Reproducing Kernel Hilbert Spaces:

We will consider real Hilbert spaces ~~as~~ vector spaces over $\mathbb{R}$.

def A Hilbert space H is a vector space (over $\mathbb{R}$) together with a function $\langle\cdot,\cdot\rangle : H^2 \to \mathbb{R}$ satisfying

(i) $\forall x,y \in H \quad \langle x,y\rangle = \langle y,x\rangle$

(ii) $\forall x \in H \quad \langle x,x\rangle \geq 0$ and $= 0$ iff $x = 0$

(iii) $\forall r \in \mathbb{R}, x,y \in H \quad \langle rx,y\rangle = r\langle x,y\rangle$

(iv) $\forall x \in H$, define $\|x\| = \sqrt{\langle x,x\rangle}$; then $\|\cdot\|$ satisfies all the axioms for a norm; in particular $\|x+y\| \leq \|x\| + \|y\|$.

(v) $\forall x,y \in H$ define $d(x,y) = \|x-y\|$; then d is a metric on H. The metric space (H,d) is Cauchy complete.

Remark A vector space with a norm satisfying (v) above is called a Banach space. Hilbert spaces are very special Banach spaces.

Example 1 Euclidean vector space $\mathbb{R}^n$ with the usual dot product is a Hilbert space

Example 2 $\quad \ell^2(\mathbb{N}) = \{(a_0, a_1, a_2, \dots) \mid \forall i\ a_i \in \mathbb{R}\ \&\ \sum_{i=0}^{\infty} |a_i|^2 < \infty\}$

This is the prototypical Hilbert space: ∞-dimensional.

Example 3 Consider $\mathbb{R}$ with the usual notion of (Lebesgue) integration which we will write simply as $\int_{-\infty}^{\infty} f\,dx$.

~~$L^2(\mathbb{R}) = \{f$~~

We say f is square-integrable if $\int_{-\infty}^{\infty} |f|^2 dx < \infty$.

We say $f \sim g$ if $\{x \in \mathbb{R} \mid f(x) \neq g(x)\}$ has measure 0. Then $\int f\,dx = \int g\,dx$. Take the set of all these equivalence classes $[f] := \{g \mid f \sim g\}$. We define

$$L^2(\mathbb{R}) = \left\{[f] \,\middle|\, \int_{-\infty}^{\infty} |f|^2 dx < \infty\right\}$$

Often one sees the casual statement "$L^2(\mathbb{R})$ is the collection of square-integrable functions." This is not quite correct.

The inner product on $L^2(\mathbb{R})$ is

$$\langle [f], [g] \rangle = \int f g \, dx$$
$$\|[f]\|^2 := \langle [f], [f] \rangle = \int f^2 dx$$

Note that it does not matter which representative from each equivalence class is chosen. Without carrying out the quotienting one would not have a Hilbert space.

This is one of the most commonly cited examples of a Hilbert space but it is <u>never</u> going to be an RKHS.

— x —

<u>Kernels</u> In SVM data cannot always be separated by a hyperplane. We learned the use of soft margins to deal with the non separable case. Another approach is to use non-linear functions but this has to be done in a tractable way. The idea is to try and identify relevant <u>features</u> of the data and use these features to separate them. Thus we embed the data in a higher dimensional space – the feature space – and try to linearly separate them there. The feature space will be a Hilbert space.

<u>def</u> Given a set X, a <u>kernel</u> over X is a map $k: X \times X \to \mathbb{R}$.

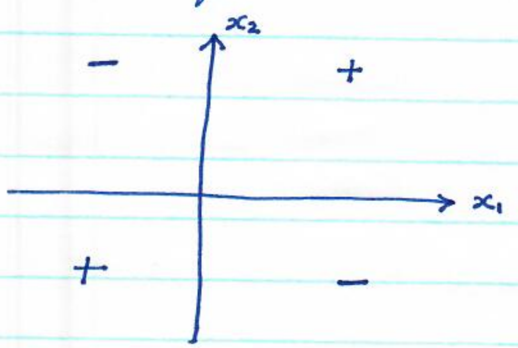
We take our feature space to be some Hilbert space H and we define a map $\Phi: X \to H$. If we consider
$$K(x, x') := \langle \Phi(x), \Phi(x') \rangle$$ we have a kernel.
We will try to define Φ & H cleverly so that K is easy to compute even when computation of Φ & $\langle \cdot, \cdot \rangle$ is expensive.

