}^n \alpha_j \langle x_j, x_1 \rangle \\ \vdots \end{bmatrix}$$

$$(*) = \min_{\alpha \in \mathbb{R}^n} \|y - G\alpha\|_2^2 + \lambda \alpha^T G \alpha$$

THE BIG PICTURE

We want to solve the finite dimensional problem

$$\min_{\alpha \in \mathbb{R}^n} \|y - G\alpha\|_2^2 + \lambda \alpha^T G \alpha \quad (*)$$

Taking the gradient of (*) and setting to 0 we obtain

True b/c G is symmetric

$\nabla(x^T A x) = (A + A$

$$\nabla_\alpha \left(y^T y - 2 y^T G \alpha + \alpha^T G^T G \alpha + \lambda \alpha^T G \alpha \right) = 0 \iff -2 G^T y + 2 G^T G \alpha + 2 \lambda G \alpha = 0$$

$$\iff G^T (G\alpha - y) + \lambda G \alpha = 0$$

Assume we pick α such that $\underbrace{(G + \lambda I)}_{\text{invertible}} \alpha = y$ $\quad$ [$(G + \lambda I)^{-1} y$ would work]

THE BIG PICTURE

For a Hilbert space $\calF$ and n pairs $(x_i,y_i)\in\calF\times \bbR$, we know how to solve the following problem with linear algebra

$$\min_{f\in\calF}\sum_{i=1}^n\abs{y_i-{\dotp{f}{x_i}}_{\calF}}^2+\lambda\norm[\calF]{f}$$

Set $\boxed{\alpha = (G+\lambda I)^{-1} y}$

$$G^T(G\alpha - y) + \lambda G\alpha = G^T(G(G+\lambda I)^{-1}y - y) + \lambda G(G+\lambda I)^{-1}y$$

$$= G^T(G(G+\lambda I)^{-1} - I)y + \lambda G(G+\lambda I)^{-1}y$$

$$= G^T(G(G+\lambda I)^{-1} - (G+\lambda I)(G+\lambda I)^{-1})y + \lambda G(G+\lambda I)^{-1}y$$

$$= [G^TG - G^T(G+\lambda I) + \lambda G](G+\lambda I)^{-1}y$$

$$= \underbrace{[G^TG - G^TG - \lambda G^T + \lambda G]}_{=0}(G+\lambda I)^{-1}y \quad \checkmark$$

THE BIG PICTURE

For a Hilbert space $\mathcal{F}$ and n pairs $(x_i, y_i) \in \mathcal{F} \times \mathbb{R}$, we know how to solve the following problem with linear algebra

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \left| y_i - \langle f, x_i \rangle_{\mathcal{F}} \right|^2 + \lambda \|f\|_{\mathcal{F}}$$

We would really like to solve the following problem for n pairs $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \left| y_i - f(\mathbf{x}_i) \right|^2 + \lambda \|f\|_{\mathcal{F}}$$

The question whether $f(\mathbf{x}_i) = \langle f, x_i \rangle_{\mathcal{F}}$ for some $x_i \in \mathcal{F}$ function of $\mathbf{x}_i$

- Can this be done?
- We can *choose* what $\mathcal{F}$ is!

Reproducing Kernel Hilbert Spaces (RKHSs) are specific Hilbert spaces where this happens to be true

- Specifcally, this is a Hilbert space of functions in whih the *sampling linear operation is a continuous linear functional*

As usual, we’re throwing definitions at out problem to make progress

LINEAR FUNCTIONS ON HILBERT SPACES

In what follows, $\mathcal{F}$ is a Hilbert space with scalar field $\mathbb{R}$

Definition.

A functional $F : \mathcal{F} \to \mathbb{R}$ associates real-valued number to an element of a Hilbert space $\mathcal{F}$

Notation can be tricky when the Hilbert space is a space of functions: F can act on a function $f \in \mathcal{F}$

Examples

Definition.

