Week 5 Tutorial Tasks

Sarat Moka

With a single predictor but a high-degree polynomial basis, the coding task illustrates

1. how a Gaussian prior prevents over-fit, and
2. how Bayesian credible bands differ from OLS confidence bands.

Theoretical Task: OLS vs Bayesian Linear Regression

We use a single predictor x but fit a degree-d polynomial basis

$$\phi(x) = [1,\, x,\, x^2, \ldots, x^d]^\top.$$

Denote $X \in \mathbb{R}^{n\times(d+1)}$ with rows $\phi(x_i)^\top$, and observe noisy targets $y_i = f(x_i) + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Assume that $d \leq n$ and x_i's are distinct.

Task 1. Derive the ordinary least squares estimator $\hat{\beta}_{\mathrm{OLS}}$ that minimizes the mean squared error $\|y - X\beta\|_2^2$.

Task 2. Introduce a Gaussian prior $\beta \sim \mathcal{N}(0, \tau^2 I)$, where I denotes the identity matrix of appropriate dimension. Derive the posterior distribution $p(\beta \mid X, y)$.

Task 3. Derive the pointwise 95% *confidence interval* under OLS and the 95% *credible interval* under the Bayesian posterior for the predicted value $\hat{f}(x_*) = \phi(x_*)^\top \beta$.

Coding Task

Consider above polynomial regression when the underlying model is

$$y = \sin(2\pi\, x) + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, 0.3)$. Taking 30 linearly spaced x_i's on $[0, 1]$, compute the above confidence intervals and credible intervals for $d = 20$.