Week 5. Sobolev Models

May 25, 2024

1 TODO

In Sieves/Gaussian Sobolev model hw, give eigenvalues and state eigenvectors are polynomials of increasing order. Optional exercise to prove it.

From there, width calc + subspace sieve stuff can just use it. Ask what the best choice of K is as a function of p.

2 Summary

So far, we've worked with the Sobolev model on even 2-periodic functions.

$$\mathcal{M} = \{\text{even 2-periodic } m(x) : \rho(m) \le B\} \quad \text{for} \quad \rho(m) = \left\| \frac{d}{dx} m \right\|_{L_2} \tag{1}$$

Because curves on $[0,1]$ are in one-to-one correspondence with even periodic functions,[1] this is a natural model for functions on the unit interval $[0,1]$. Because $\|u\|_{L_2}^2 = \mathrm{E}\, u(X)^2$ for X uniformly distribution on $[0,1]$, our constraint requires $|m'(X)|$ to be small most of the time for X with this distribution.

In this homework, we'll see what happens when we impose the same constraint for X with another distribution: the standard normal distribution. This is a bit more natural if our covariates X_i tend to be near zero but aren't bounded. And it saves us a little trouble, as we'll be modeling functions on all of $\mathbb{R}$ directly, so we won't need to think about periodic extension. Here's the model.

$$\begin{aligned} \mathcal{M} &= \left\{ m : \lim_{x \to \pm\infty} m(x)^2 \phi(x) = 0 \;\text{ and }\; \rho(m) \le B \right\} \\ &\text{for} \quad \rho(m) = \sqrt{\int_{-\infty}^{+\infty} \left\{ \frac{d}{dx} m(x) \right\}^2 \phi(x)\; dx} \end{aligned} \tag{2}$$

where $\phi(x) = (1/\sqrt{2\pi}) e^{-x^2/2}$ is the probability density for the standard normal distribution. This is a space of smooth functions m from $\mathbb{R} \to \mathbb{R}$ that don't grow too fast as x approaches $\pm\infty$.

We'll find a Fourier-series characterization of this model, use it to bound the gaussian width of a neighborhood, and use that to find a rate of convergence for the least squares estimator in this model. As we did when analyzing the Sobolev model (1), we'll start by making some simplifying assumptions. We'll bound the gaussian width of a neighborhood of zero when our covariate X_i has a standard normal distribution.

Throughout, we'll use the gaussian inner product and associated norm.

$$\langle u, v\rangle = \int_{-\infty}^{+\infty} u(x)v(x)\phi(x)dx \quad \text{and} \quad \|v\| = \sqrt{\langle v, v\rangle}. \tag{3}$$

If we prefer, we can think of these in terms of expectations involving a standard normal random variable.

$$\langle u, v\rangle = \mathrm{E}\, u(X)v(X) \quad \text{and} \quad \|v\| = \sqrt{\mathrm{E}\, v(X)^2} \quad \text{for} \quad X \sim N(0,1). \tag{4}$$

3 Fourier Series

Our first step is characterizing the adjoint of the differential operator $\frac{d}{dx}$. This isn't $-\frac{d}{dx}$, as it was when we were talking about the model (1), because we're working with a different vector space of functions with a different inner product.

Exercise 1 *Show that if we're using the gaussian inner product, the adjoint of the first derivative operator* $Lv(x) = \frac{d}{dx}v(x)$ *is* $L^\star u(x) = xu(x) - \frac{d}{dx}u(x)$.

Hint. The adjoint $L^\star$ satisfies $\langle L^\star u,\ v\rangle = \langle u,\ Lv\rangle$. Use integration by parts on $\int \{\phi(x)u(x)\}v'(x)dx$.

Hint. Why is it important that the functions and their derivatives don't grow to fast as $x \to \pm\infty$? How does that relate to our periodicity restriction from Lecture 8?

