

# Kernel Methods in Machine Learning - Course Notes

Hugo Cisneros

## Contents

<b>1</b>	<b>Kernels and RKHS</b>	<b>1</b>
1.1	Positive Definite Kernels . . . . .	1
1.2	Reproducing Kernel Hilbert Spaces (RKHS) . . . . .	2
1.3	Examples . . . . .	3
<b>2</b>	<b>Kernel tricks</b>	<b>4</b>
2.1	Kernel trick . . . . .	4
2.2	Representer theorem . . . . .	4
<b>3</b>	<b>Kernel Methods: Supervised Learning</b>	<b>5</b>
<b>4</b>	<b>Kernel Methods: Unsupervised Learning</b>	<b>5</b>
<b>5</b>	<b>The Kernel Jungle</b>	<b>5</b>
5.1	Green, Mercer, Herglotz, Bochner and friends . . . . .	5
<b>6</b>	<b>Open Problems and Research Topics</b>	<b>6</b>
<b>A</b>	<b>Proofs</b>	<b>6</b>
A.1	Kernels and RKHS . . . . .	6
A.2	The Kernel Jungle . . . . .	7

## 1 Kernels and RKHS

### 1.1 Positive Definite Kernels

#### Definition 1

A kernel  $K$  is a comparison function  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ .  
With  $n$  data point  $\{x_1, x_2, \dots, x_n\}$  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  $x$

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

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  $\Phi : \mathcal{X} \rightarrow \mathbb{R}^d$ , 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**  $\Phi : \mathcal{X} \rightarrow \mathcal{H}$  such that, for any  $x, x'$  in  $\mathcal{X}$ :

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

Proof.

(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

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

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

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

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

is **continuous**.

Proof.

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

#### 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.

#### Theorem 4

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

Proof.

### 1.3 Examples

#### 1.3.1 Steps for finding the RKHS of a Kernel

1. Look for an **inner product** ( $K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$ )
2. Propose a **candidate RKHS**  $\mathcal{H}$
3. Check that the candidate  $\mathcal{H}$  is a **Hilbert space** ( inner product and complete)
4. Check that  $\mathcal{H}$  is **the RKHS**
  - $\mathcal{H}$  contains all the functions  $K_x : t \mapsto K(x, t)$
  - For all  $f \in \mathcal{H}$  and  $x \in \mathcal{X}$ ,  $f(x) = \langle f, K_x \rangle_{\mathcal{H}}$ .

#### 1.3.2 Linear Kernel

##### Definition 3 – Linear Kernel

In  $\mathbb{R}^d$ , the linear kernel is defined by  $K(x, y) = \langle x, y \rangle_{\mathbb{R}^d}$

##### Theorem 5 – RKHS of a linear Kernel

The RKHS of the linear kernel is the set of linear functions of the form  $f_w(x) = \langle w, x \rangle_{\mathbb{R}^d}$  for  $w \in \mathbb{R}^d$ , endowed with the inner product  $\langle f_w, f_v \rangle_{\mathcal{H}} = \langle w, v \rangle_{\mathbb{R}^d}$

#### 1.3.3 Polynomial Kernel

##### Definition 4 – Polynomial Kernel

In  $\mathbb{R}^d$ , the polynomial kernel is defined by  $K(x, y) = \langle x, y \rangle_{\mathbb{R}^d}^2$

##### Theorem 6 – RKHS of a polynomial Kernel

The RKHS  $\mathcal{H}$  of the polynomial kernel is the set of quadratic functions of the form  $f_S(x) = x^T S x$  for  $S \in \mathcal{S}^{d \times d}$

#### 1.3.4 Properties of kernels

If  $K_1, K_2$  are p.d. kernels,

- $K_1 + K_2$  is a p.d. kernel
- $K_1 \cdot K_2$  is a p.d. kernel

- $cK_1$  for  $c \geq 0$  is a p.d. kernel
- The point-wise limits of a sequence of p.d. kernels is a p.d. kernel.
- $\exp(K_1)$  is a p.d. kernel

**Small norms in the RKHS space means slow variations in the original space  $\mathcal{X}$  with respect to the geometry defined by the kernel.**

## 2 Kernel tricks

### 2.1 Kernel trick

**Statement:** All expression of vectors that can be written in terms of pairwise inner products can be transposed to a infinite dimensional space by replacing inner products with kernel evaluations.

### 2.2 Representer theorem

#### Theorem 7 – Representer theorem

Let  $\mathcal{X}$  a set with a p.d. kernel  $K$  and corresponding RKHS  $\mathcal{H}$ ,  $S = \{x_1, \dots, x_n\} \subset \mathcal{X}$  a set of points of  $\mathcal{X}$ .

Let  $\Phi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  a function strictly increasing w.r.t. the last variable.

