Reproducing Kernel Hilbert Spaces (RKHS), cont.

Review: $\mathcal{X} \subseteq \mathbb{R}^d$ (closed subset)

Mercer kernel

- $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- symmetric: $K(x, x') = K(x', x)$
- cont. in each variable:

  $x_n \to x \implies K(x_n, x') \to K(x, x')$
- positive semidefinite:

  $\forall\, x_1, x_2, \dots, x_n \in \mathcal{X}$

$$G = \begin{pmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_n) \\ K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_n,x_1) & K(x_2,x_n) & \cdots & K(x_n,x_n) \end{pmatrix}$$

  is positive semidefinite: $v^T G v \geq 0 \;\; \forall v \in \mathbb{R}^n$

e.g.

- take $\varphi_1, \varphi_2, \dots, \varphi_N : \mathcal{X} \to \mathbb{R}$ cont. fcns, lin. ind.

$$K(x, x') = \sum_{j=1}^{N} \varphi_j(x)\varphi_j(x') = \varphi(x)\varphi(x')^T$$

where $\varphi(x) := \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \\ \varphi_N(x) \end{pmatrix}$

- or $k : \mathbb{R}^d \to \mathbb{R}$ be reflection symmetric,

  $k(x) = k(-x)$, and

  $\lim\limits_{\|x\| \to \infty} k(x) = 0$

take $K(x, x') = k(x - x')$; Merker kernel if k has nonneg. Fourier transform.

e.g. $k(x) = e^{-a\|x\|^2}$ or $k(x) = \dfrac{1}{1 + a\|x\|^2}$ $(a > 0)$

Key Result: Mercer's Thm

Let K be a Mercer kernel on $\mathcal{X} \subseteq \mathbb{R}^d$.

Then $\exists$ a "unique" Hilbert space $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_K)$ of cont. fcns $f: \mathcal{X} \to \mathbb{R}$ s.t.:

1) linear span of $K_x = K(x, \cdot)$, $x \in \mathcal{X}$, is dense in $\mathcal{H}_K$
2) $K(x, x') = \langle K_x, K_{x'} \rangle_K \quad \forall x, x' \in \mathcal{X}$
3) $\forall f \in \mathcal{H}_K$, $\langle f, K_x \rangle_K = f(x)$ (reproducing kernel property)

Re 1): $f \in \mathcal{H}_K$

$\forall \varepsilon > 0 \qquad \exists c_1, \dots, c_N \in \mathbb{R}$
$\qquad\qquad\qquad x_1, \dots, x_N \in \mathcal{X}$

$$\text{s.t. } \Big\| f - \underbrace{\sum_{n=1}^{N} c_n K_{x_n}}_{\tilde{f}} \Big\|_K < \varepsilon$$

$$\tilde{f}(x) = \sum_{n=1}^{N} c_n K(x_n, x)$$

Re 3): $f \in \mathcal{H}_K$

$$f(x) = \langle f, K_x \rangle_K$$

let's say $f(x) = \sum_{n=1}^{N} c_n K_{x_n}(x)$

$$\langle f, K_x \rangle_K = \Big\langle \sum_{n=1}^{N} c_n K_{x_n}, K_x \Big\rangle_K$$

$$= \sum_{n=1}^{N} c_n \langle K_{x_n}, K_x \rangle_K$$

$$= \sum_{n=1}^{N} c_n K(x_n, x) \qquad (\text{by } 2))$$

$$= f(x)$$

Examples

- basis fcns

$\varphi_1, \dots, \varphi_N : \mathcal{X} \to \mathbb{R}$ cont. fcns, lin. ind.

— $\sum_{n=1}^{N} c_n \varphi_n(x) = 0 \ \forall x \in \mathcal{X} \iff c_1 = \dots = c_N = 0$

$$K(x, x') := \sum_{n=1}^{N} \varphi_n(x) \varphi_n(x') = \varphi(x) \varphi(x)^T$$

$$\mathcal{H}_\varphi = (\text{closure of span of } \varphi_1, \dots, \varphi_N) = \mathcal{H}_K$$

$$\langle f, g \rangle_K = \left\langle \sum_{n=1}^{N} c_n \varphi_n , \sum_{n=1}^{N} c_n' \varphi_n \right\rangle_K$$

