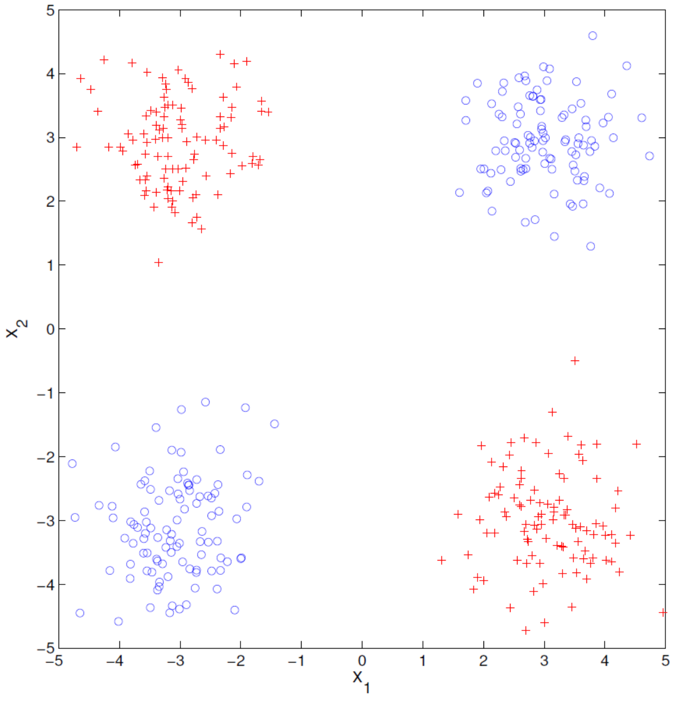
Introduction to Kernels

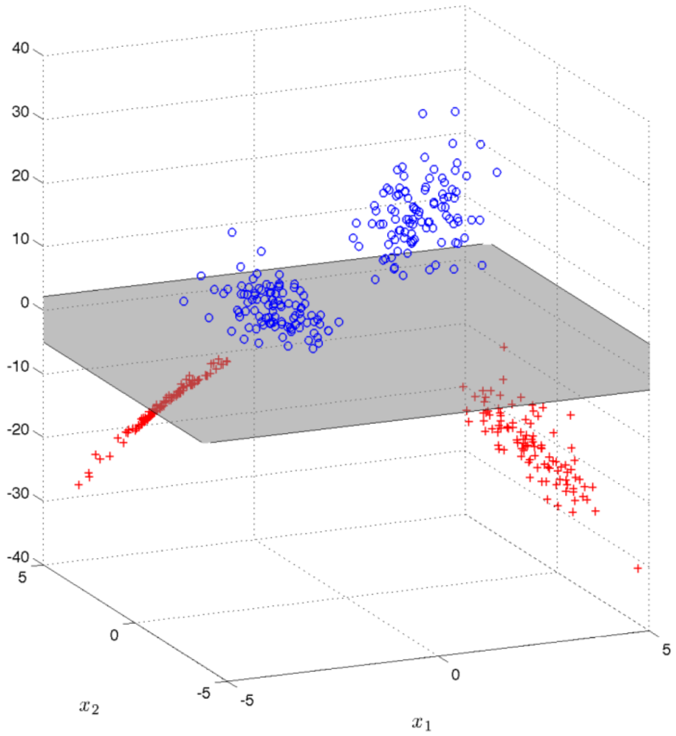
Overview

Motivation

- Unable to find a linear separation for XOR data ☹



- Use map projection: $\phi(x) = [x_1 \;\; x_2 \;\; x_1x_2]^T$



Definitions

- **Inner product**: given a vector space $\mathcal{H}$ defined over $\mathbb{R}$, we define an inner product to be a function $\langle \cdot,\cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ if the following conditions hold:
  1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{H}} = \alpha_1 \langle f_1, g \rangle_{\mathcal{H}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{H}}$
  2. $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$
  3. $\langle f, f \rangle_{\mathcal{H}} \geq 0$ and $\langle f, f \rangle_{\mathcal{H}} = 0$ iff $f = 0$
- **Inner product space**: a vector space that is equipped with an inner product.
- **Hilbert space**: a complete inner product space. Complete means every Cauchy sequence converges to a limit that is also contained in the space.
- **Kernel**: given a non-empty set $\mathcal{X}$, we define a kernel as a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ if there exists a Hilbert space $\mathcal{H}$ and a map $\phi: \mathcal{X} \to \mathcal{H}$ such that $\forall x, x' \in \mathcal{X}$,
$$k(x, x') \coloneqq \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

Definitions (Cont.)

