Kernel Methods in Machine Learning - Course Notes

Hugo Cisneros

1 Kernels and RKHS

1.1 Positive Definite Kernels

Definition 1

A kernel K is a comparison function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.
With n data point $\{x_1, x_2, ..., x_n\}$ a $n \times n$ matrix $\mathbf{K}$ can be defined by $\mathbf{K}_{ij} = K(x_i, x_j)$.
A kernel K is **positive definite** (p.d.) if it is **symmetric** ($K(x, x') = K(x', x)$) and for all sets of a and x

$$\boxed{\sum_i \sum_j a_i a_j K(x_i, x_j) \geq 0}$$

This is equivalent to the kernel matrix being **positive semi-definite**.

Examples:

- Kernel on $\mathbb{R} \times \mathbb{R}$ defined by $K(x, x') = xx'$ is p.d. ($xx' = x'x$ and $\sum_i \sum_j a_i a_j K(x_i, x_j) = (\sum_i a_i x_i)^2 \geq 0$).
- Linear kernel ($K(x, x') = \langle x, x' \rangle_{\mathbb{R}^d}$) is p.d
- More generally for any set $\mathcal{X}$, and function $\Phi : \mathcal{X} \to \mathbb{R}^d$, the kernel defined by $K(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathbb{R}^d}$ is p.d.

Theorem 1: Aronszajn, 1950

K is a p.d. kernel on the set $\mathcal{X}$ if and only if there exists a **Hilbert space** $\mathcal{H}$ **and a mapping** $\Phi : \mathcal{X} \to \mathcal{H}$ such that, for any x, x' in $\mathcal{X}$:

$$\boxed{K(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathcal{H}}}$$

(A Hilbert space is a vector space with an inner product and complete for the corresponding norm).

1.2 Reproducing Kernel Hilbert Spaces (RKHS)

Let $\mathcal{X}$ be a set and $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ a class of functions forming a Hilbert space.

Definition 2: Reproducing kernel

A kernel K is called a **reproducing kernel** (r.k.) of $\mathcal{H}$ if

- $\mathcal{H}$ contains all functions of the form

$$\boxed{\forall x \in \mathcal{X}, K_x : t \to K(x,t)}$$

- For every $x \in \mathcal{X}$ and $f \in \mathcal{H}$, $\boxed{f(x) = \langle f, K_x \rangle_{\mathcal{H}}}$

If there exists a r.k., $\mathcal{H}$ is called a RKHS.

Theorem 2: Equivalent Definition of RKHS

$\mathcal{H}$ is a RKHS if and only if for any $x \in \mathcal{X}$, the mapping

$$\begin{aligned} F :& \mathcal{H} \to \mathbb{R} \\ & f \mapsto f(x) \end{aligned}$$

is **continuous**.

As a corollary, convergence in a RKHS implies point-wise convergence.

Proof 1: of Theorem 2

($\Rightarrow$) If a r.k. exists in $\mathcal{H}$ then for any $(x,f) \in \mathcal{X} \times \mathcal{H}$:

$$\begin{aligned} |f(x)| &= |\langle f, K_x \rangle_{\mathcal{H}}| \\ &\leq \|f\|_{\mathcal{H}} \cdot \|K_x\|_{\mathcal{H}} && \text{(Cauchy-Schwarz)} \\ &\leq \|f\|_{\mathcal{H}} \cdot K(x,x)^{\frac{1}{2}} \end{aligned}$$

Therefore, $f \in \mathcal{H} \to f(x) \in \mathbb{R}$ is a linear continuous mapping because F is linear and $\lim_{f \to 0} F(f) = 0$

($\Leftarrow$) F is continuous, by the Riesz representation theorem: there exists a unique $g_x \in \mathcal{H}$ such that $f(x) = \langle f, g_x \rangle_{\mathcal{H}}$.
The function $K : (x,y) \mapsto g_x(y)$ is then a r.k. for $\mathcal{H}$

Theorem 3: Uniqueness of RKHS

If $\mathcal{H}$ is a RKHS, it has a **unique r.k.**, and a function K can be **the r.k of at most one RKHS**.

Proof 2: of Theorem 3

(Uniqueness) If K and K' are two r.k. of a RKHS, then for any x

$$\|K_x - K'_x\|^2 = K_x(x) - K'_x(x) - K_x(x) + K'_x(x) = 0$$

So $K_x = K'_X$

Theorem 4

A function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is **p.d. if and only if it is a r.k.**.

Proof 3: of Theorem 4

($\Leftarrow$) A r.k. is symmetric, and $\sum_{i,j} a_i a_j K(x_i, x_j) = \|\sum_i a_i K_{x_i}\|_{\mathcal{H}}^2 \geq 0$

($\Rightarrow$)

1.3 Examples

2 Kernel tricks

3 Kernel Methods: Supervised Learning

4 Kernel Methods: Unsupervised Learning

5 The Kernel Jungle

6 Open Problems and Research Topics