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

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

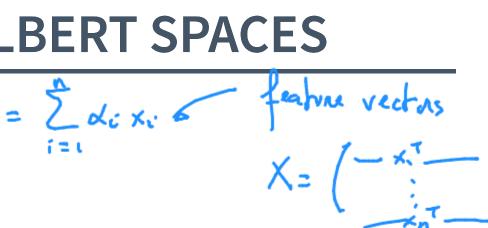
Let \mathcal{F} be a Hilbert space and let $f \in \mathcal{F}$ be the function we are trying to estimate

- We will estimate $f\in \mathcal{F}$ using noisy observations $\langle f,x_i
 angle$ 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

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

has solution

$$f = \sum_{i=1}^n lpha_i x_i$$
 with $oldsymbol{lpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}$ $\mathbf{K} = [\langle x_i, x_i \rangle + \lambda \mathbf{I})^{-1} \mathbf{y}$



 $\mathfrak{r}_{i}\rangle|_{1 < i.i < n}$

$$\frac{\operatorname{Poof}:}{\operatorname{lat}} \operatorname{time} \operatorname{we} \operatorname{poved} \operatorname{that} \operatorname{wlog} \operatorname{we} \operatorname{can} \operatorname{assume} f \in \operatorname{Span}(\operatorname{lxi})_{i=1}^{n}$$

$$\operatorname{let} \hat{f} \text{ be our minimizer, we} \operatorname{can} \operatorname{wnize} \hat{f} = \sum_{i=1}^{n} \widehat{\alpha}_{i} \times i \quad \left\{ \widehat{\alpha}_{i} \right\} \operatorname{GR}$$

$$\frac{\operatorname{let} \hat{f} \text{ be our minimizer, we}}{\operatorname{Span}(2\times \operatorname{s}_{i=1}^{n})} \operatorname{Hen} \quad u = \sum_{i=1}^{n} \operatorname{ci}_{i} \times i \quad \left\{ \widehat{\alpha}_{i} \right\} \operatorname{GR}$$

$$\frac{\operatorname{Nole}:}{\operatorname{let}} \operatorname{let} u, v \in \operatorname{Span}(2\times \operatorname{s}_{i=1}^{n}) \operatorname{Hen} \quad u = \sum_{i=1}^{n} \operatorname{ci}_{i} \times i \quad v = \sum_{i=1}^{n} \operatorname{di}_{i}$$

$$\frac{\operatorname{Vo}_{i} \times \operatorname{Span}(2\times \operatorname{s}_{i})_{i=1}^{n}) \operatorname{Hen} \quad u = \sum_{i=1}^{n} \operatorname{ci}_{i} \times i \quad v = \sum_{i=1}^{n} \operatorname{di}_{i}$$

$$\frac{\operatorname{Vo}_{i} \times \operatorname{Span}(2\times \operatorname{s}_{i})_{i=1}^{n}) \operatorname{Hen} \quad u = \sum_{i=1}^{n} \operatorname{ci}_{i} \times i \quad v = \sum_{i=1}^{n} \operatorname{di}_{i}$$

$$\frac{\operatorname{Vo}_{i} \times \operatorname{Span}(2\times \operatorname{s}_{i})_{i=1}^{n}) \operatorname{Hen} \quad u = \sum_{i=1}^{n} \operatorname{ci}_{i} \times i \quad v = \sum_{i=1}^{n} \operatorname{di}_{i}$$

$$\frac{\operatorname{Vo}_{i} \times \operatorname{Span}(2\times \operatorname{s}_{i})_{i=1}^{n}) \operatorname{di}_{i} = \operatorname{d}_{i} \operatorname{Gc} \operatorname{wi} \operatorname{d}_{i} = \operatorname{Gi}_{i} \operatorname{di}_{i} \times \operatorname{sin} \operatorname{d}_{i}$$

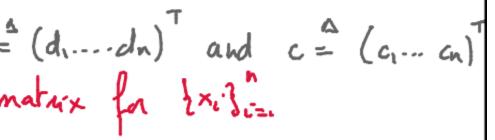
$$\frac{\operatorname{Vo}_{i} \times \operatorname{Span}(2\times \operatorname{s}_{i})_{i=1}^{n}}{\operatorname{Gi}_{i} \times (1\times \operatorname{s}_{i})_{i=1}^{n}} = \operatorname{d}_{i} \operatorname{Gc} \operatorname{wi} \operatorname{d}_{i} = \operatorname{Gi}_{i} \operatorname{di}_{i} \times \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{d}_{i} = \operatorname{Gi}_{i} \operatorname{di}_{i} \times \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{sin} \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{sin} \operatorname{d}_{i} \times \operatorname{sin} \operatorname{sin}$$

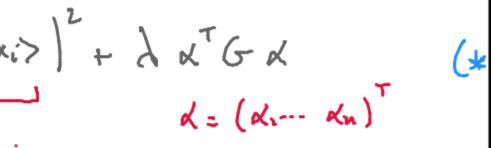
_,)

