Reproducing Kernel Hilbert Spaces(RKHS)

Roberto Alcover Couso

29/9/2018

1 Introduction

In this document we will study a special type of Hilbert spaces, RKHS, with a kernel which meets the reproducing property. Understanding this is key for our study due to it's relevance in statistical models and the ideas behind algorithims such as RDC,HSIC... Furthermore we will define a homogeneity test based on embeddings of probability distributions on RKHSs, where the distance between distributions corresponds to the distance between their embeddings. We will see that the unit ball of an RKHS is a rich enough space so that the expression for the discrepancy vanishes only if the two probability distributions are equal. At the same time it is restrictive enough for the empirical estimate at the discrepancy to converge quickly to its population counterpart as the sample size increases.

1.1 Preliminar knowledge

1. Feature map: ϕ is known as a feature map if is a function which maps the data to a Hilbert space $\mathcal{H}$ (feature space).

$$\begin{aligned} \varphi : \mathcal{X} &\to \mathcal{H} \\ x &\mapsto \varphi \end{aligned}$$

2. Kernel function: k is called a kernel function if it is the dot product defined on a feature space.

   Then, we can rewrite the dot product of the space in terms of this mapping:

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\to \mathcal{H} \\ (x, x') \mapsto k(x, x') &= < \phi(x), \phi(x') > \end{aligned}$$

3. Reproducing kernel: a function k is a reproducing kernel of the Hilbert space $\mathcal{H}$ is and only if it satisfies:

   (a) k(x,.)$\in \mathcal{H}, \forall x \in \mathcal{X}$

   (b) *Reproducing property*: $< f, k(x, \cdot) >= f(x) \forall f \in \mathcal{H}, \forall x \in \mathcal{X}$

Proposition 1. If k is a reproducting kernel then: k(x,x') = ¡k(x,.),k(x',.)¿

2 Maximum mean discrepancy

In this section it'll be shown how RKHSs can be used to define a homogeneity test in terms of the embeddings of the probability measures. This test consist in maximizing the measure of discrepancy between functions that belong to a certain family $\mathcal{F}$ which must be rich enough to detect all the possible differences between the two probability measures.

2.1 Mean embedding

Lemma 1. Given two Borel probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equal if and only if $\mathbb{E}f(X) = \mathbb{E}f(Y)$ $\forall f \in \mathcal{C}(\mathcal{X})$

$$X \sim \mathbb{P} \text{ and } Y \sim \mathbb{Q}$$

This condition is pretty dificult to prove therefore we will keep our study in order to simplify this evaluation.

Definition 4. MMD Let $\mathcal{F}$ be a class of functions f:$X \to \mathbb{R}$ the MMD based on $\mathcal{F}$ is

$$\gamma(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathcal{F}} \{\mathbb{E}f(X) - \mathbb{E}f(Y)\}$$

This $\mathcal{F}$ must be rich enough for it to ensure that $\mathbb{P} = \mathbb{Q} \leftrightarrow \gamma(\mathbb{P}, \mathbb{Q}) = 0$. And restrictive enough for the empirical estimate to converge quickly as the sample size increases. This will be done through RKHS with a characteristic kernel K

Definition 5. Characteristic kernel A reproducing kernel k is a characterisctic kernel if the induced γ_k is a metric.

Riesz representation theorem If T is a bounded linear operator on a Hilbert space $\mathcal{H}$, then there exist some $g \in \mathcal{H}$ such that $\forall f \in \mathcal{H}$:

$$T(f) =< f, g >_{\mathcal{H}}$$

Lemma 2. Given a K(s,) semi positive definite, measurable and $\mathbb{E}\sqrt{k(X,X)} < \infty$, where X$\sim$ $\mathbb{P}$ then $\mu_p \in \mathcal{H}$ exist and fulfulls the next condition $\mathbb{E}f(X) =< f, \mu_p >$ for all f$\in$ $\mathcal{H}$