Using this, we can characterize the *self-adjoint operator* $L^\star L$. One advantage of this alternate definition is that it allows us to define a *family of models*, $\mathcal{M}^p$ for positive integers p, based on powers of the operator.[2]

$$\begin{aligned} \mathcal{M}^p = \left\{ m : \lim_{x \to \pm\infty} m(x)^2\phi(x) = 0 \ \text{ and } \ \rho_p(m) \le B \right\} \\ \text{for} \quad \rho_p(m) = \sqrt{\langle (L^\star L)^p m, m\rangle} \end{aligned} \tag{5}$$

Exercise 2 *What is the self-adjoint operator $L^\star L$ in this case? In terms of it, write an alternate definition of the seminorm ρ in* (2). *Something like ρ_1 from* (5)*, but more concrete. Then write one for the seminorm ρ_2 from* (5).

Having this definition (5) also reduces Fourier series representation to the calculation of eigenvalues and eigenvectors of $L^\star L$. Let's do it.

Exercise 3 *Find the eigenvectors and eigenvalues of $L^\star L$. Then write a Fourier series representation of $\mathcal{M}^p$, i.e., an equivalent characterization of the form $\mathcal{M}^p = \{m(x) = \sum_j m_j\phi_j(x) \;\; : \;\; \ldots\}$ where $\langle \phi_j, \phi_k\rangle = 1$ if $j = k$ and 0 otherwise.*

Rather than finding an explicit expression for the eigenvectors, which is hard to do and ultimately not all that useful, define them recursively. That is, write a formula that lets us compute each one given the one before it. I'll walk you through it below. This approach will give you a recursive formula for the eigenvalues, too, but it's simple enough that you should be able to turn it into an explicit expression $\lambda_j = \ldots$.

Walkthrough. Let's start by observing that we have some polynomial eigenvectors. The functions 1, x, and $x^2 - 1$ are eigenvectors with corresponding eigenvalues of 0, 1, and 2. We'll show that we've got polynomial eigenvectors of all orders. As a first step, show that if we've got one eigenvector $v(x)$, we can get another: $v'(x)$ will also be one. To do this, differentiate both sides of the identity $L^\star L\ v(x) = \lambda v(x)$. This will be useful, but if we think about polynomials, this helps us get lower-order ones from higher order ones, which is sort of the opposite of what we need. To *increase* the other of a polynomial $v(x)$, we can multiply it by x. The result, $xv(x)$, won't be an eigenvector. But it's close. Calculate $L^\star L\ xv(x)$ for an eigenvalue $v(x)$ and see if there's some other function $u(x)$ you can add so that $xv(x) + u(x)$ *is* an eigenvector.

Hint. If you group the terms of $L^\star L\{xv(x)\}$ the right way, you'll see a multiple of $L^\star Lv(x)$.

Hint. If $L^\star Lu(x) = \lambda_u u(x)$ and $L^\star Lw(x) = \lambda_w w(x) + (\lambda_u - \lambda_w)u(x)$, what's $L^\star L\ \{w(x) - u(x)\}$?.

Tip. The eigenvectors you'll get this way aren't unit-length. For example, consider x^2-1. Its squared length is $\mathrm{E}(X^2-1)^2 = \mathrm{E}\,X^4 - 2\,\mathrm{E}\,X^2 + 1 = 3-2+1$ for gaussian X. In the Fourier series representation, you'll want to scale them to so their length is one. For this assignment, it's fine to leave this abstract — once you've found a recursive formula for non-unit-length eigenvectors $\tilde{\phi}_j$, you can write your model in terms of $\phi_j = \tilde{\phi}_j / \|\tilde{\phi}_j\|$ without a formula for the denominator. But if you want to implement this model, you'll need to work out a way of computing this denominator. If you're interested, it may be helpful to have a formula for the moments of the gaussian distribution: $\mathrm{E}\,X^k$ for $X \sim N(0,1)$. See here.