- **Positive Definite**: if $\forall n \geq 1, \forall (a_1, \dots, a_n) \in \mathbb{R}^n, \forall (x_1, \dots, x_n) \in \mathcal{X}^n$, then the symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite if it satisfied the following:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

  - Note that we can easily show that an inner product is positive definite.

Key Mathematical Properties

- If $\mathcal{H}$ is a Hilbert space, $\mathcal{X}$ is a non-empty set, and $\phi : \mathcal{X} \to \mathcal{H}$, then $k(x, y)$ is a p.d. function:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} \langle a_i \phi(x_i), a_j \phi(x_j') \rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{n} a_j \phi(x_j') \right\rangle = \left\| \sum_{i=1}^{n} a_i \phi(x_i) \right\|_{\mathcal{H}}^2 \geq 0$$

- If $k(x, y)$ is a p.d. kernel, then there exists a feature space $\phi : \mathcal{X} \to \mathcal{H}$ s.t. the kernel is a dot product between features (Moore–Aronszajn).

Key Mathematical Properties (Cont.)

- Dealing with infinite dimensions
  - The ℓ_2 space is the space of all sequences that are square summable. Given $a \coloneqq (a_i)_{i \geq 1}$ in ℓ_2:

$$\|a\|_{\ell_2}^2 = \sum_{i=1}^{\infty} a_i^2 < \infty$$

  - Given $\left(\phi_i(x)\right)_{i \geq 1}$ in ℓ_2 where $\phi_i : \mathcal{X} \to \mathbb{R}$ is the *i*th coordinate of $\phi(x)$, then we can use Cauchy-Schwarz to show:

$$|k(x, x')| = \left| \sum_{i=1}^{\infty} \phi_i(x) \phi_i(x') \right| \leq \|\phi(x)\|_{\ell_2} \|\phi(x')\|_{\ell_2} < \infty$$

Kernel Construction

- Examples of kernels:
  - Linear: $k(x,z) = x^T z$
  - RBF: $k(x,z) = e^{-\frac{(x-z)^2}{\sigma}}$
  - Polynomial: $k(x,z) = \left(1 + x^T z\right)^d$
- Kernel composition[1]
  - $k(x,z) = x^T z$
  - $k(x,z) = ck_1(x,z)$
  - $k(x,z) = k_1(x,z) + k_2(x,z)$
  - $k(x,z) = g\big(k_1(x,z)\big)$
  - $k(x,z) = k_1(x,z)k_2(x,z)$
  - $k(x,z) = f(x)k_1(x,z)f(z)$
  - $k(x,z) = e^{k_1(x,z)}$
  - $k(x,z) = x^T A z$

[1] Given k_1 and k_2 are well-defined kernels, $c \geq 0$, g is a polynomial function with positive coefficients, A is p.s.d., f is any function.

Reproducing Kernel Hilbert Space

- We want to demonstrate that we can use kernels to define functions on $\mathcal{X}$. The space of these function is known as the RKHS.
- Given f : $\mathbb{R}^2 \rightarrow \mathbb{R}$ as in $f(x) = f_1x_1 + f_2x_2 + f_3x_1x_2$, we can represent f in terms of its coefficients, $f(\cdot) = [f_1 \;\; f_2 \;\; f_3]^T$, and we can evaluate it at a particular point $f(x) = f(\cdot)^T\phi(x) = \langle f(\cdot), \phi(x)\rangle_{\mathcal{H}}$
- Provides a duality for $\phi(x)$ where $\phi(x)$ represents both a mapping from $\mathbb{R}^2 \rightarrow \mathbb{R}^3$and $\mathbb{R}^2 \rightarrow \mathbb{R}$. Can write $\phi(x)$=k$(\cdot, x)$ and $\phi(y)$=k$(\cdot, y)$.