Proof Lets define the linear operator $T_{\mathbb{P}}f \equiv \mathbb{E}(\sqrt{k(X,X)}) < \infty \forall f \in \mathcal{H}$

$$|T_{\mathbb{P}}f| =^1 |\mathbb{E}(f(X))| \leq^2 \mathbb{E}(|f(X)|) =^3 \mathbb{E}|< f, k(\cdot, X) >_{\mathcal{H}}| \leq^4$$
$$\|f\|_{\mathcal{H}} \cdot \sqrt{K(X,X)} <^5 \infty$$

Then using the Riesz representation theorem applied to T_p, there exist a $\mu_p \in \mathcal{H}$ such that $T_p f =< f, \mu_p >_{\mathcal{H}}$

Definition 6. Mean embedding Given a probability distribution $\mathbb{P}$ we will define the mean embedding of $\mathbb{P}$ as an element $\mu_p \in \mathcal{H}$ such that

$$\mathbb{E}(f(X)) =< f, \mu_{\mathbb{P}} >_{\mathcal{H}}, \forall f \in \mathcal{H}$$

If $f \in \mathcal{H}$ and $\mu_{\mathbb{P}} \in \mathbb{R}$ $\mathbb{E}(f(X)) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n)$

Applying the Riesz representation theorem to represent $f(x_n)$

$\forall x_n$ then:

$f(x_n) =< f, K(\cdot, x_n) >_{\mathcal{H}}$

then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} < f, K(\cdot, x_n >_{\mathcal{H}} =< f, \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} K(\cdot, x_n) >_{\mathcal{H}}$$

which leads to the final conclussion:

$\mu_{\mathbb{P}} \equiv \mathbb{E}_{X \sim \mathbb{P}}(K(t, X))$ $t \in [0, T]$

Second interpretation of the mean embedding

$$\mu_{\mathbb{P}} = \mathbb{E}(K(\cdot, X))$$

2.2 Introduction to MMD

Lemma 3. Given the conditions of Lemma 2 ($\mu_{\mathbb{P}}$and$\mu_{\mathbb{Q}}$ exist) then:
$X \sim \mathbb{P}\mu_{\mathbb{P}} \equiv \mathbb{E}_{X\sim\mathbb{P}}(K(\cdot,X))$ $Y \sim \mathbb{Q}\mu_{\mathbb{Q}} \equiv \mathbb{E}_{Y\sim\mathbb{Q}}(K(\cdot,Y))$ and:
$MMD(\mathcal{F},\mathbb{P},\mathbb{Q}) = \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}$
Proof

$$MMD \equiv \sup_{f\in\mathcal{H}\ \|f\|\le 1} \{\mathbb{E}(f(x)) - \mathbb{E}(f(y))\}$$
$$= \sup_{f\in\mathcal{H}\ \|f\|\le 1} \{< f, \mu_{\mathbb{P}} > - < f, \mu_{\mathbb{Q}} >\}$$
$$= \sup_{f\in\mathcal{H}\ \|f\|\le 1} < f, (\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}) >$$
$$\le^1 \sup_{f\in\mathcal{H}\ \|f\|\le 1} \{\|f\|_{\mathcal{H}}, \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}\}$$
$$\le \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}$$

But on the other side, if we choose f as:

$$f = \frac{1}{\|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|}(\mu_{\mathbb{P}} - \mu_{\mathbb{Q}})$$

then we have:

$$\sup_{f\in\mathcal{H}\ \|f\|\le 1} \{\|f\|_{\mathcal{H}}, \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}\} \ge \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}$$

therefore

$$MMD = \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|_{\mathcal{H}}$$

Proposition 1
Given:

$$X, X' \sim \mathbb{P} \text{ and } Y, Y' \sim \mathbb{Q}$$

then:

$$MMD^2(\mathcal{F},\mathbb{P},\mathbb{Q}) = \mathbb{E}(K(X,X')) + \mathbb{E}(K(Y,Y')) - 2\mathbb{E}K(X,Y).$$