Any solution to the optimization problem

$$\min_{f \in \mathcal{H}} \Phi(f(x_1), \dots, f(x_n), \|f\|_{\mathcal{H}})$$

admits a representation in the form

$$\forall x \in \mathcal{X}, f(x) = \sum_{i=1}^n \alpha_i K(x_i, x)$$

Proof

### 3 Kernel Methods: Supervised Learning

### 4 Kernel Methods: Unsupervised Learning

### 5 The Kernel Jungle

#### 5.1 Green, Mercer, Herglotz, Bochner and friends

##### 5.1.1 Green Kernel

###### Theorem 8 – Green Kernel in dimension 1

The set defined by

$$\mathcal{H} = \{f : [0, 1] \rightarrow \mathbb{R}, \text{absolutely continuous}, f' \in L^2([0, 1]), f(0) = 0\}$$

endowed with the inner product  $\forall(f, g) \in \mathcal{F}^2 \langle f, g \rangle = \int_0^1 f'(u)g'(u)du$ , is a RKHS with with r.k.

$$\forall(x, y) \in [0, 1]^2, K(x, y) = \min(x, y)$$

.

###### Theorem 9 – General Green Kernel

If  $D$  is a differential operator on a class of functions of  $\mathcal{H}$  such that the inner product  $\langle f, g \rangle_{\mathcal{H}} = \langle Df, Dg \rangle_{L^2(\mathcal{X})}$

Then  $\mathcal{H}$  is a RKHS and admits for r.k. the Green function of the operator  $D^*D$

##### 5.1.2 Mercer Kernels

###### Definition 5 – Mercer Kernels

A kernel  $K$  on a set  $\mathcal{X}$  is called a Mercer kernel if:

- $\mathcal{X}$  is a compact metric space (typically, a closed bounded subset of  $\mathbb{R}^d$ )
- $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a continuous p.d kernel (w.r.t the Borel topology)

## 6 Open Problems and Research Topics

### A Proofs

#### A.1 Kernels and RKHS

##### Proof of Theorem 2

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

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

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

( $\Leftarrow$ )  $F$  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  $\mathcal{H}$

##### Proof 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  $K_x = K'_x$

##### Proof of Theorem 4

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

( $\Rightarrow$ ) Let  $\mathcal{H}_0$  be the subspace spanned by the functions  $(K_x)_{x \in \mathcal{X}}$ . If  $f = \sum_i a_i K_{x_i}$  and  $g = \sum_j b_j K_{y_j}$ . Let (not an inner product yet)

$$\begin{aligned} \langle f, g \rangle_{\mathcal{H}_0} &= \sum_{i,j} a_i b_j K(x_i, y_j) \\ &= \sum_i a_i g(x_i) \\ &= \sum_j b_j f(y_j) \end{aligned}$$

( $\langle f, g \rangle_{\mathcal{H}_0}$  does not depend on the expansion of  $f$  or  $g$ ) For any  $x \in \mathcal{X}$  and  $f \in \mathcal{H}_0$ ,  $\langle f, K_x \rangle_{\mathcal{H}_0} = f(x)$ .

$$\|f\|_{\mathcal{H}_0}^2 = \sum_{i,j} a_i a_j K(x_i, x_j) \geq 0$$

And since Cauchy-Schwarz is valid,

$$|f(x)| = |\langle f, K_x \rangle_{\mathcal{H}_0}| \leq \|f\|_{\mathcal{H}_0} \cdot K(x, x)^{\frac{1}{2}}$$

Therefore  $\|f\|_{\mathcal{H}_0} = 0 \implies f = 0$ .  $\langle \cdot, \cdot \rangle$  is an inner product on  $\mathcal{H}_0$ .

For a Cauchy sequence  $(f_n)_{n \geq 0}$ ,

$$|f_m(x) - f_n(x)| \leq \|f_m - f_n\|_{\mathcal{H}_0} \cdot K(x, x)^{\frac{1}{2}}$$

For any  $x$  the sequence  $(f_n(x))$  is Cauchy in  $\mathbb{R}$  and therefore converges.

If the functions defined as the point-wise limits of Cauchy sequences are added  $\mathcal{H}_0$ , it becomes a Hilbert space with  $K$  as r.k..

### Proof of Aronszajn's theorem

If  $K$  is p.d. over a set  $\mathcal{X}$ , it is the r.k. of a Hilbert space  $\mathcal{H}$ . The mapping  $\Phi$  is defined by  $\forall x \in \mathcal{X}, \quad \Phi(x) = K_x$ .

By the reproducing property

$$\forall (x, y) \in \mathcal{X}^2, \quad \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} = \langle K_x, K_y \rangle_{\mathcal{H}} = K(x, y)$$

### Proof of the Representer theorem

## A.2 The Kernel Jungle

### Proof