

Espacio Pre-Hilbert.: espacio dotado de producto interno con bases ortonormales

$$\{x_k \in \mathbb{H}\}_{k=1}^{\infty} \text{ tal que } \langle x_k, x_l \rangle_{\mathbb{H}} = \begin{cases} 1 & k=l \\ 0 & k \neq l \end{cases}$$

Desde las bases, se pueden representar vectores x (que pueden o no $\in \mathbb{H}$), como:

$$x = \sum_{k=1}^{\infty} a_k x_k$$

abarcados (spanned) por $\{x_k \in \mathbb{H}\}_{k=1}^{\infty}$, $a_k \in \mathbb{R}$.

Sea el vector $y_n = \sum_{k=1}^n a_k x_k$; la distancia Euclídea entre y_n y y_m ($n > m$) se define:

$$\begin{aligned} \|y_n - y_m\|_{\mathbb{H}}^2 &= \left\| \sum_{k=1}^n a_k x_k - \sum_{k=1}^m a_k x_k \right\|_{\mathbb{H}}^2 \\ &= \left\| \sum_{k=1}^m a_k x_k + \sum_{k=m+1}^n a_k x_k - \sum_{k=1}^m a_k x_k \right\|_{\mathbb{H}}^2 \\ &= \left\| \sum_{k=m+1}^n a_k x_k \right\|_{\mathbb{H}}^2 = \left\langle \sum_{k=m+1}^n a_k x_k, \sum_{k=m+1}^n a_k x_k \right\rangle_{\mathbb{H}} \\ \|y_n - y_m\|_{\mathbb{H}}^2 &= \sum_{k=m+1}^n \sum_{l=m+1}^n a_k a_l^* \langle x_k, x_l \rangle_{\mathbb{H}} = \sum_{k=m+1}^n |a_k|^2 \end{aligned}$$

Para que la definición de $x = \sum_{k=1}^{\infty} a_k x_k$ sea coherente:

$$1. \sum_{k=m+1}^n |a_k|^2 \rightarrow 0 \quad (n, m \rightarrow \infty)$$

$$2. \sum_{k=1}^{\infty} |a_k|^2 < \infty$$

$$\Rightarrow \{y_n = \sum_{k=1}^n a_k x_k \in \mathbb{H}\}_{n=1}^{\infty}$$

Secuencia de Cauchy.

→ Un vector \mathbf{x} se puede representar desde las bases $\{\mathbf{x}_k \in \mathcal{H}\}_{k=1}^{\infty}$ si y solo si \mathbf{x} es una combinación lineal de las bases y sus coeficientes asociados $\{a_k \in \mathbb{R} \cup \mathbb{C}\}_{k=1}^{\infty}$ se pueden sumar en magnitud cuadrática (converge)

Pre-Hilbert \mathcal{H} → Hilbert \mathcal{H}

normado
producto punto

normado
producto punto
completo (secuencia de Cauchy converge)

Mercer Kernel (Núcleo de Mercer):

→ función continua, simétrica y definida positiva

$\mathcal{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{X} : dominio de entrada

$$\mathcal{X} \subset \mathbb{R}^p$$

$$\text{Ej: } \mathcal{K}(\mathbf{x}, \mathbf{x}') = A \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}'\|_{\mathcal{X}}^2\right) = A \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_{\mathcal{X}}^2}{2\sigma^2}\right)$$

$\sigma = \frac{1}{2\sigma^2} \in \mathbb{R}$ precisión; σ : escala (ancho del kernel) Gaussiano

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = (\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathcal{X}} + 1)^L \quad L \in \mathbb{Z}.$$

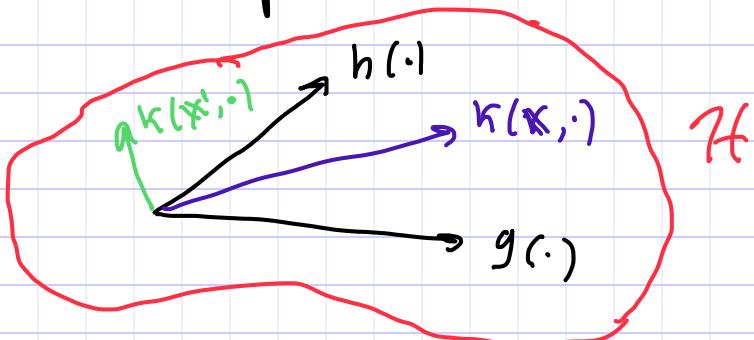
POLINOMIAL.

$$\mathbf{x}, \mathbf{x}' \in \mathcal{X}$$

Reproducing Kernel Hilbert Space (RKHS)

Espacio de Hilbert con Kernel reproductivo.

Sea \mathcal{H} un espacio vectorial de funciones con rango real (real-valued functions) de \mathbb{X} generadas por el kernel (núcleo) $k(x, \cdot)$.



Sean $h, g \in \mathcal{H}$ dos funciones en RKHS. \mathcal{H} , representadas como :

$$h = \sum_{i=1}^l a_i k(x_i, \cdot)$$

$$g = \sum_{j=1}^m b_j k(x_j, \cdot) ; \quad a_i, b_j \in \mathbb{R}$$

$$x_i, x_j \in \mathbb{X}$$

La forma bilineal (producto punto entre h y g):

$$\langle h, g \rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^l a_i k(x_i, \cdot), \sum_{j=1}^m b_j k(x_j, \cdot) \right\rangle_{\mathcal{H}}$$

$$= \sum_{i=1}^l \sum_{j=1}^m a_i b_j k(x_i, x_j)$$

Propiedades de la forma bilineal en \mathcal{H} :

1. Simetría $\langle h, g \rangle = \langle g, h \rangle$

2. Escalamiento y propiedad distributiva:

$$\langle (cf + dg), h \rangle_{\mathcal{H}} = c \langle f, h \rangle_{\mathcal{H}} + d \langle g, h \rangle_{\mathcal{H}}$$

3. Norma $\|f\|^2 = \langle f, f \rangle_{\mathcal{H}}, 0$

Por ende, la forma bilineal es un producto interno.