$$= \sum_{n=1}^{N} c_n c_n'$$

$$K_x(\cdot) = \sum_{n=1}^{N} \underbrace{\varphi_n(x)}_{c_n} \varphi_n(\cdot) \in \overbrace{\text{Span}\{\varphi_1, \dots, \varphi_n\}}^{\mathcal{H}_\varphi}$$

$$\langle K_x, K_{x'} \rangle = \left(\sum_{n=1}^{N} \underbrace{\varphi_n(x)}_{c_n} \varphi_n(\cdot), \sum_{n=1}^{N} \underbrace{\varphi_n(x')}_{c_n'} \varphi_n(\cdot) \right)$$

$$= \sum_{n=1}^{N} \varphi_n(x) \varphi_n(x') = K(x, x')$$

$\mathcal{H}_K$: subspace of $\mathcal{H}_\varphi = (\text{span}\{\varphi_1, \dots, \varphi_N\}, \langle \cdot, \cdot \rangle_\varphi)$

Assume

$\exists\, g \in \mathcal{H}_\psi$ s.t. $\underbrace{\langle g, f \rangle_\psi = 0 \quad \forall f \in \mathcal{H}_K}$

$$g = \sum_{n=1}^{N} c_n \psi_n \Big\} \longrightarrow \langle g, K_x \rangle_\psi = 0 \quad \forall x$$

$$\Big\langle \sum_{n=1}^{N} c_n \psi_n, \sum_{n=1}^{N} \psi_n(x) \psi_n \Big\rangle_\psi$$

$$= \sum_{n=1}^{N} c_n \psi_n(x) = 0 \implies c_1 = \cdots = c_N = 0$$

by lin. ind.

- can be extended to countably many $\psi_1, \psi_2, \ldots : \mathcal{X} \to \mathbb{R}$ under additional conditions

$c \in \ell^2$ and $\sum_{n=1}^{\infty} c_n \psi_n(x) = 0 \quad \forall x$

$$\Big(c = (c_1, c_2, \ldots),\ \sum_{n=1}^{\infty} c_n^2 < \infty \Big)$$

$$\iff c_1 = c_2 = \cdots = 0$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ (inner-product kernel, linear classifiers)

$\psi_0(x) \equiv 1$

$\psi_j(x) = x_j$ for $j = 1, \ldots, d$

$$K(x, x') = \sum_{j=0}^{d} \psi_j(x) \psi_j(x') = 1 + x_1 x_1' + \cdots + x_d x_d'$$
$$= 1 + \langle x, x' \rangle$$

$$\mathcal{H}_K = \{ \langle w, x \rangle + b : w \in \mathbb{R}^d, b \in \mathbb{R} \}$$

$$f(x) = b + \sum_{j=1}^{d} w_j x_j = b \cdot \psi_0(x) + \sum_{j=1}^{d} \psi_j(w) \psi_j(x)$$

$$\|f\|_K^2 = b^2 + w_1^2 + \cdots + w_d^2 = b^2 + \|w\|^2$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ (polynomial kernel)

$$K(x, x') := (1 + \langle x, x' \rangle)^m \qquad m \in \mathbb{N}$$

$$= (1 + x_1 x_1' + \cdots + x_d x_d')^m$$

$$= \sum_{\substack{j_0, j_1, \dots, j_d \in \mathbb{Z}_+ \\ j_0 + \cdots + j_d = m}} \underbrace{\binom{m}{j_0 \ j_1 \ \cdots \ j_d}}_{\text{mult. coeffs.}} \prod_{i=1}^{d} x_i^{j_i} (x_i')^{j_i}$$

$$\underline{j} = (j_0, \dots, j_d) \in \mathbb{Z}_+^{d+1}, \quad j_0 + \cdots + j_d = m$$

$$\psi_{\underline{j}}(x) = \sqrt{\binom{m}{j_0 \ j_1 \ \cdots \ j_d}} \, x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}$$

