1D "Warm Up" - Approximation Theory

Aim: Illustrate key ideas about model accuracy in "simple" 1D context.

Problem 2. Polynomial Regression

Given samples from a continuous map,

(*) $$y_i = G(x_i) \quad \text{for } i = 0, \ldots, n,$$

where $G : [0,1] \to \mathbb{R}$, find a degree $m \le n$ polynomial that "best fits" the data,

(**) $$p_m = \underset{p \in \mathbb{P}_m}{\operatorname{argmin}} \| Y - p(X) \|,$$

where $Y = [y_0, \ldots, y_n]^T$ and $X = [x_0, \ldots, x_n]^T$.

$\Rightarrow$ Focus on $\|X\| = \sqrt{\sum_{k=0}^{n} x_i^2}$ today, corresponding to "least-squares" fit.

Algorithm 2. "Dictionary" Regression

1. Select basis $\{e_0(x), \ldots, e_m(x)\} \in \mathbb{P}_m$.

2. Minimize objective function

$$R(C) = \left\| \begin{bmatrix} e_0(x_0) & \cdots & e_m(x_0) \\ \vdots & & \vdots \\ e_0(x_n) & & e_m(x_n) \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_m \end{bmatrix} - \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix} \right\|$$

where $C = [c_0, \ldots, c_m]^T$.

3. The best degree $\leq m$ polynomial is

$$p_m(x) = c_0 e_0(x) + \cdots + c_m e_m(x).$$

Q: How accurate is the "fit"?

"Model Error" $\quad E_{n,m} = \sup_{0 \leq x \leq 1} |G(x) - p_m(x)|$

Q: Given n, how do choose m?

⇒ See demo02.m for numerical experiments w/polynomial regression.

⇒ Remarkably, when $n \gg m$, the model error closely tracks the best $\mathbb{P}_m$ approximant - similar to interpolation at Chebyshev nodes.

Analysis of Regression. (View 1)

Consider a "continuous" analogue of Step 2 in Algorithm 2:

Discrete least-squares

$$\underset{c_0,\ldots,c_m}{\text{minimize}} \sum_{j=0}^{n}\left[y_j - \sum_{k=0}^{m} c_k e_k(x_j)\right]^2$$

continuous least-squares

$$\underset{c_0,\ldots,c_m}{\text{minimize}} \overbrace{\int_0^1 \left[G(x) - \sum_{k=0}^{m} c_k e_k(x)\right]^2 dx}^{\text{"}L^2\text{-norm" of } G - p_m}$$

⇒ The minimizer of the continuous problem is the best degree n polynomial approx. to G when measured in the $L^2[0,1]$ norm.

⇒ Just like the $C[0,1]$ problem, the minimizer exists, is unique, and converges algebraically with power dependent on smoothness of G.

⇒ The discrete objective is essentially a quadrature approximation of the continuous objective. As $n \to \infty$ (m fixed), the quadrature is "refined" and the minimizer of the discrete problem looks increasingly like the minimizer of the continuous problem!

If data is equispaced with $h = |x_{j+1} - x_j|$,

$$\underbrace{h \sum_{j=0}^{n} \left[y_j - \sum_{k=0}^{m} c_k e_k(x_j) \right]^2}_{\text{Riemann Sum}} \xrightarrow{n \to \infty} \int_0^1 [G(x) - p(x)]^2 \, dx$$

If data is not equispaced, $h_j = |x_{j+1} - x_j|$,

$$\underbrace{\sum_{j=0}^{n} h_j \left[y_j - \sum_{k=0}^{m} c_k e_k(x_j) \right]^2}_{\text{Riemann Sum}} \xrightarrow{n \to \infty} \int_0^1 [G(x) - p(x)]^2 \, dx$$

provided that $\limsup_{j \to \infty} h_j = 0$.

The modified objective for the discrete problem leads to weighted least-squares.

Analysis of Regression. (View 2)

The sensitivity of Algorithm 2 outputs to small perturbations in the inputs is measured by the condition #

$$K(A) = \frac{\sigma_1(A)}{\sigma_m(A)}, \qquad A = \begin{bmatrix} e_0(x_0) & \cdots & e_m(x_0) \\ \vdots & & \vdots \\ e_0(x_n) & \cdots & e_m(x_n) \end{bmatrix},$$

where $\sigma_1(A) \geq \cdots \geq \sigma_m(A)$ are the singular values of the matrix A.

Just as we saw for dictionary interpolation ($m=n \to$ Algorithm 1), there are two primary factors:

$\Rightarrow$ The quality of the basis for $\mathbb{P}_n$

$\Rightarrow$ The quality of the data X.

If a "good" basis like Legendre polynomials is used, the quality of the data is crucial in influencing $\kappa(A)$.

Remarkably, the influence of the data on $\kappa(A)$ diminishes when $m \ll n$.

A rigorous analysis bounds the condition #'s of the sampled dictionary matrix.

This is a form of regularization: trade accuracy for better conditioning of the underlying mathematical problem.