Ahora, si $g(\cdot) = k(x, \cdot)$

$$\langle h, g \rangle_{\mathcal{H}} = \langle h, k(x, \cdot) \rangle_{\mathcal{H}} = \sum_{i=1}^l a_i k(x, x_i).$$

Note que si $h(\cdot) = k(x, \cdot)$; $g(\cdot) = k(x', \cdot)$

$$\langle h, g \rangle_{\mathcal{H}} = \langle k(x, \cdot), k(x', \cdot) \rangle_{\mathcal{H}} = k(x, x')$$

y si $h(\cdot) = \sum_{i=1}^l a_i k(x_i, \cdot)$ y $g(\cdot) = k(x, \cdot)$

$$\langle h, g \rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^l a_i k(x_i, \cdot), k(x, \cdot) \right\rangle_{\mathcal{H}}$$

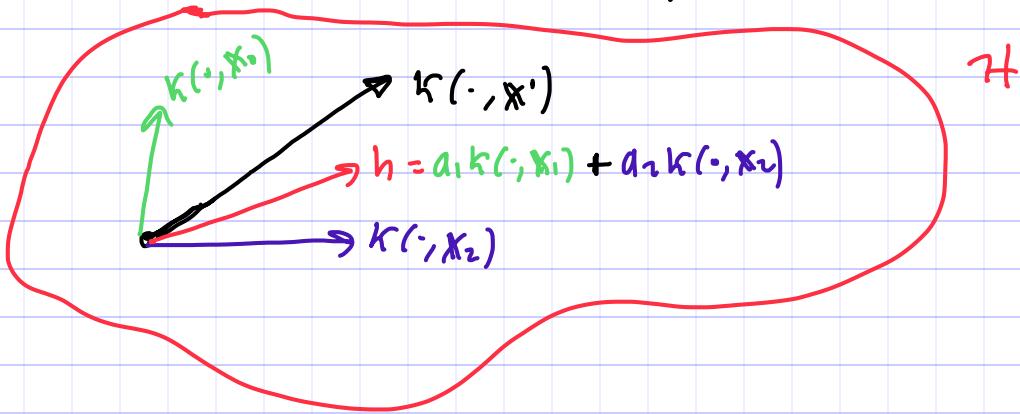
$$\langle h, g \rangle_{\mathcal{H}} = \sum_{i=1}^l a_i \langle k(x_i, \cdot), k(x, \cdot) \rangle_{\mathcal{H}} = \sum_{i=1}^l a_i k(x, x_i)$$

Reproducing property $\Rightarrow, g(x) = h(x)$

$k(x, x')$ → kernel reproductivo

→ Para un $x \in \mathcal{X}$, $k(\cdot, x) \in \mathcal{H}$

→ $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ satisface la propiedad reproductiva.



$$h(x) = \langle h, k(\cdot, x') \rangle_{\mathcal{H}} = \langle a_1 k(\cdot, x_1) + a_2 k(\cdot, x_2), k(\cdot, x') \rangle_{\mathcal{H}}$$

$$h(x) = a_1 k(x_1, x') + a_2 k(x_2, x')$$

Teorema de Mercer: $k(x, x')$ se puede expresar como:

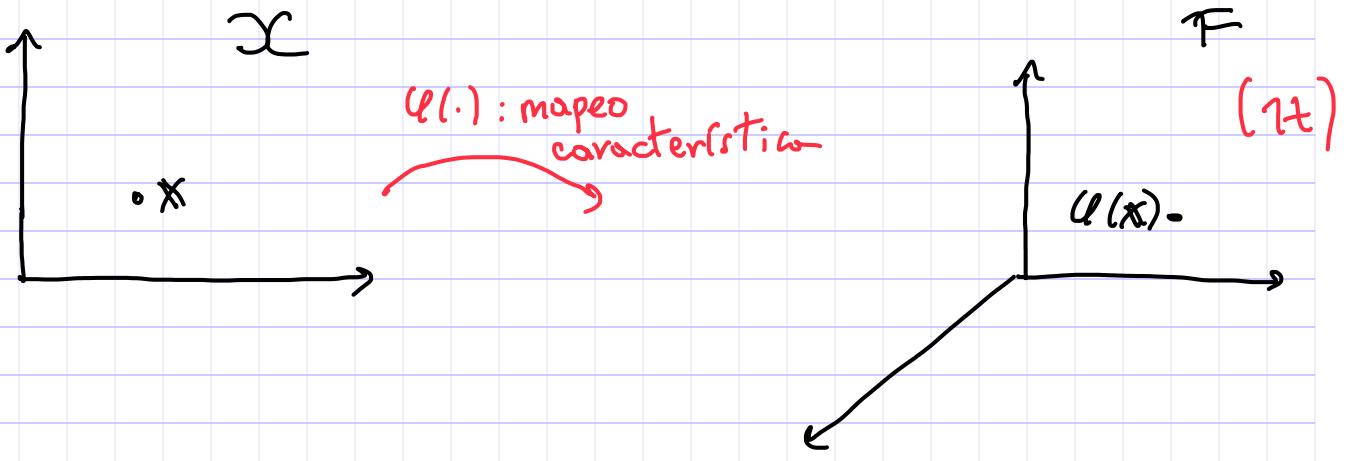
$$k(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x')$$

