Week 9 Homework: Misspecification and Self-Adjoint Operators

QTM 385-1: Machine Learning and Nonparametric Regression

In this homework, we'll do two things. To develop on our understanding of misspecification, we'll do some exercises on projection onto convex sets and its implications. To prepare for our discussion of Sobolev models next week, we'll talk a bit about using ideas you're familiar with in the context of linear operators on finite-dimensional vector spaces, i.e. matrices, in a new context—*differential operators* like $\frac{d^2}{dx^2}$ on vector spaces of functions.

1 Misspecified Models and Projection

Let $\mathcal{M}$ be any convex set of curves and μ be any curve outright.

Exercise 1 *Show that the model's closest curve to μ,*

$$\mu^\star = \operatorname*{argmin}_{m \in \mathcal{M}} \|m - \mu\|^2_{L_2(\mathrm{P_n})}$$

satisfies

$$\langle \mu^\star - \mu, m - \mu^\star \rangle_{L_2(\mathrm{P_n})} \geq 0 \quad \textit{for all curves} \quad m \in \mathcal{M}. \tag{1}$$

And draw a diagram or two to illustrate how your argument works and help you remember the result.

Hint. Focus on what goes on in a small neighborhood of $\mu^\star$. Instead of looking at the inner product $\langle \mu^\star - \mu, m - \mu^\star \rangle_{L_2(\mathrm{P_n})}$ for $m \in \mathcal{M}$, look at $\langle \mu^\star - \mu, m_\lambda - \mu^\star \rangle_{L_2(\mathrm{P_n})}$ for a curve $m_\lambda = \mu^\star + \lambda(m - \mu^\star)$ and take $\lambda \to 0$.
Hint. We know that $\|\mu^\star - \mu\|^2_{L_2(\mathrm{P_n})} \leq \|m_\lambda - \mu\|^2_{L_2(\mathrm{P_n})}$. Why? And why is convexity important?
Hint. Recall that $\|u\|^2_{L_2(\mathrm{P_n})} = \langle u, u \rangle_{L_2(\mathrm{P_n})}$. If we expand $\|m_\lambda - \mu\|^2_{L_2(\mathrm{P_n})}$ into a quadratic $a\lambda^2 + b\lambda + c$, $\langle \mu^\star - \mu,\ m - \mu^\star \rangle_{L_2(\mathrm{P_n})}$ will show up.

Exercise 2 *Explain why our least squares estimator estimator,*

$$\hat{\mu} = \operatorname*{argmin}_{m \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^{n} \{Y_i - m(X_i)\}^2,$$

will converge to $\mu^\star$, even if our model is misspecified in the sense that $\mu \notin \mathcal{M}$. In particular, explain the role the property (1) *plays.*

Exercise 3 *Why is it not necessarily the case that* (1) *is true when $\mathcal{M}$ isn't convex? Think about how this affects our least squares estimator $\hat{\mu}$. Can you think of a case — a curve μ and a non-convex model $\mathcal{M}$ — in which $\hat{\mu}$ doesn't converge to anything at all? It may help to draw some pictures.*

Exercise 4 *See if you can generalize* (1) *and subsequent arguments to say something about the* weighted least squares estimator*, $\operatorname{argmin}_{m\in\mathcal{M}}(1/n)\sum_{i=1}^n w(X_i)\{Y_i - m(X_i)\}^2$. Why might we want to use weighted least squares? If you like, think about the case that we're estimating a treatment effect using a regression discontinuity design.*

Hint. Is (1) specific to the inner product $\langle u, v\rangle_{L_2(\mathrm{P_n})}$ and the corresponding norm $\|v\|_{L_2(\mathrm{P_n})}$ or does it work for all inner products $\langle u, v\rangle$ and closest points in terms of corresponding norm $\|v\| = \sqrt{\langle v, v\rangle}$?

2 Self-adjoint Operators

In this problem, we'll generalize of the idea of a symmetric matrix.