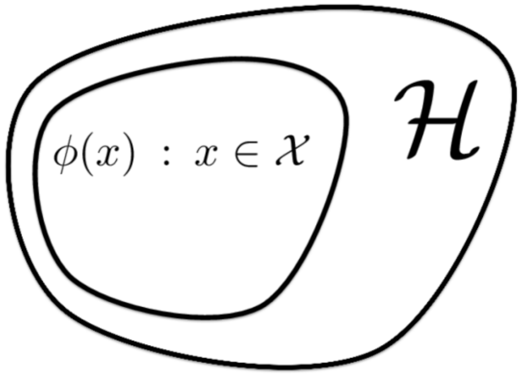
$$f(\cdot) = \mathrm{k}(\cdot, x) = [x_1 \;\; x_2 \;\; x_1x_2]^T = \phi(x)$$

$$\langle f(\cdot), \phi(y)\rangle_{\mathcal{H}} = \langle \mathrm{k}(\cdot, x), \phi(y)\rangle_{\mathcal{H}} = \mathrm{k}(x, y)$$

Reproducing Kernel Hilbert Space (Cont.)

- The feature map of every point is in the feature space:

$$\forall x \in \mathcal{X}, \mathrm{k}(\cdot, x) \in \mathcal{H}$$



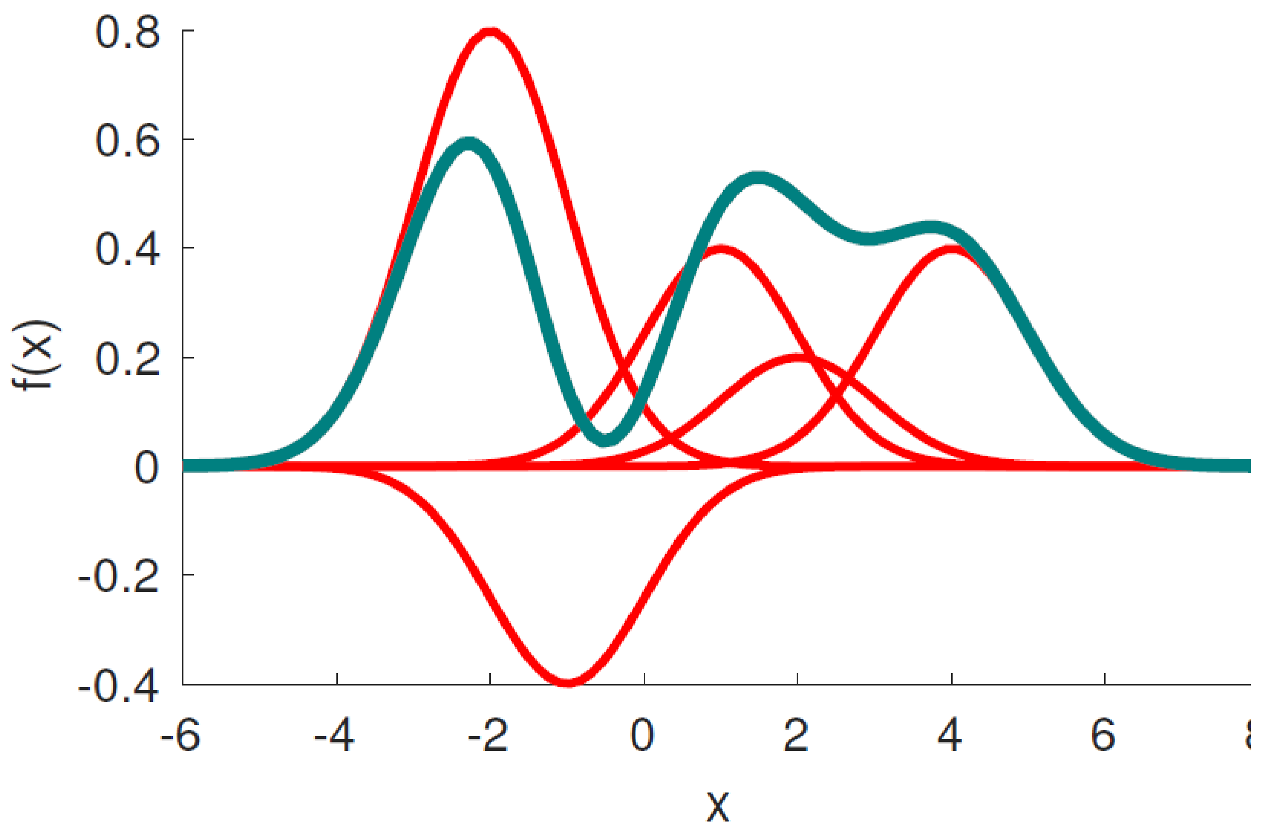
- The reproducing property:

$$\forall x \in \mathcal{X}, \forall \mathrm{f} \in \mathcal{H}, \langle f(\cdot), \mathrm{k}(\cdot, x) \rangle_{\mathcal{H}} = f(x)$$

$$\mathrm{k}(x, y) = \langle \mathrm{k}(\cdot, x), \mathrm{k}(\cdot, y) \rangle_{\mathcal{H}}$$