$\lambda_i \in \mathbb{R} \cup \mathbb{C}$ → valores propios; $\phi_i \in \mathbb{R} \cup \mathbb{C}$ → funciones propias

$$\varphi: \mathcal{X} \rightarrow \mathcal{F}$$

$$\varphi(x) = [\sqrt{\lambda_1} \phi_1(x), \sqrt{\lambda_2} \phi_2(x), \dots]$$

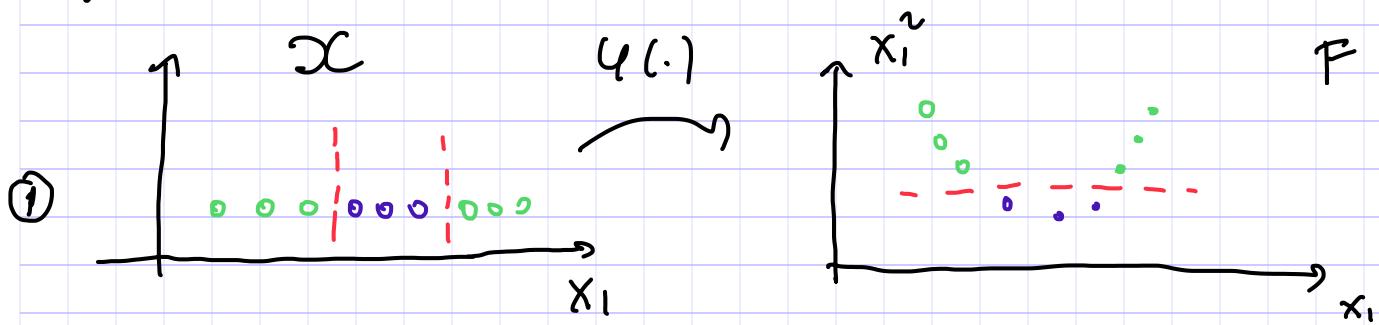
→ La dimensión de \mathcal{F} se determina por el número de valores propios positivos → pueden ser ∞ para el kernel Gaussiano.



→ F es el RKHS inducido por k , con

$$q(x) = k(x, \cdot) \rightarrow \text{las bases de } F \text{ y } H$$

Ejemplos:



$$\textcircled{2} \quad k(x, c) = (1 + x^T c)^2; \quad x = [x_1, x_2]; \quad c = [c_1, c_2].$$

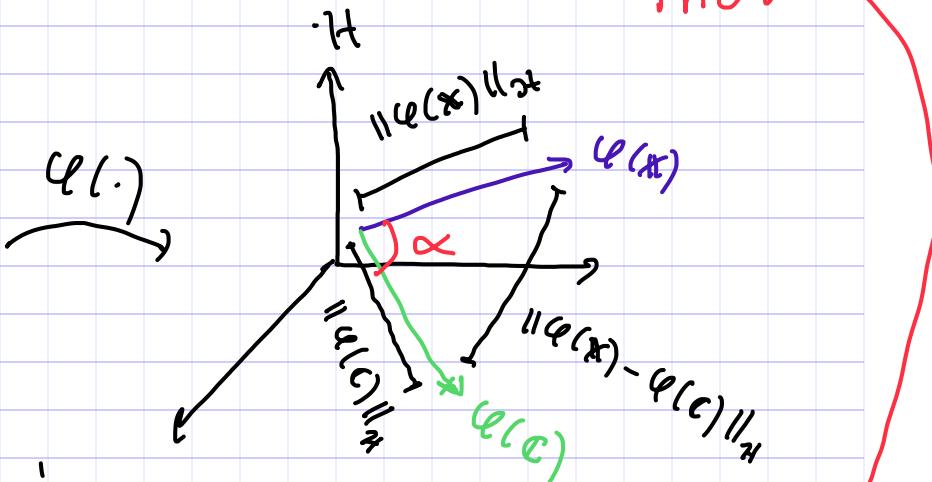
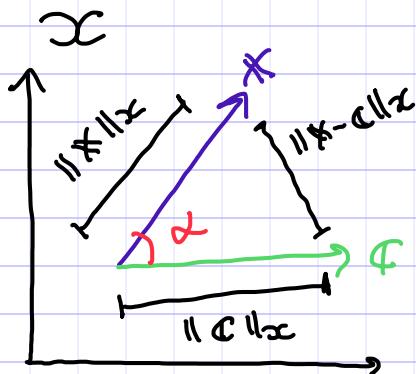
$$\begin{aligned} k(x, c) &= \left(1 + [x_1, x_2] \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}\right)^2 = \left(1 + c_1 x_1 + c_2 x_2\right)^2 \\ &= (1 + c_1 x_1)^2 + 2(1 + c_1 x_1)(c_2 x_2) + (c_2 x_2)^2 = 1 + 2c_1 x_1 + c_1^2 x_1^2 + c_2^2 x_2^2 \\ &\quad + 2(c_2 x_2 + c_1 c_2 x_1 x_2) \\ &= c_1^2 x_1^2 + c_2^2 x_2^2 + 2c_1 x_1 + 2c_2 x_2 + 2c_1 c_2 x_1 x_2 + 1 \end{aligned}$$

$$\begin{aligned} \text{Así: } q(x) &= [1, x_1^2, \sqrt{2} x_1 x_2, x_2^2, \sqrt{2} x_1, \sqrt{2} x_2]^T \\ q(c) &= [1, c_1^2, \sqrt{2} c_1 c_2, c_2^2, \sqrt{2} c_1, \sqrt{2} c_2]^T \end{aligned}$$

$$\text{Por ende: } K(x, c) = (1 + x^T c)^2 = \langle \varphi(x), \varphi(c) \rangle_H = \varphi(x)^T \varphi(c).$$

Ilustración geométrica del RKFHS.

