Week 12. Misspecification

May 26, 2024

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, $\mathrm{argmin}_{m \in \mathcal{M}} (1/n) \sum_{i=1}^{n} w(X_i)\{Y_i - m(X_i)\}^2$. *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 $\|v\| = \sqrt{\langle v, v \rangle}$?