Function Approximation from Scattered Data

Goal: Approximate \( f: \Omega \rightarrow \mathbb{R} \) from \( f_X = \{ f(x_1), \ldots, f(x_n) \}^T \).

Approach: Choose \( s(x) = \sum_{i=1}^n k(x, x_i)c \) with kernel \( k: \Omega \times \Omega \rightarrow \mathbb{R} \).

(often \( k(x, y) = \phi(||x - y||) \) for some radial basis function \( \phi \))

To fit: solve \( (K_{XX} + \lambda I)c = f_X \) where \( (K_{XX})_{ij} = k(x_i, x_j) \).

▶ Computational issue: \( K_{XX} \) is dense and ill-conditioned.
▶ Theoretical issue: How to choose kernel?

Kernel Regression Stories

Feature map

Data-dependent basis

Energy minimization

Gaussian process

Minimize

\[ \lambda \|s\|^2_H + \|s - f_X\|^2 \]

where \( s(x) = \langle d, \psi(x) \rangle_{\mathcal{H}} \) for some feature map \( \psi: \Omega \rightarrow \mathcal{H} \).

Gives \( d = \sum_{i=1}^n c_i \psi(x_i) \), kernel is \( k(x, y) = \langle \psi(x), \psi(y) \rangle_{\mathcal{H}} \).

Can reconstruct features if needed from eigenpairs of \( \mathcal{H} U \)

\[ u = \int_{\Omega} k(x, y)u(y)dy \]

Or treat as regularized regression with a data-dependent basis determined by sample locations (overcomes Mairhuber-Curtis).

Or Gaussian process: Gaussian random variables indexed by \( \Omega \), kernel gives covariance, regression gives posterior mean.

Approximation by Chebyshev Features

Alternate idea: Use a kernel-independent \( \mathcal{U} \subset \mathcal{H} \) but kernel determines the inner product.

Concrete 1D case: \( k(x, y) = \phi(x - y) = T(x)^T M T(y) \), where

\[ T(x) = [T_0(x), T_1(x), \ldots]^T \] (Chebyshev features)

\( M \) determined from \( k \)

Truncated expansion gives polynomial \( s(x) = T(x)d \) with \( (T_k^T T_k + \lambda M^{-1})d = T_k^T f_X \).

Constructing the Inner Product

Goal: \( \phi(x - y) = T(x)^T M T(y) \).

Approach: Compute \( D_k: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) s.t. \( T_k((x - y)/2) = T(x)^T D_k T(y) \).

Then

\[ \phi(x - y) = \sum_{k=0}^\infty a_k T_k((x - y)/2) \]

\[ = T(x) \left( \sum_{k=0}^\infty a_k D_k \right) T(y). \]

Rewrite recurrence on \( T_k(x) \) as operator on \( T(x) \) vector:

\[ xT_k(x) = \frac{1}{2} T_{k+1}(x) + T_{k-1}(x), \quad k > 0 \]

\[ = \frac{1}{2} T_k(x), \quad k = 0 \]

\[ xT(x) = \frac{1}{2} ST(x), S \equiv \text{tridiag} \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \ldots \end{pmatrix} \]

Then \( T_{k+1}(z) = 2zT_k(z) - T_{k-1}(z) \) for \( z = (x - y)/2 \) yields

\[ T_{k+1}((x - y)/2) = T(x) \left( \frac{1}{2} ST D_k - \frac{1}{2} D_k S - D_{k-1} \right) T(y) \]

\( \Rightarrow \)

\[ D_{k+1} = \frac{1}{2} ST D_k - \frac{1}{2} D_k S - D_{k-1}. \]

with starting values

\[ D_0(0:0.0:0) = 1, \quad D_1(0:1.0:1) = \begin{bmatrix} 0 & 1 \end{bmatrix} \]

Splitting the Kernel

Common case: not low rank, lacks regularity near zero. Write

\[ \phi(r) \approx \phi_{\text{smooth}}(r) + \phi_{\text{fp}}(r) \]

where

\[ \phi_{\text{smooth}}(r) \] is an even polynomial (treat as above)

\( \phi_{\text{fp}}(r) \) is supported only near origin

Resulting kernel matrix looks like

\[ K_{XX} \approx T_k M T_X + B, \]

where first term is low rank (as above), second term is sparse.

Matérn and SE kernels

Low-Rank Approximation of Kernels

Smooth kernels \implies eigenvalues of \( K_{XX} \) decay fast.

Approximate \( K_{XX} = UU^T \), regression \equiv regularized LS with \( U \):

\[ (U^T U + \lambda I)d = U^T f_X, \quad \lambda = \lambda^T (f_X - Ud). \]

Useful idea: approximate kernel function, not kernel matrix.

(Or devise an approximate feature map, like rows of \( U \).)

Examples:

▶ Using inducing points: \( k(x, y) = k_{xx} K_{2Z}^{-1} k_{2Y} \)

▶ Leading eigenpairs of associated integral operator \( \mathcal{K} \) (Mercer)

▶ Random Fourier features: \( k(x, y) = E_{\omega} \exp(\omega^T x) \exp(\omega^T y)) \), \( \omega \sim \text{Fourier transform of (scaled) kernel. Then MC quadrature.} \)

For each: reduced approximation space \( \mathcal{W} \subset \mathcal{H} \) and inner product on \( \mathcal{W} \) depend on kernel.