"Kernel Trick"



$$\|x\|_x^2 = \langle x, x \rangle_x$$

$$\|\varphi(x)\|_H^2 = \langle \varphi(x), \varphi(x) \rangle_H = k(x, x)$$

$$\|x - c\|_x = \langle x - c, x - c \rangle_x$$

$$\|\varphi(x) - \varphi(c)\|_H^2 = \langle \varphi(x) - \varphi(c), \varphi(x) - \varphi(c) \rangle_H$$

$$= \langle x, x \rangle - 2 \langle x, c \rangle + \langle c, c \rangle$$

$$= \|\varphi(x)\|_H^2 - 2 \langle \varphi(x), \varphi(c) \rangle_H + \|\varphi(c)\|_H^2$$

$$= \|x\|_x^2 - 2 \langle x, c \rangle_x + \|c\|_x^2$$

$$= k(x, x) - 2 k(x, c) + k(c, c)$$

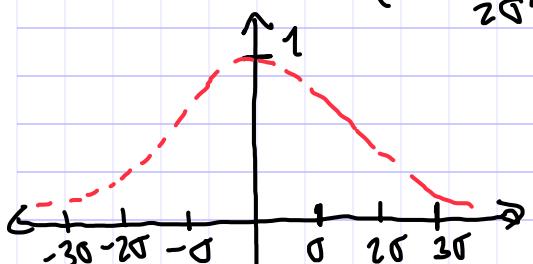
$$\cos(\alpha) = \frac{\langle x, c \rangle_x}{\|x\|_x \|c\|_x}$$

$$\cos(\alpha) = \frac{\langle \varphi(x), \varphi(c) \rangle_H}{\|\varphi(x)\|_H \|\varphi(c)\|_H}$$

$$\cos(\alpha) = \frac{k(x, c)}{\sqrt{k(x, x) k(c, c)}}$$

Ej: Kernel Gaussiano normalizado

$$k(x, c) = \exp\left(-\frac{\|x - c\|_2^2}{2\sigma^2}\right); \quad \|\varphi(x) - \varphi(c)\|_H^2 = 1 - 2k(x, c) + 1$$



$$\delta(\varphi(x), \varphi(c)) = \sqrt{2(1 - k(x, c))}$$

$$\cos(\alpha) = k(x, c)$$

NOTA: $\varphi(\cdot)$ NO SÉ CONOCE EN KERNEL GAUSSIANO.

Aplicaciones del RKHS en métodos clásicos basados en riesgo empírico.

1. Regresión rígida Kernel:

$$y_i \in \mathbb{R}^N; \quad X \in \mathbb{R}^{N \times p}; \quad \varphi: \mathbb{R}^p \rightarrow \mathcal{H}; \quad \mathcal{H} \subseteq \mathbb{R}^Q; \quad Q \rightarrow \infty.$$

$$\hat{y}_i = f(\varphi(x) | w) = \phi(x)w; \quad f: \mathbb{R}^Q \rightarrow \mathbb{R}; \quad w \in \mathcal{H} \subseteq \mathbb{R}^Q.$$

$$w^* = \arg \min_w L(y_i, f(\varphi(x) | w)) + R(f) \lambda$$

$$= \arg \min_w \frac{1}{N} \|y_i - \phi(x)w\|_2^2 + \lambda \|w\|_2^2$$

$\lambda \in \mathbb{R}^+$: parámetro de regularización (fijo = rígido).

$$\begin{aligned} \|y_i - \phi(x)w\|_2^2 &= \langle y_i - \phi(x)w, y_i - \phi(x)w \rangle \\ &= y_i^T y_i - 2 y_i^T \phi(x)w + (\phi(x)w)^T (\phi(x)w) \\ &= y_i^T y_i - 2 y_i^T \phi(x)w + w^T \phi^T(x) \phi(x)w \end{aligned}$$

NOTA: $\phi(x) = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_Q(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_Q(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_N) & \varphi_2(x_N) & \cdots & \varphi_Q(x_N) \end{bmatrix}_{N \times Q}.$

$$\frac{1}{N} \phi(x)^T \phi(x) \in \mathbb{R}^{Q \times Q} \approx \mathbb{E} \{\phi(x) \phi(x)^T\} = R_{\varphi(x) \varphi(x)}$$

CORRELACIÓN EN RKHS; $\underline{Q \rightarrow \infty}$

NO SIEMPRE SE PUEDE CALCULAR.

$$\frac{\partial}{\partial w} \left\{ L(y, f) + R(f|x) \right\} = -\frac{2}{N} \underset{Q \times N \times N \times 1}{\phi^T(x)} y_1 + \frac{2}{N} \underset{Q \times Q \times 1}{\phi^T(x) \phi(x)} w + 2 \lambda w$$

$$\nabla_w = \frac{2}{N} \left[(\phi^T(x) \phi(x) + N\lambda) w - \phi^T(x) y \right]$$

Solución analítica:

$$\frac{2}{N} \left[(\phi^T(x) \phi(x) + N\lambda I) w \right] = \frac{2}{N} \phi^T(x) y.$$

$$w^* = \left[\phi^T(x) \phi(x) + N\lambda I \right]^{-1} \phi^T(x) y$$

NOTA: $\phi^T(x) \phi(x) \in \mathbb{R}^{Q \times Q}$; $Q \rightarrow \infty$.