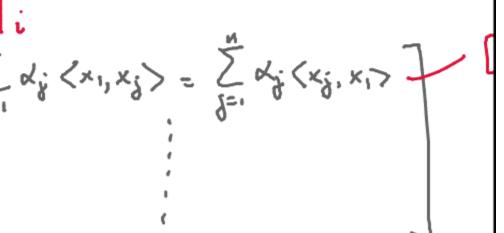
{xi]i= ETE

GIR

Xi







THE BIG PICTURE

6/14

We want to solve the finite dimensional publicity
while
$$\|'y - Gx\|_{z}^{2} + d x^{T}Gx$$
 (4)
 $d \in \mathbb{R}^{n}$
Taking the gradient of (*) and setting to 0 we obtain
 $\nabla_{x}(y_{y}^{T} - 2y_{y}^{T}Gx' + d^{T}G^{T}Gx' + d d^{T}Gx') = 0 \iff -2GT_{y} + 2G^{T}Gx' + 2$
 $d = 2G^{T}(Gx - y_{y}) + d Gx'$
Assume we pick x such that $(G + dI)x = y$ $[(G + dI)^{T}y = 1)^{T}y$

True ble G is symmetric $\nabla(x^TAx) = (A + A)$ 2 + A = 0X = 0 world work]

THE BIG PICTURE

For a Hilbert space $(\langle n \rangle)$ and $(n \rangle)$ pairs $((x_i, y_i) \rangle (n calF \rangle)$, we know how to solve the following problem with linear algebra \[\min_{f\in\calF}\sum_{i=1}^n\abs{y_i-{\dotp{f}} $x_i}_{\langle calF}^2+\langle ambda norm[\langle calF]{f} \rangle$

$$\begin{aligned} S = & \left[\mathcal{K} = (G + \delta \mathbf{I})^{T} \mathcal{Y} \right] \\ & \mathcal{K} = (G + \delta \mathbf{I})^{T} \mathcal{K} = \mathcal{K}^{T} (G + \delta \mathbf{I})^{T} \mathcal{Y} - \mathcal{Y} \right] + \mathcal{K} G + \delta \mathbf{I}^{T} \mathcal{Y} \\ & = \mathcal{K}^{T} (G + \delta \mathbf{I})^{T} - \mathbf{I} \mathcal{Y} + \mathcal{K} G + \delta \mathbf{I}^{T} \mathcal{Y} \\ & = \mathcal{K}^{T} (G + \delta \mathbf{I})^{T} - (G + \delta \mathbf{I}) (G + \delta \mathbf{I})^{T} \mathcal{Y} + \mathcal{K} G + \delta \mathbf{I}^{T} \mathcal{Y} \\ & = \left[\mathcal{K}^{T} - \mathcal{K}^{T} (G + \delta \mathbf{I}) + \mathcal{K} - \mathcal{K}^{T} \right] (G + \delta \mathbf{I})^{T} \mathcal{Y} \\ & = \left[\mathcal{K}^{T} - \mathcal{K}^{T} - \mathcal{K} - \mathcal{K}^{T} + \mathcal{K} - \mathcal{K}^{T} \right] (G + \delta \mathbf{I})^{T} \mathcal{Y} \\ & = \left[\mathcal{K}^{T} - \mathcal{K}^{T} - \mathcal{K} - \mathcal{K} - \mathcal{K} - \mathcal{K}^{T} + \mathcal{K} - \mathcal{K}^{T} \right] (G + \delta \mathbf{I})^{T} \mathcal{Y} \\ & = \left[\mathcal{K}^{T} - \mathcal{K}^{T} - \mathcal{K} -$$

THE BIG PICTURE

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

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

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

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

The question whether $f(\mathbf{x}_i) = \langle f, x_i
angle_{\mathcal{F}}$ 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

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

Definition.

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

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

Proposition. All norms are continuous functionals $F:\mathcal{F} o\mathbb{R}:x\mapsto\langle x,c
angle$ 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 $F(x)=\langle x,c
angle$ 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 $F(x)=\langle x,c
angle$ 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

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

Definition. (Reproducing Kernel Hilbert Spaces)

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

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

$$f(oldsymbol{ au}) = ig\langle f, k_{oldsymbol{ au}} ig
angle_{\mathcal{H}} ext{ for all } f \in \mathcal{H}$$

Definition. (Kernel)

The kernel of an RKHS is

$$k: \mathbb{R}^d imes \mathbb{R}^d o \mathbb{R}: (\mathbf{t}, oldsymbol{ au}) \mapsto k_{oldsymbol{ au}}(\mathbf{t})$$

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

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



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

$$x = \sum_{n \geq 1} lpha_n \phi_n$$
 with $lpha_n riangleq \langle x, \widetilde{\phi}_n
angle$

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

$$k(\mathbf{t},oldsymbol{ au})\sum_{n=1}^{\infty}\phi_n(oldsymbol{ au})\widetilde{\phi}_n(\mathbf{t})$$

iff $orall oldsymbol{ au} \in \mathbb{R}^d \sum_{n=1}^\infty \left| \phi_n(au)
ight|^2 < \infty$