Kernel Trick

$$
\begin{aligned}
f(x) &= \sum_{i=1}^{\infty} f_i \phi_i(x) \\
&= \sum_{i=1}^{\infty} \left(\sum_{j=1}^{m} \alpha_j \phi_j(x_j) \right) \phi_i(x) \\
&= \left\langle \sum_{j=1}^{m} \alpha_j \phi(x_j), \phi(x) \right\rangle_{\mathcal{H}} \\
&= \sum_{j=1}^{m} \alpha_j k(x_j, x)
\end{aligned}
$$



RKHS in Action

- Fourier series representation

$$f(x) = \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell e^{i\ell x} = \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell (\cos(\ell x) + i\sin(\ell x))$$

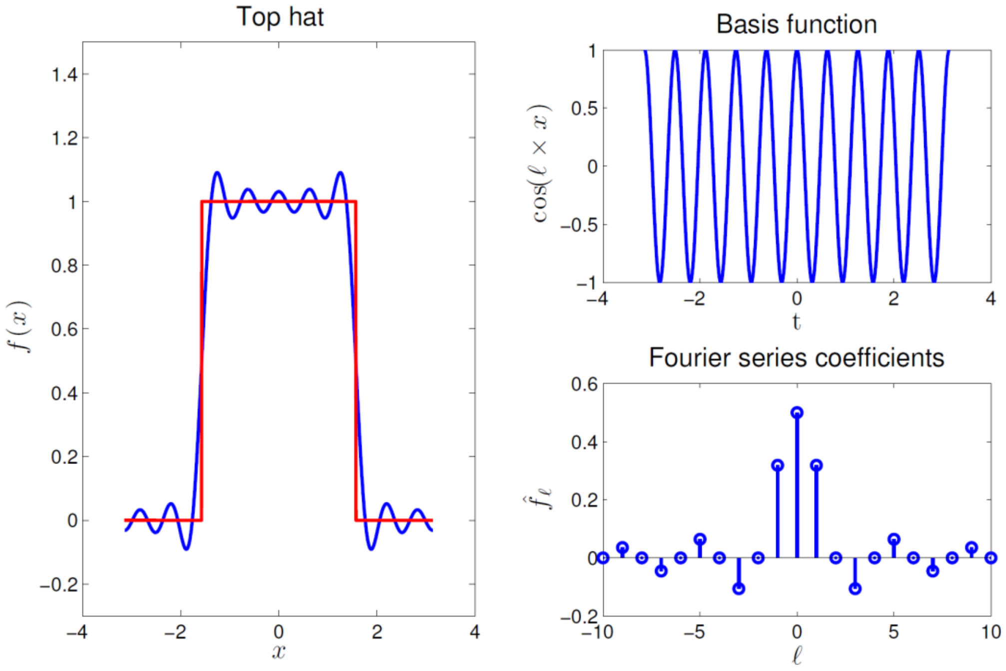
- Top hat function

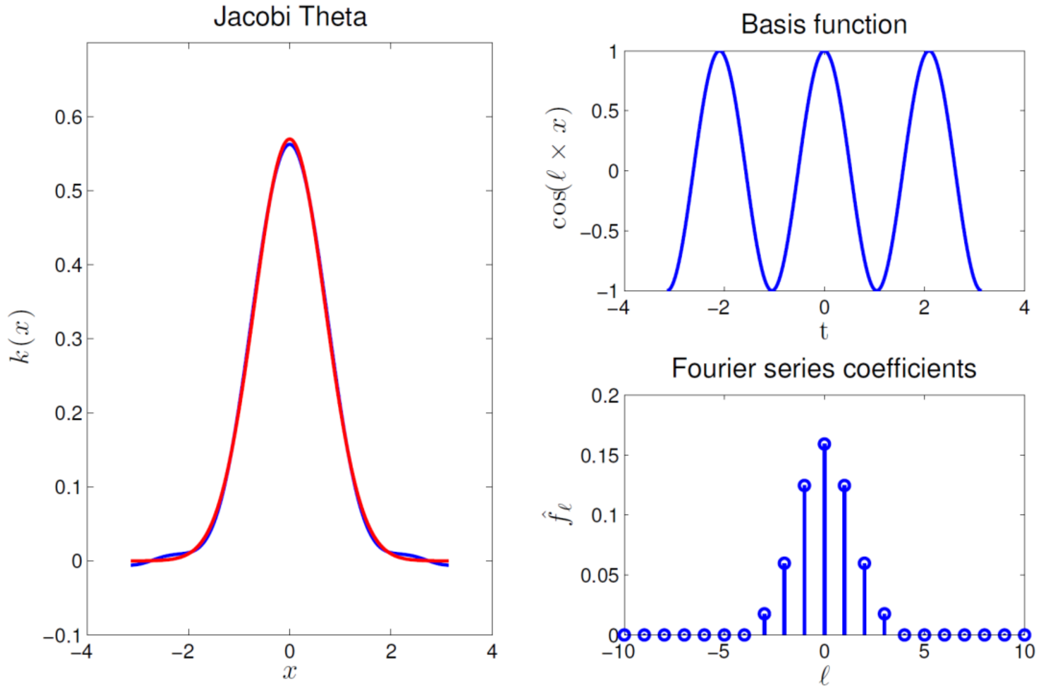
$$f(x) = \sum_{\ell=0}^{\infty} 2\hat{f}_\ell \cos(\ell x), \qquad \hat{f}_\ell = \frac{\sin(\ell T)}{\ell \pi}$$

- Jacobi theta kernel

$$k(x-y) = \frac{1}{2\pi} \vartheta\left(\frac{(x-y)}{2\pi}, \frac{i\sigma^2}{2\pi}\right), \qquad \hat{k}_\ell = \frac{1}{2\pi} e^{\left(\frac{-\sigma^2 \ell^2}{2}\right)}$$

RKHS in Action (Cont.)





Roughness Penalty

- Dot product in L_2

$$\langle f, g \rangle_{L_2} = \left\langle \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell e^{i\ell x}, \sum_{m=-\infty}^{\infty} \overline{\hat{g}_m e^{imx}} \right\rangle_{L_2}$$

$$= \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \hat{f}_\ell \, \bar{\hat{g}}_m \langle e^{i\ell x}, e^{-imx} \rangle_{L_2}$$