Propiedades de matriz inversa:

$$\begin{aligned} (I + AB)^{-1} A &= A (I + BA)^{-1} \\ (\phi^T \phi + N\lambda I)^{-1} \phi^T &= \left[N \lambda \left(\frac{1}{N} \phi^T \phi + I \right) \right]^{-1} \phi^T \\ &= \frac{1}{N\lambda} \left(I + \frac{\phi^T \phi}{N\lambda} \right)^{-1} \phi^T = \phi^T \frac{1}{N\lambda} \left(I + \frac{\phi^T \phi}{N\lambda} \right)^{-1} \end{aligned}$$

$$= \phi^T \left[N \lambda \left(I + \frac{\phi^T \phi}{N\lambda} \right) \right]^{-1} = \phi^T \left[N \lambda I + \phi \phi^T \right]^{-1}$$

$$w^* = \phi^T(x) \left[N \lambda I + \phi(x) \phi^T(x) \right]^{-1} y$$

$$\begin{aligned}\hat{y} &= f(\varphi(x_{\text{new}}) | w) = \varphi^T(x_{\text{new}}) w \\ &= \varphi^T(x_{\text{new}}) \Phi^T(X) [N\lambda I + \Phi(X) \Phi^T(X)]^{-1} w\end{aligned}$$

$\varphi(x_{\text{new}}) \in \mathbb{R}^Q$.

$$\Phi(X) = [\varphi(x_1) \quad \varphi(x_2) \quad \dots \quad \varphi(x_N)]^T_{N \times Q}$$

$$\Phi(X) \Phi^T(X) = \begin{bmatrix} \varphi^T(x_1) \\ \varphi^T(x_2) \\ \vdots \\ \varphi^T(x_N) \end{bmatrix}_{N \times Q} [\varphi(x_1) \quad \varphi(x_2) \quad \dots \quad \varphi(x_N)]_{Q \times N}$$

$$K \in \mathbb{R}^{N \times N} = \begin{bmatrix} \varphi^T(x_1) \varphi(x_1) & \varphi^T(x_1) \varphi(x_2) \dots & \varphi^T(x_1) \varphi(x_N) \\ \varphi^T(x_2) \varphi(x_1) & \varphi^T(x_2) \varphi(x_2) \dots & \varphi^T(x_2) \varphi(x_N) \\ \vdots & \vdots & \ddots \end{bmatrix}_{N \times N}$$

$$\begin{bmatrix} \varphi^T(x_N) \varphi(x_1) & \varphi^T(x_N) \varphi(x_2) & \varphi^T(x_N) \varphi(x_N) \end{bmatrix}_{N \times N}$$

$$K = \left[K(x_n, x_m) \right]_{n,m=1}^N \Rightarrow \text{POR KERNE TRICK}.$$

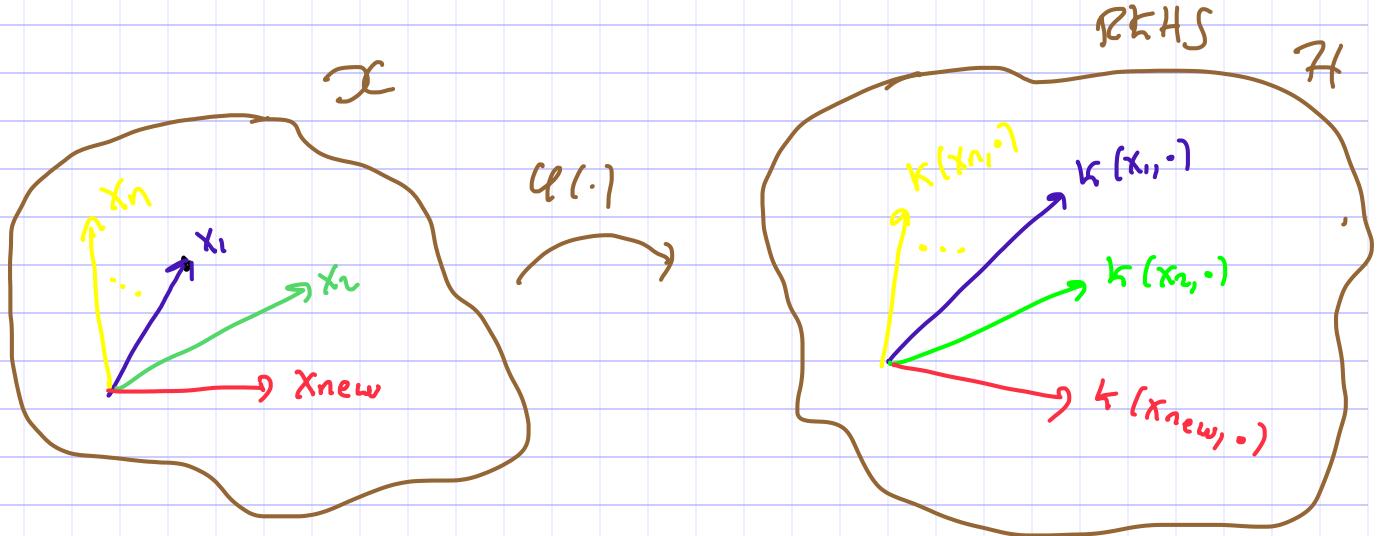
$$\varphi^T(x_{\text{new}}) \Phi^T(X) = \begin{bmatrix} \varphi^T(x_{\text{new}}) \end{bmatrix}_{1 \times Q} \begin{bmatrix} \varphi(x_1) \quad \varphi(x_2) \dots \quad \varphi(x_N) \end{bmatrix}_{Q \times N}$$

$$K_{\text{new}}^T \in \mathbb{R}^{1 \times N} = \left[K(x_{\text{new}}, x_n) \right]_{n=1}^N$$

$$\hat{y} = f(x_{\text{new}}) = \|k_{\text{new}}^T \left(N \lambda I + K \right)^{-1} y \|_{N \times N \times N \times 1}$$

$$\hat{y} = \|k_{\text{new}}^T \alpha\| ; \quad \alpha \in \mathbb{R}^{N \times 1}$$

$$\hat{y} = \sum_{n=1}^N \alpha_n k(x_{\text{new}}, x_n) = \langle f, \varphi(x_{\text{new}}) \rangle_H = \langle f, k(x_{\text{new}}, \cdot) \rangle_H$$



$$f(\cdot) = \sum_{n=1}^N \alpha_n k(x_n, \cdot) ; \quad \varphi(\cdot) = k(x_{\text{new}}, \cdot)$$

$$f(x_{\text{new}}) = \langle f, \varphi \rangle_H = \left\langle \sum_{n=1}^N \alpha_n k(x_n, \cdot), k(x_{\text{new}}, \cdot) \right\rangle_H$$

$$f(x_{\text{new}}) = \sum_{n=1}^N \alpha_n k(x_n, x_{\text{new}}) \rightarrow \text{Reproducing property.}$$

NOTA: $\alpha_n \in \mathbb{R}$ se encuentra mediante min. cuadrado regularizado en RKHS.

