Approximation of functions (from data)

1.) Local approximation via Taylor series :

Taylor polynomial of degree n, $f : \mathbb{R} \to \mathbb{R}$:

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

$$= f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \frac{f'''(x_0)}{3!}(x - x_0)^3 +$$

- $T_n(x)$ is a polynomial, but the function f that it approximates need not be one.

If f is indeed a polynomial, then $T_n(x)$ is an exact representation of f, if the degree of f is $k \leq n$.

But, this assumes that we can evaluate $f(x_0)$, and all its n derivatives at x_0.

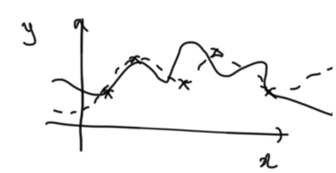
In practice, usually all we have is a set scattered observations;

$$\mathcal{D} := \{x_i, y_i\}, \quad i = 1, \dots, N, \quad x_i \in \mathbb{R}^d, \quad y_i \in \mathbb{R}^m$$

we want to "learn" the function $f : \mathbb{R}^d \to \mathbb{R}^m$ that generated the observed data, i.e. $y_i = f(x_i)$

unknown.

In classical scientific computing it is common to seek "parametric" approximations of the form :



$$f_\theta(x) = \sum_{k=1}^{n} w_k \varphi_k(x),$$

where $\theta := \{w_1, \dots, w_n\}$, and $\varphi_k(x)$ are known

unknown weights / parameters

features / basis functions

encodes any

with desirable properties (prior info ... may have

Key questions :

1.) How do we choose $\varphi_k(x)$?

2.) How do we determine θ ?

3.) How to assess the quality of our prediction ?

2.) Lagrange interpolation :

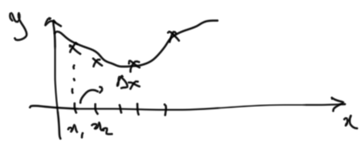
Assume : $f(x) = \sum_{k=1}^{n} W_k \varphi_k(x)$,

- $W_k := y_k \rightarrow$ k-th observed output (corr. to input ...
- $\varphi_k(x) = \prod_{\substack{0 \le k, j \le n \\ k \neq j}} \frac{x - x_j}{x_k - x_j}$

One can theoretically prove that the maximum approximation error

$$\max_x |E_n f(x)| \le \frac{\max_x |f^{(n+1)}(x)|}{4(n+1)} h^{n+1}, \quad h := \Delta x = x_i - \ldots$$

(*) Here we have assumed that the input data x_i live on a regular grid



(*) Runge's phenomenon can be avoided / mitigated if a suitable distribution on input points / nodes is used !

In a generar interval $[a, b]$:

- $x_i = \frac{a+b}{2} + \frac{b-a}{2} \hat{x}_i$, $\hat{x}_i = -\cos\left(\frac{\pi i}{n}\right)$, $i = 0, \ldots, n$

$\hookrightarrow$ Chebyshev – Gauss – Lobatto nodes

- $x_i = \frac{a+b}{2} - \frac{b-2}{2} \cos\left(\frac{2i+1}{n+1} \frac{\pi}{2}\right)$, $i = 0, \ldots, n$

$\hookrightarrow$ Chebyshev – Gauss nodes

3.) Interpolation with trigonometric polynomials (Fourier features) :

[for approximating periodic functions $f : [0, 2\pi] \rightarrow \mathbb{R}$]. FFT on a regular grid

<u>Assume</u>:

$$f(x) = \sum_{k=-m}^{m} w_k \varphi_k(x) \quad \text{(2m features)}, \qquad \begin{cases} \cdot\ w_k = \dfrac{1}{n+1}\displaystyle\sum_{j=0}^{n} f(x_j)\, e^{-ikj} \\ \cdot\ \varphi_k(x) = e^{ikx} \end{cases}$$

↪ Fourier featu

+ Exponential convergence for smooth periodic

General comments:

In all case the model parameters/weights can be identified via optimization by minimizing the mean-square prediction error:

$$\vartheta^* = \underset{\vartheta}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left[y_i - f_{\vartheta}(x_i) \right]^2$$

{ this is known as the "least-squares" method.

- <u>Pros</u>: + rigorous theory, well-understood behavior, error/convergence estimates.
- <u>Cons</u>: − rigid prior assumptions, scalability to high-dimensions, data corrupted by noise, outliers, etc.