<u>Example</u> $\Phi : \mathbb{R}^2 \to \mathbb{R}^6$

$$\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}\, x_1 x_2 \\ \sqrt{2}\, x_1 \\ \sqrt{2}\, x_2 \\ 1 \end{pmatrix}$$

$$\begin{aligned} K(\vec{x}, \vec{x}') &= \Phi(\vec{x}) \cdot \Phi(\vec{x}') \\ &= x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_2 x_1' x_2' + 2x_1 x_1' + 2x_2 x_2' + 1 \\ &= (1 + \vec{x} \cdot_2 \vec{x}')^2 \end{aligned}$$

So here we can compute K without embedding into $\mathbb{R}^6$ first or using the inner product of $\mathbb{R}^2$.



These data are not linearly separable in $\mathbb{R}^2$. But it is linearly separable when embedded in $\mathbb{R}^6$.

<u>Example</u> Polynomial kernels: Suppose we are in $\mathbb{R}^n$ so we have variables $x_1, \dots x_n$, let us consider polynomials of degree d.

$$I = \left\{ (k_1, \cdots k_n) \,\middle|\, \sum_{i=1}^{n} k_i \leq d \right\}.$$

Note $|I| \geq \binom{n+d-1}{d}$ [Not obvious!]

$$\Phi : \mathbb{R}^n \to \mathbb{R}^I$$

$\Phi(\vec{x}) \in \mathbb{R}^I$ let $(k_1, \cdots k_n) \in I$ so

$\Phi(\vec{x})(k_1, \cdots k_n) \in \mathbb{R}$ and is given by

$$\sqrt{c(k_1, \dots k_n)}\; x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$$

where $c(k_1, \cdots k_n) = \dfrac{d!}{k_0! k_1! \cdots k_n!}$ where $k_0 = d - \sum_{i=1}^{n} k_i$

The associated kernel is

$$K(\vec{x},\vec{y}) = \Phi(\vec{x})\cdot\Phi(\vec{y}) \quad \text{using} \cdot \text{product of } \mathbb{R}^I$$
$$= \sum_{(k_1\cdots k_n)\in I} c(k_1\cdots k_n)\,(x_1y_1)^{k_1}\cdots(x_ny_n)^{k_n}$$
$$= (1+\vec{x}\cdot\vec{y})^d \quad \text{using} \cdot \text{product of } \mathbb{R}^n$$

The expression $(1+\vec{x}\cdot\vec{y})^d$ is linear in n & d when we count the arithmetic operations. Wheras the dot product in $\mathbb{R}^I$ has $\geq \binom{n+d-1}{d}$ components which is exponential in d.

— x —

How can we design H & Φ? This may be hard, but the general theory of RKHS shows how to construct H & Φ starting from a kernel. So given $k: X\times X\to\mathbb{R}$ we find H & Φ. We will look for kernels with good mathematical properties.

<u>def</u> Fix a set X. A kernel on X is a map $k: X\times X\to\mathbb{R}$ st

(i) $k(x,y) = k(y,x)$ SYMMETRY

(ii) $\forall x_1,\dots x_n\in X$ the matrix $K_{ij}=k(x_i,x_j)$ is positive semidefinite i.e. $\forall v\in\mathbb{R}^n \quad v^TKv\geq 0$

<u>Prop</u> (i) $k(x,x)\geq 0$ (ii) $k(u,v)\leq\sqrt{k(u,u)k(v,v)}$ [CAUCHY-SCHWARTZ for k]

<u>Proof</u> (i) Just take $n=1$ in condition (ii) of the def.