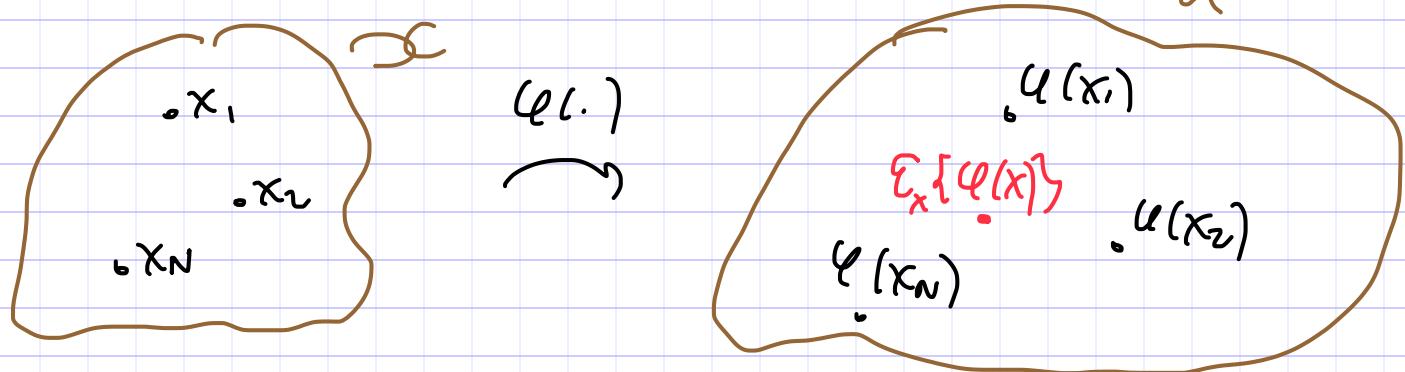
2. Centered Kernel Alignment - CKA

Alineamiento centrado kernel.

Sea $\varphi: \mathcal{X} \rightarrow \mathcal{H}$ función de mapeo no lineal con kernel asociado $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$.

El mapeo centrado $\ell(x) = \varphi(x) - \mathbb{E}_x\{\varphi(x)\}$, se obtiene desde la función de distribución P_x ; $x \sim P_x$.

$$\mathbb{E}_x\{\ell(x)\} = \int \ell(x) dP_x$$



$k_c: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ se define como:

$$\begin{aligned} k_c(x, x') &= \langle \varphi(x) - \mathbb{E}_x\{\varphi(x)\}, \varphi(x') - \mathbb{E}_{x'}\{\varphi(x')\} \rangle_{\mathcal{H}} \\ &= \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}} - \langle \varphi(x), \mathbb{E}_{x'}\{\varphi(x')\} \rangle_{\mathcal{H}} \\ &\quad - \langle \mathbb{E}_x\{\varphi(x)\}, \varphi(x') \rangle_{\mathcal{H}} + \langle \mathbb{E}_x\{\varphi(x)\}, \mathbb{E}_{x'}\{\varphi(x')\} \rangle_{\mathcal{H}} \end{aligned}$$

$$\begin{aligned} k_c(x, x') &= k(x, x') - \mathbb{E}_{x'}\{\langle \varphi(x), \varphi(x') \rangle\} - \mathbb{E}_x\{\langle \varphi(x), \varphi(x') \rangle\} \\ &\quad + \mathbb{E}_{x, x'}\{\langle \varphi(x), \varphi(x') \rangle\} \end{aligned}$$

$$k_c(x, x') = k(x, x') - \mathbb{E}_x \{ k(x, x') \} - \mathbb{E}_{x'} \{ k(x, x') \} \\ + \mathbb{E}_{x, x'} \{ k(x, x') \}.$$

NOTA: $\mathbb{E}_{x, x'} \{ k(x, x') \} = \iint_{x \in X, x' \in X} k(x, x') dP_x dP_{x'}$

Si P_x admite función de densidad $dP_x = p(x) dx$.

$$\mathbb{E}_{x, x'} \{ k(x, x') \} = \iint_{x \in X, x' \in X} k(x, x') p(x) p(x') dx dx'$$

Estimación del CKA:

Sea un set de datos $\{x_n \in \mathbb{R}^p\}_{n=1}^N$;

con $x_n \sim U(X) = \frac{1}{N}$ ($p(x) = U(x/N)$);