$\mathcal{H}_K$: all polynomials in $x_1, \dots, x_d$ of deg m

- Gaussian kernel

$\mathcal{X} = \mathbb{R}$, $\quad K(x, x') = e^{-\frac{1}{2}(x - x')^2}$

$$e^{-\frac{1}{2}(x-x')^2} = e^{-\frac{1}{2}(x^2 + (x')^2)} e^{xx'}$$

$$= e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}(x')^2} \cdot \sum_{n=0}^{\infty} \frac{x^n (x')^n}{n!}$$

$$= \sum_{n=0}^{\infty} \psi_n(x) \psi_n(x')$$

where $\varphi_n(x) = e^{-x^2/2} \dfrac{x^n}{\sqrt{n!}}$

$$\mathcal{H}_K = \overline{\operatorname{span}\{\varphi_0, \varphi_1, \varphi_2, \dots\}}$$

$$= \operatorname{span}\left\{ \sum_{i=1}^{m} c_i\, e^{-\frac{1}{2}(x-x_i)^2} : \begin{array}{l} m \in \mathbb{N} \\ c_1, \dots, c_m \in \mathbb{R} \end{array} \right\}$$

- Extension to $K(x,x') = \exp(-\frac{1}{2}\|x-x'\|^2)$ on $\mathbb{R}^d$ using polynomial basis and above example

$$K(x,x') = e^{-\|x\|^2/2}\, e^{-\|x'\|^2/2} \sum_{n=0}^{\infty} \frac{\langle x, x'\rangle^n}{n!}$$

$$\langle x, x'\rangle^n = (x_1 x_1' + \dots + x_d x_d')^n \qquad \dots$$

$\mathcal{H}_K$: sums of shifted Gaussians and all their limits.

$$K(x,x') = e^{-a\|x-x'\|^2} \qquad (a > 0)$$

- General idea: RKHS / weighted inner products

$\tilde{\varphi}_1, \tilde{\varphi}_2, \dots : \mathcal{X} \to \mathbb{R}$

classifiers: $x \mapsto \operatorname{sgn}\left(\sum_j a_j \tilde{\varphi}_j(x)\right)$

choose $\lambda_1, \lambda_2, \dots > 0$ s.t. $\sum_j \lambda_j < \infty$

$[\lambda_j \to 0 \text{ as } j \to \infty]$

$$\psi_j(x) := \sqrt{\lambda_j}\,\tilde{\psi}_j(x) \qquad \forall j$$

$$K(x,x') = \sum_j \psi_j(x)\psi_j(x')$$

$$= \sum_j \lambda_j \tilde{\psi}_j(x)\tilde{\psi}_j(x')$$

classifier: $x \mapsto \operatorname{sgn} \hat{f}(x)$

where $\hat{f}(x) = \sum_j a_j \tilde{\psi}_j(x)$

$$= \sum_j \frac{a_j}{\sqrt{\lambda_j}} \psi_j(x) \quad \in \mathcal{H}_K$$

$$\|\hat{f}\|_K^2 = \sum_j \frac{a_j^2}{\lambda_j}$$

Constrain to $\hat{f}$ s.t. $\|\hat{f}\|_K^2 \leq R$

— $\hat{f}(x) = \sum_j \frac{a_j}{\sqrt{\lambda_j}} \psi_j(x)$

s.t. $\sum_j \frac{a_j^2}{\lambda_j} \leq R$ $\qquad \lambda_j \to 0$ as $j \to \infty$

Recap of $e^{-\frac{1}{2}(x-x')^2}$

$$f(x) = \sum_{i=1}^{m} \alpha_i e^{-\frac{1}{2}(x-x_i)^2}$$

$$= \sum_{i=1}^{m} \alpha_i \sum_{n=0}^{\infty} \psi_n(x_i)\psi_n(x)$$

$$= \sum_{n=0}^{\infty} \underbrace{\left(\sum_{i=1}^{m} \alpha_i \psi_n(x_i) \right)}_{c_n} \psi_n(x)$$

$\in \mathrm{span}\{\psi_0, \psi_1, \dots\}$

Preview for next lecture:

- fix Mercer kernel K
- look for classifiers in a bdd subset of $\mathcal{H}_K$

$$\hat{f}_n \in \underbrace{\{ f \in \mathcal{H}_K : \|f\|_K^2 \leq R \}}_{\mathcal{F}_R}$$

$$R_n(\mathcal{F}_R) \lesssim \sqrt{\frac{R}{n}}$$

- representer thm
- reduction to convex programs (surrogate loss!)