(ii) Take $n=2$. Let $v=\begin{bmatrix}k(v,v)\\ -k(u,v)\end{bmatrix}$ and apply psd

$$K=\begin{pmatrix}k(u,u) & k(u,v)\\ k(v,u) & k(v,v)\end{pmatrix}\geq 0 \Leftrightarrow v^TKv\geq 0$$

i.e. $k(v,v)\left[k(u,u)k(v,v)-k(u,v)^2\right]\geq 0 \underset{\substack{\uparrow\\ \text{Because of (i)}}}{\Longrightarrow} k(u,u)k(v,v)\geq k(u,v)^2.$

Such a matrix is called a Gram matrix.

Now we are ready to define RKHS and build them.

Let X be a set; write $\mathcal{F}(X)$ for the vector space of functions from X to $\mathbb{R}$.

Def A subset $H \subset \mathcal{F}(X)$ is called a reproducing kernel Hilbert space on X if

(i) H is a vector space of $\mathcal{F}(X)$

(ii) H has an inner product making it a Hilbert space

(iii) For every $x \in X$ the linear evaluation map $eval_x : H \to \mathbb{R}$ given by $eval_x(f) = f(x)$ is bounded.

Remark: A bounded linear map is continuous. In fact a linear map on H is bounded if and only if it is continuous.

Remark: Condition (iii) cannot even be stated for $L^2(\mathbb{R})$.

A major theorem in Hilbert space theory is the Riesz representation theorem: For any bounded (= continuous) linear map $\alpha : H \to \mathbb{R}$ there is an element $a \in H$ s.t.

$$\alpha(f) = \langle a, f \rangle.$$

Hence for every $x \in X$ there is a unique $k_x \in H$ s.t.

$$\forall f \in \mathcal{H} \quad \langle k_x, f \rangle = f(x) = eval_x(f)$$

Def k_x is called the reproducing kernel for x. The function $K : X \times X \to \mathbb{R}$ given by $K(x,y) = k_y(x)$ is called the reproducing kernel for H.

Note $K(x,y) = k_y(x) = \langle k_x, k_y \rangle = \langle k_y, k_x \rangle = k_x(y) = K(y,x)$.

Also $\|k_x\|^2 = \langle k_x, k_x \rangle = K(x,x) = k_x(x)$.

If we have $H \subseteq \mathcal{F}(X)$ and $\Phi : X \to H$ $\Phi(x) = k_x$ then $k(x,y) = \langle k_x, k_y \rangle = \langle \Phi(x), \Phi(y) \rangle$ is a positive semidefinite kernel.

How do we know that some kernel we have designed comes from a ~~He~~ RKHS? We write $K(x,\cdot)$ for k_x & note

$$\langle K(x,\cdot), K(y,\cdot)\rangle = K(x,y)$$

<u>Thm</u> Let $K: X\times X \to \mathbb{R}$ be a positive semidefinite kernel. Then there exists a unique Hilbert space $H \subseteq \mathcal{F}(X)$ with K as its reproducing kernel. The subspace $H_0 \subseteq H$ spanned by the set $\{K(x,\cdot) \mid x \in X\}$ is dense in H and H is the completion of H_0 with inner product

$$\langle f, g\rangle_{H_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j K(y_j, x_i)$$

where $f = \sum_{i=1}^{n} \alpha_i K(\cdot, x_i)$ and $g = \sum_{j=1}^{m} \beta_j K(\cdot, y_j)$

So we can write the inner product as

$$\langle f, g\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j f(y_j)\, g(x_i)$$

I will not prove this fundamental result due to MOORE and later to ARONSZAJN.

I will show $\langle f,f\rangle = 0 \Leftrightarrow f = 0$.

~~As~~ $\forall f \in H \quad |f(x)|^2 \geq 0$ hence

$$0 \leq |f(x)|^2 = |\langle K(x,\cdot), f\rangle|^2 \leq \langle K(x,\cdot), K(x,\cdot)\rangle \cdot \langle f, f\rangle \quad \text{CAUCHY SCHWARTZ}$$

So if $\langle f,f\rangle = 0$ we have $|f(x)|^2$ for all $x \in X$

Thus f is indeed the zero function.