$$\mathbb{E}_x \{ \ell(x_n) \} = \frac{1}{N} \sum_{n=1}^N \ell(x_n).$$

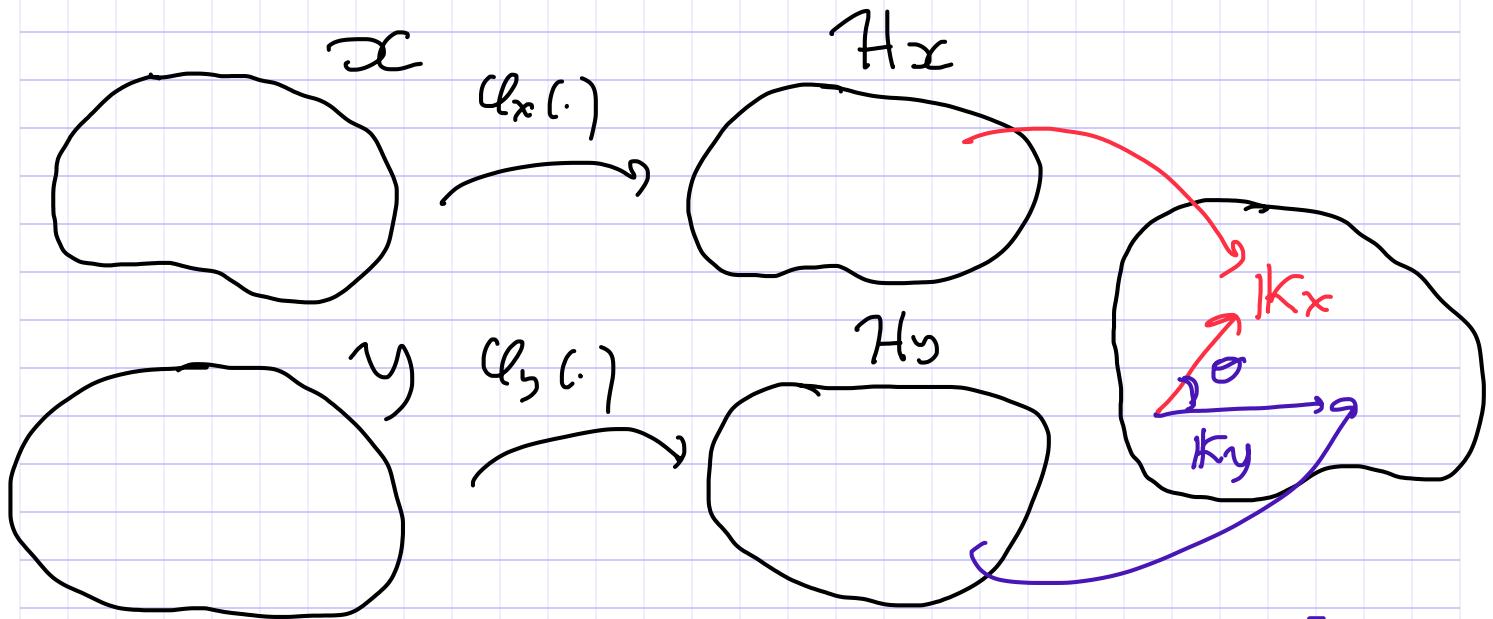
$$k_c(x_n, x_m) = k(x_n, x_m) - \frac{1}{N} \sum_{n=1}^N k(x_n, x_m) - \frac{1}{N} \sum_{m=1}^N k(x_n, x_m) \\ + \frac{1}{N^2} \sum_{n,m=1}^N k(x_n, x_m).$$

Por ende la matriz $K_c \in \mathbb{R}^{N \times N} = [k_c(x_n, x_m)]_{n,m=1}^N$

$$\text{Así: } \|K_c\| = \|H\| \|K\| \|H\| ; \quad \|H\| = I - \frac{11^T}{N}$$

TAREA: DEMOSTRAR ↗

Función de alineamiento



$$f(\bar{K}_x, \bar{K}_y) = \cos(\theta) = \frac{\langle \bar{K}_x, \bar{K}_y \rangle_F}{\|\bar{K}_x\|_F \|\bar{K}_y\|_F}$$

$\bar{K}_x, \bar{K}_y \in \mathbb{R}^{N \times N}$: versiones centradas
de K_x y K_y .

$$f(K_x, K_y) = \frac{\sum \{ \bar{K}_x \bar{K}_y \}}{\sqrt{\sum \{ \bar{K}_x^2 \} \sum \{ \bar{K}_y^2 \}}}$$

Por la igualdad de Cauchy-Schwarz:

$$|\varepsilon\{\bar{K}_x \bar{K}_y\}| \leq \sqrt{\varepsilon\{\bar{K}_x^2\} \varepsilon\{\bar{K}_y^2\}}$$

$$\hat{f}(K_x, K_y) \in [0, 1], \quad \|K\|_F^2 = \text{tr}(K K^*) \\ = \text{tr}(K^* K)$$

Algoritmos desde CKA:

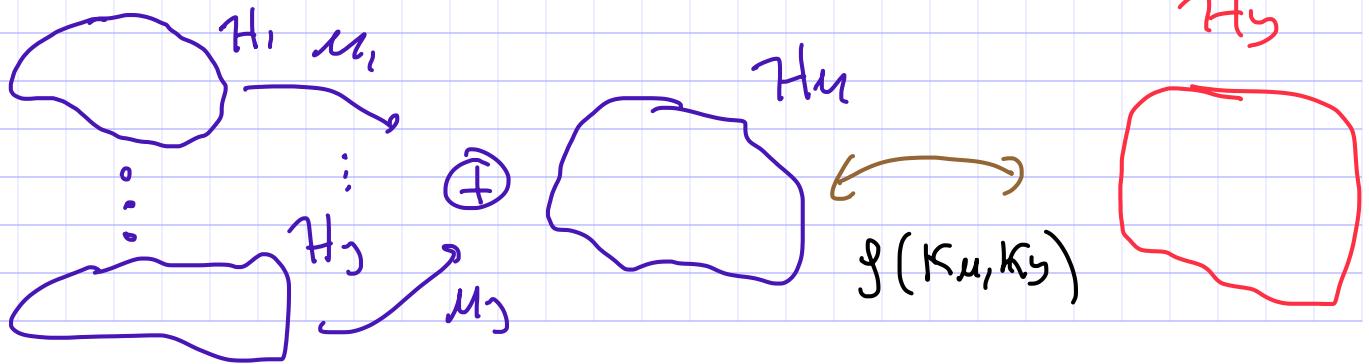
a. Alineamiento independiente:

$$\left\{ K_j \in \mathbb{R}^{N \times N} \right\}_{j=1}^J, \quad K_y \in \mathbb{R}^{N \times N} \rightarrow \text{Kernel objetivo.}$$

$$K_M \propto \sum_{j=1}^J \mu_j K_j, \quad \mu_j = \hat{f}(K_j, K_y)$$

$$K_M = \sum_{j=1}^J \hat{f}(K_j, K_y) K_j = \sum_{j=1}^J \frac{\langle \bar{K}_j, \bar{K}_y \rangle_F}{\| \bar{K}_j \|_F \| \bar{K}_y \|_F} K_j$$

$$K_M = \frac{1}{\| \bar{K}_y \|_F} \sum_{j=1}^J \frac{\langle \bar{K}_j, \bar{K}_y \rangle_F}{\| \bar{K}_j \|_F} K_j$$



