1D "Warm Up" - Approximation Theory

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

Problem 1. Polynomial Interpolation.

Given samples from a continuous map,

(*) $$y_i = G(x_i), \quad \text{for } i = 0, 1, 2, \ldots, n,$$

where $G : [0,1] \to \mathbb{R}$, find the unique degree $\leq n$ polynomial that reproduces the observed data, i.e., interpolates G.

Algorithm 1. "Dictionary" Interpolation

1. Select basis $\{e_0(x), \ldots, e_n(x)\} \subseteq \mathbb{P}_n$

2. Solve $$\begin{bmatrix} e_0(x_0) & \cdots & e_n(x_0) \\ \vdots & & \vdots \\ e_0(x_n) & \cdots & e_n(x_n) \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}$$

3. Interpolant is $F_n(x) = c_0 e_1(x) + \cdots + c_n e_n(x)$

Q: How accurate is the interpolant?

"Model Error" $$E_n = \sup_{x \in [0,1]} |G(x) - F_n(x)|$$

Analysis of Interpolation

To analyze the accuracy of the interpolant, we follow a strategy that is common in approximation theory.

Step 1. Best Approximation in $\mathbb{P}_n$.

Q: How well is G approximated by polynomials in $\mathbb{P}_n$?

Step 2. Suboptimality of $\mathbb{P}_n$-interpolant.

Q: How "far from best" is the interpolant?

Best Approximation

We want to find $p_* \in \mathbb{P}_n$ that solves

$$(*) \qquad p_* = \underset{p \in \mathbb{P}_n}{\operatorname{argmin}} \|p - G\|,$$

where $\|p\| = \sup_{x \in [0,1]} |p(x)|$ is the "sup" norm.

Key Fact 1. Existance & Uniqueness

If $G \in C[0,1]$ (continuous on $[0,1]$), then there is a unique best approximant $p_* \in \mathbb{P}_n$ that satisfies $(*)$.

pf| Existence: the space $\mathbb{P}_n$ is a complete n-dimensional vector space and the subset

$$B = \{p \in \mathbb{P}_n \text{ s.t. } \|p - G\| \leq \|G\|\}$$

is closed & bounded, hence, compact.

The error function $p \to \|p - G\|$ is continuous w.r.t. $p \in B$ and, therefore, achieves its minimum on B.

Uniqueness: Chebyshev Equioscillation Thm.

<u>Key Fact 2.</u> Estimates for Smooth Functions

The error in the best polynomial approximation to G is tightly linked to the "smoothness" or "regularity" of G.

<u>Jackson Thm.</u> Let $G \in C^k[0,1]$ $(k \geq 1)$ and $n \geq k-1$. Then,

$$\inf_{p \in \mathbb{P}_n} \|p - G\| \leq \left(\frac{\pi}{2}\right)^k \frac{1}{(n+1)\cdots(n-k+2)} \|G^{(k)}\|.$$

Essentially, the error decreases algebraically

$$\Rightarrow \quad \inf_{p \in \mathbb{P}_n} \|p - G\| \leq C_k \frac{\|G^{(k)}\|}{n^k}$$

__Pf/__ Integration-by-parts n-times & a few "tricks." See attached reading.

Relation to Interpolant

How far is interpolant from "best"?

Given distinct points $x_0, \ldots, x_N \in [0,1]$, every continuous function $f:[0,1] \to \mathbb{R}$ has a unique polynomial interpolant of deg $\leq N$:

$$p = P_x f.$$

The map $P_x : C[0,1] \to \mathbb{P}_n$ is a linear projection of $C[0,1]$ onto $\mathbb{P}_n$ with

$$\|P_x\| = \sup_{f \in C[0,1]} \frac{\|P_x f\|}{\|f\|} \geq 1.$$

This amplification factor, the operator norm of P_X determines how far the interpolants can be from "best."

<u>Claim:</u> Given $X = \{x_0, \dots, x_n\} \subset [-1,1]$, $f \in C[0,1]$, and $p_* = \underset{p \in \mathbb{P}_n}{\operatorname{argmin}} \|p - f\|$,

$$\|f - P_X f\| \leq (1 + \|P_X\|)\|f - p_*\|$$

interp. error $\qquad\qquad$ best fit error

<u>pf</u> $\quad \|f - P_X f\| \leq \|f - p_*\| + \|p_* - P_X f\|$

Now, $P_X p_* = p_*$ because $p_* \in \mathbb{P}_n$, so

$$\|p_* - P_X f\| = \|P_X(p_* - f)\| \leq \|P_X\| \|f - p_*\|$$

$$\Rightarrow \quad \|f - P_X f\| \leq (1 + \|P_X\|)\|f - p_*\|.$$

For "good" sets of interpolation nodes

(slow growth) $\quad \|P_X\| \sim \log(\#\text{ nodes}) \quad$ as $n \to \infty$.

For equally-spaced nodes,

(exponential growth) $\|P_x\| \sim 2^{n+1}(n\log n)^{-1}$ as $n \to \infty$.

<u>Lemma</u> The operator norm of P_x is called the Lebesgue constant and can be calculated:

$$\|P_x\| = \max_{x\in[0,1]} \sum_{k=0}^{n} \left| \prod_{\substack{i=0\\ i\neq k}}^{n} \frac{x-x_i}{x_k-x_i} \right|.$$

(sketch) <u>Pf/</u> The Lagrange form of the interpolant

$$[P_x f](x) = \sum_{k=0}^{n} f(x_k)\, l_k(x), \text{ where}$$

"nodal polynomial" $$l_k(x) = \prod_{\substack{i=0\\ i\neq k}}^{n} \frac{x-x_i}{x_k-x_i} = \begin{cases} 1 & x = x_k \\ 0 & x = x_j \quad j\neq k \end{cases}$$

$$\Rightarrow \sup_{f\in C[0,1]} \frac{\|P_x f\|}{\|f\|} \leq \max_{x\in[0,1]} \sum_{k=0}^{n} |l_k(x)|$$

$\Leftarrow$ For $f(x) = \underline{1}$, "$\leq$" is "$=$"

Model Accuracy for Poly. Interp.

Collecting the results of our 2-step analysis of interpolation yields:

Thm Let $G \in C^k[0,1]$ for some $k \geq 1$ and let $p_n \in \mathbb{P}_n$ interpolate G at distinct points $x_0, \ldots, x_n \in [0,1]$. For $n \geq k-1$,

$$\|G - p_n\| \leq \|P_x\| \|G - p_*\| \leq C_k \|P_x\| \|G^{(k)}\| n^{-k}$$

where
$$p_* = \underset{p \in \mathbb{P}_n}{\operatorname{argmin}} \|G - p\|$$

is the best degree $\leq n$ approximation of G.

$\Rightarrow$ $\|G - p_*\|$ describes how well the "truth" is approximated by a degree $\leq n$ polynomial model.

$\Rightarrow$ This is controlled by the "smoothness" of G and depends both on G and our choice of a $\mathbb{P}_n$-model.

$\Rightarrow$ $\|P_x\|$ describes the influence of the data distribution on $[0,1]$ on the accuracy of the approximation.

Regression: Oversampling

In data-driven settings, we may not have control over the sampling points, which can severely limit the accuracy of our model because $\|P_x\|$ often grows rapidly as n increases.

Q: What can we do if we are stuck with "bad" sample points?

Idea: Try using a lower degree polynomial model and find the "best fit" to the available data.

Interpolation $\Rightarrow$ Regression