$$= \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell \bar{\hat{g}}_\ell$$

- Roughness penalty for dot product in $\mathcal{H}$

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{\ell=-\infty}^{\infty} \frac{\hat{f}_\ell \bar{\hat{g}}_\ell}{\hat{k}_\ell}, \qquad \|f\|_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}} = \sum_{\ell=-\infty}^{\infty} \frac{\left|\hat{f}_\ell\right|^2}{\hat{k}_\ell}$$

Reproducing Property

- Given the following kernel representation:

$$g(x) \coloneqq k(x-z) = \sum_{\ell=-\infty}^{\infty} e^{i\ell x} \hat{k}_\ell e^{-i\ell z} = \sum_{\ell=-\infty}^{\infty} \hat{g}_\ell e^{i\ell x}$$

- Given a function $f(\cdot) \in \mathcal{H}$:

$$\begin{aligned}\langle f(\cdot), g\rangle_{\mathcal{H}} &= \langle f(\cdot), k(\cdot, z)\rangle_{\mathcal{H}} \\ &= \sum_{\ell=-\infty}^{\infty} \frac{\hat{f}_\ell \bar{\hat{g}}_\ell}{\hat{k}_\ell} \\ &= \sum_{\ell=-\infty}^{\infty} \frac{\hat{f}_\ell \hat{k}_\ell e^{-i\ell z}}{\hat{k}_\ell} \\ &= \sum_{\ell=-\infty}^{\infty} \hat{f}_\ell e^{i\ell z} = f(z)\end{aligned}$$

Takeaway

- Small RKHS norm → smooth functions

$$f^* = \arg\min_{f \in \mathcal{H}} \left(\sum_{i=1}^{n} (y_i - \langle f, \phi(x_i) \rangle_{\mathcal{H}})^2 + \lambda \|f\|_{\mathcal{H}}^2 \right).$$