A functional $F : \mathcal{F} \to \mathbb{R}$ is continuous if

$$\forall \epsilon > 0 \exists \delta > 0 \text{ such that } \|x - y\|_{\mathcal{F}} \leq \delta \Rightarrow |F(x) - F(y)| \leq \epsilon.$$

Proposition. All norms are continuous functionals $F : \mathcal{F} \to \mathbb{R} : x \mapsto \langle x, c \rangle$ for some $c \in \mathcal{F}$ is continuous

Definition.
A functional F is linear if $\forall a, b \in \mathbb{R}\, \forall x, y \in \mathcal{F}\; F(ax + by) = aF(x) + bF(y)$.

REPRESENTATION OF (CONTINUOUS) LINEAR FUNCTIONALS

Proposition.

Let $F : \mathcal{F} \to \mathbb{R}$ be a linear functional on an n-dimensional Hilbert space $\mathcal{F}$.

Then there exists $c \in \mathcal{F}$ such that $F(x) = \langle x, c \rangle$ for every $x \in \mathcal{F}$

Linear functional over finite dimensional Hilbert spaces are continuous!

This is *not* true in infinite dimension

Theorem (Riesz representation theorem)

Let $F : \mathcal{F} \to \mathbb{R}$ be a *continuous* linear functional on a (possible infinite dimensional) separable Hilbert space $\mathcal{F}$.

Then there exists $c \in \mathcal{F}$ such that $F(x) = \langle x, c \rangle$ for every $x \in \mathcal{F}$

Proposition. If $\{\psi_n\}_{n \geq 1}$ is an orthobasis for $\mathcal{H}$, then we can construct c above as

$$c \triangleq \sum_{n=1}^{\infty} F(\psi_n)\psi_n$$

REPRODUCING KERNEL HILBERT SPACES

Definition. (Reproducing Kernel Hilbert Spaces)

An RKHS is a Hilbert space $\mathcal{H}$ of real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$ in which the sampling operation $\mathcal{S}_{\boldsymbol{\tau}} : \mathcal{H} \to \mathbb{R} : f \mapsto f(\boldsymbol{\tau})$ is continuous for every $\boldsymbol{\tau} \in \mathbb{R}^d$.

In other words, for each $\boldsymbol{\tau} \in \mathbb{R}^d$, there exists $k_{\boldsymbol{\tau}} \in \mathcal{H}$ s.t.

$$f(\boldsymbol{\tau}) = \langle f, k_{\boldsymbol{\tau}} \rangle_{\mathcal{H}} \text{ for all } f \in \mathcal{H}$$

Definition. (Kernel)
The kernel of an RKHS is

$$k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} : (\mathbf{t}, \boldsymbol{\tau}) \mapsto k_{\boldsymbol{\tau}}(\mathbf{t})$$

where $k_{\boldsymbol{\tau}}$ is the element of $\mathcal{H}$ that defines the sampling at $\boldsymbol{\tau}$.

Proposition. A (separable) Hilbert space with orthobasis $\{\psi_n\}_{n\geq 1}$ is an RKHS iff $\forall \boldsymbol{\tau} \in \mathbb{R}^d$ $\sum_{n=1}^{\infty} |\psi_n(\tau)|^2 < \infty$

RKHS AN NON ORTHOGONAL BASIS

If $\{\phi_n\}_{n\geq 1}$ is a Riesz basis for $\mathcal{H}$, we know that every $x \in \mathcal{H}$ can be written

$$x = \sum_{n\geq 1} \alpha_n \phi_n \text{ with } \alpha_n \triangleq \langle x, \widetilde{\phi}_n \rangle$$

where $\{\widetilde{\phi}_n\}_{n\geq 1}$ is the dual basis.

> **Proposition.** A (separable) Hilbert space with Riesz basis $\{\phi_n\}_{n\geq 1}$ is an RKHS with kernel
>
> $$k(\mathbf{t}, \boldsymbol{\tau}) \sum_{n=1}^{\infty} \phi_n(\boldsymbol{\tau}) \widetilde{\phi}_n(\mathbf{t})$$
>
> iff $\forall \boldsymbol{\tau} \in \mathbb{R}^d \sum_{n=1}^{\infty} |\phi_n(\tau)|^2 < \infty$