You can think of an $n \times n$ matrix $A \in \mathbb{R}^{n\times n}$ as a linear operator on the vector space $\mathbb{R}^n$, i.e. a function from $\mathbb{R}^n$ to $\mathbb{R}^n$ that's linear in the sense that that $A(\alpha u + \beta v) = \alpha A u + \beta A v$ for any $\alpha, \beta \in \mathbb{R}$ and $u, v \in \mathbb{R}^n$. And we can talk about linear operators on other vectors spaces. For example, $\frac{d}{dx}$ is a linear operator on the space of infinitely-differentiable functions, as $\frac{d}{dx}\{\alpha u(x) + \beta v(x)\} = \alpha \frac{d}{dx} u(x) + \beta \frac{d}{dx} v(x)$.

When we're working with an inner product $\langle u, v\rangle$ on our vector space $\mathcal{V}$, we can define the *adjoint* $A^\star$ of a linear operator A to be another linear operator satisfying $\langle A^\star u, v\rangle = \langle u, Av\rangle$ for all vectors u and v. Here are some examples.

2.1 Operators on finite dimensional spaces

When we're working with the dot product $\langle u, v\rangle_2 = u^T v$ on $\mathbb{R}^n$, the adjoint of a matrix $A \in \mathbb{R}^{n\times n}$ is its *transpose* A^T.

$$\langle A^T u, v\rangle_2 = (A^T u)^T v = u^T A v = \langle u, Av\rangle_2.$$

When we're working with the dot product $\langle u, v\rangle_2 = u^T \bar{v}$ on $\mathbb{C}^n$, the adjoint of a matrix $A \in \mathbb{C}^{n\times n}$ is its *conjugate transpose* $\bar{A}^T$. That is, it's the matrix whose elements are the complex conjugates of the elements in A^T.

$$\langle \bar{A}^T u, v\rangle_2 = (\bar{A}^T u)^T \bar{v} = u^T \bar{A}\bar{v} = u^T \overline{Av} = \langle u, Av\rangle.$$

Why we use complex spaces. Even when we really intend to work with real-valued vectors, it's useful to think about matrices as operators on $\mathbb{C}^n$ and think of the dot product $\langle u, v\rangle_2$ as $u^T\bar{v}$. We have to deal with complex numbers in any case, as matrices $A \in \mathbb{R}^{n\times n}$, can have complex eigenvalues and eigenvectors. And the inner product $u^T v$ we use on $\mathbb{R}^n$ isn't an inner product on complex vectors at all, as the norm $\|v\|^2 = \langle v, v\rangle$ associated with an inner product must be positive and $u^T u$ will be negative for imaginary vectors.

2.2 Operators on spaces of fuctions

When we're working with the inner product $\langle u, v\rangle = \int_{-\infty}^{\infty} u(x)v(x)dx$ on the vector space of infinitely-differentiable functions v with $v(x) \to 0$ as $x \to \pm\infty$, the adjoint of the linear operator $\frac{d}{dx}$ is $-\frac{d}{dx}$. To see this, we integrate by parts.

$$\begin{aligned}
\left\langle u, \frac{d}{dx}v \right\rangle &= \int_{-\infty}^{\infty} u(x)v'(x)dx && \text{by definition} \\
&= u(x)v(x)|_{-\infty}^{\infty} \quad - \quad \int_{-\infty}^{\infty} u'(x)v(x)dx && \text{because} \quad (uv)' = u'v + uv' \\
&= 0 \quad - \quad \int_{-\infty}^{\infty} u'(x)v(x)dx && \text{because} \quad u(x)v(x) \underset{x\to\pm\infty}{\to} 0 \\
&= \left\langle -\frac{d}{dx}u, v \right\rangle.
\end{aligned}$$

Note that it's important that our vector space includes only functions that go to zero as $x \to \pm\infty$; otherwise our 'boundary term' $u(x)v(x)|_{-\infty}^{\infty}$ would be nonzero and we could not say that $-\frac{d}{dx}$ was the adjoint of $\frac{d}{dx}$. Specifying the vector space and inner product we're using is more important when talking about operators on spaces of functions than operators on finite-dimensional vectors. We can essentially get away with assuming we're talking about $\mathbb{C}^n$ and $\langle u, v\rangle = u^T\bar{v}$ in the latter case because that's what everyone always does; we don't have unspoken defaults like this for operators on functions.