proof

$$MMD^2(\mathcal{F},\mathbb{P},\mathbb{Q}) = \|\mu_{\mathbb{P}} - \mu_{\mathbb{Q}}\|^2_{\mathcal{H}}$$
$$=< \mu_{\mathbb{P}} - \mu_{\mathbb{Q}}, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} >_{\mathcal{H}}$$
$$=< \mathbb{E}(K(\cdot,X)) - K(\cdot,Y)), \mathbb{E}(K(\cdot,X')) - K(\cdot,Y')) >$$
$$= \mathbb{E}(< K(\cdot,X), K(\cdot,X') > + < K(\cdot,Y), K(\cdot,Y') > -2 < K(\cdot,X)K(\cdot,Y) >)$$
$$=^1 \mathbb{E}(K(X,X') + K(Y,Y') - 2K(X,Y))$$
$$= \mathbb{E}(K(X,X')) + \mathbb{E}(K(Y,Y')) - 2\mathbb{E}(K(X,Y))$$
$$= \int\int K(s,t)\ \underbrace{d(\mathbb{P}-\mathbb{Q})(s)}_{Signedmeasure}\ d(\mathbb{P}-\mathbb{Q})(t)$$

1) is due to the reproductive property of the kernel.

Prooving that MMD defines an homogeneity test

Definition 7. A reproducing kernel k is a characteristic kernel if and only if the induced γ_k is a metric.

Theorem 2. If X is a compact metric space, k is continuous and $\mathcal{H}$ is dense in $\mathcal{C}$(X) with respect to the supremum norm, then $\mathcal{H}$ is characteristic.

Proof. Being characteristic means that $MMD(\mathcal{F}, \mathbb{P}, \mathbb{Q}) = 0 \leftrightarrow \mathbb{P} = \mathbb{Q}$

$\rightarrow$

By lemma 1 we know that $\mathbb{P}$ and $\mathbb{Q}$ are equal if and only if $\mathbb{E}f(X) = \mathbb{E}f(Y)$ $\forall f \in \mathcal{C}(\mathcal{X})$

Given that $\mathcal{H}$ is dense in $\mathcal{C}$(X) then:

$$\forall \epsilon > 0, f \in \mathcal{C}(X), \exists g \in \mathcal{H} : \|f - g\|_\infty < \epsilon$$

$$\begin{aligned}|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| &= |\mathbb{E}(f(X)) - \mathbb{E}(g(X)) + \mathbb{E}(g(X)) - \mathbb{E}(g(Y)) + \mathbb{E}(g(Y)) - \mathbb{E}(f(Y))| \\ &\leq |\mathbb{E}(f(X)) - \mathbb{E}(g(X))| + |\mathbb{E}(g(X)) - \mathbb{E}(g(Y))| + |\mathbb{E}(g(Y)) - \mathbb{E}(f(Y))| \\ &= |\mathbb{E}(f(X)) - \mathbb{E}(g(X))| + |< g, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} >_{\mathcal{H}}| + |\mathbb{E}(g(Y)) - \mathbb{E}(f(Y))| \\ &\leq \mathbb{E}|f(X) - g(X)| + |< g, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} >_{\mathcal{H}}| + \mathbb{E}|g(Y) - f(Y)| \\ &\leq^1 \|f - g\|_\infty + |< g, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} >_{\mathcal{H}}| + \|f - g\|_\infty \\ &\leq |< g, \mu_{\mathbb{P}} - \mu_{\mathbb{Q}} >_{\mathcal{H}}| + 2\epsilon\end{aligned}$$

By lemma 3 we know that if MMD = 0 then $\mu_{\mathbb{P}} = \mu_{\mathbb{Q}}$. Hence:

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(Y))| \leq 2\epsilon$$

Then by lemma 1 $\mathbb{P}$ and $\mathbb{Q}$ are equal.

$\leftarrow$

By definition of MMD.