The adjoint is still $-\frac{d}{dx}$ if we're thinking about $\frac{d}{dx}$ as a linear operator on the space of *complex-valued* infinitely-differentiable functions with $v(x) \to 0$ as $x \to \pm\infty$ with the inner product $\langle u, v\rangle = \int_{-\infty}^{\infty} u(x)\overline{v(x)}dx$. It's useful to think this way for the same reason it's useful to think about $\mathbb{C}^n$ instead of $\mathbb{R}^n$.

You may not be familiar with derivatives of complex-valued functions. That's no big deal. For a complex-valued function u, $u(x) = u_r(x) + iu_i(x)$ where u_r and u_i are real-valued functions, $\frac{d}{dx}u(x) = \frac{d}{dx}u_r(x) + i\frac{d}{dx}u_i(x)$ and $\int u(x)dx = \int u_r(x)dx + i\int u_i(x)dx$. We can show that $-\frac{d}{dx}$ is the adjoint of $\frac{d}{dx}$ by using integration by parts as above on the real and imaginary components separately.

2.3 Self-adjointness

A self-adjoint operator on a vector space $\mathcal{V}$ with an inner product $\langle u, v\rangle$ is, as you would expect, an operator that is its own adjoint. That is, we say an operator

A is self-adjoint if $\langle Au, v\rangle = \langle u, Av\rangle$. Symmetric matrices, i.e. matrices A with $A^T = A$, are self-adjoint on $\mathbb{R}^n$ with the usual inner product $\langle u, v\rangle = u^T v$. *Conjugate-symmetric* matrices, i.e. matrices A with $\bar{A}^T = A$, are self-adjoint on $\mathbb{C}^n$ with the usual inner product $\langle u, v\rangle = u^T \bar{v}$.

Now let's talk about self-adjoint operators on spaces of functions. A classic example is the differential operator $-\frac{d^2}{dx^2}$ on the space of complex-valued twice-differentiable functions on $[-1, 1]$ that are periodic in the sense that $v(-1) = v(1)$ with inner product $\langle u, v\rangle = (1/2)\int_{-1}^{1} u(x)\overline{v(x)}dx$.

Exercise 5 *Prove that the operator $-\frac{d^2}{dx^2}$ on this space is self-adjoint. That is, prove that $\langle -\frac{d^2}{dx^2}u,\ v\rangle = \langle u, -\frac{d^2}{dx^2}v\rangle$ for periodic functions u and v.*

Hint. Integrate by parts twice. Why is it important that u and v be periodic?

2.4 Diagonalizing self-adjoint operators

A linear operator has eigenvalues and eigenvectors, just like a matrix.[1] In our example, they are defined by the differential equation $-\frac{d^2}{dx^2}v = \lambda v$. And like a symmetric matrix, a self-adjoint linear operator's eigenvalues are real and the eigenvectors corresponding to distinct eigenvalues are orthogonal.

Exercise 6 *Prove that. And having done this, explain why this implies that, for integers j and k with $j \neq k$,*

$$\int_{-1}^{1} \sin(\pi kx)\sin(\pi jx) = \int_{-1}^{1} \cos(\pi kx)\cos(\pi jx) = \int_{-1}^{1} \cos(\pi kx)\sin(\pi jx)dx = 0.$$

Hint. Recall from our Vector Spaces Homework that for any vectors u and v,

$$\langle u, v\rangle = \overline{\langle v, u\rangle}, \quad \langle u + \alpha v, w\rangle = \langle u, w\rangle + \alpha\langle v, w\rangle, \quad \text{and} \quad \langle u, u\rangle \geq 0.$$

Let u and v be eigenvectors of A and use the property that $\langle Au, v\rangle = \langle u, Av\rangle$ for $u = v$ and for $u \neq v$.
Hint. What are $\frac{d^2}{dx^2}\sin(\pi kx)$ and $\frac{d^2}{dx^2}\cos(\pi kx)$?

Next week. We'll use the results we've proven to talk about the models defined using the *Sobolev seminorm* $\rho(v) = \sqrt{\int_0^1 |v'(x)|^2 dx}$ and its generalizations.

[1] The eigenvectors of operators on vector spaces of functions are sometimes called eigenfunctions.