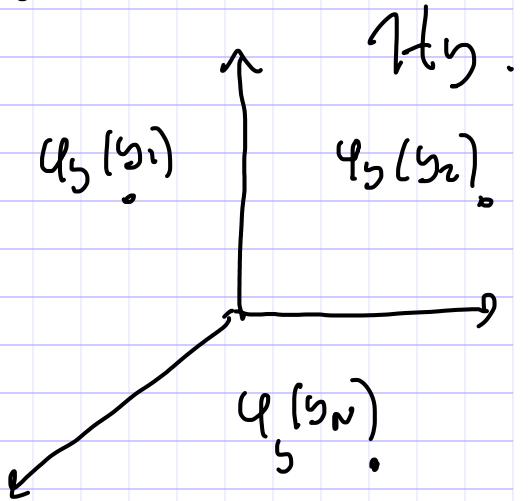
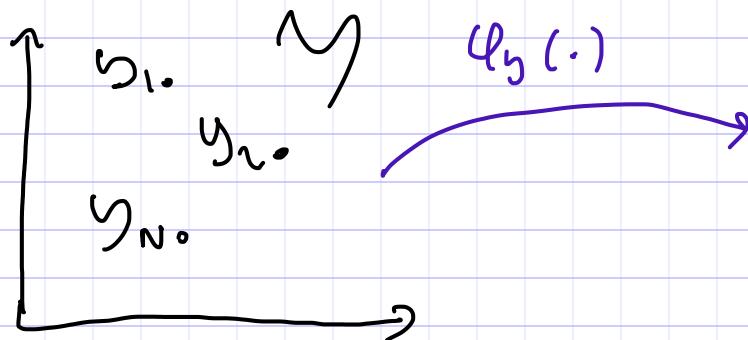
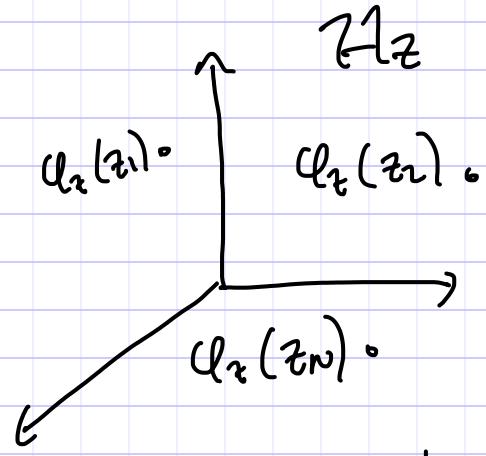
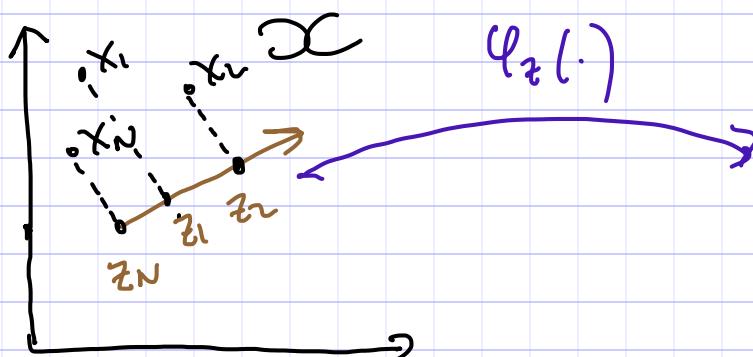
b. Alineamiento restringido

$$\max_{\mathbf{M}} \hat{f} \left(\sum_{j=1}^J \mu_j \|K_j, K_y\| \right)$$

$$\text{s.t. } \|\mathbf{M}\|_q^q \leq \varsigma$$

TAREA: Resolver por Lagrange para $q=2$
 $\Rightarrow \|\mathbf{M}\|_2 = 1, \quad M > 0.$
(combinación convexa).

c. Aprendizaje de métrica. (Metric Learning).



$$\mathcal{F}(K_z, K_y) = \frac{\langle K_z, K_y \rangle_p}{\|K_z\|_p \|K_y\|_p}$$

S; $z = XA$; $X \in \mathbb{R}^P$; $A \in \mathbb{R}^{P \times Q}$; $Q \leq P$.
 $z \in \mathbb{R}^Q$; $y \in \mathbb{Y}$.

$$S; K_z(z_n, z_m) = \exp(-\gamma \|z_n - z_m\|_2^2)$$

$$K_z(X_n A, X_m A) = \exp(-\gamma \|X_n A - X_m A\|_2^2)$$

$$= \exp(-\gamma \langle X_n A - X_m A, X_n A - X_m A \rangle)$$

$$d_z^2(z_n, z_m) = \|z_n - z_m\|_2^2 = \langle X_n A - X_m A, X_n A - X_m A \rangle$$

$$= (X_n A - X_m A) \quad (X_n A - X_m A)^T$$

$$= (X_n - X_m) \underset{1 \times P}{A} \underset{P \times Q}{((X_n A)^T - (X_m A)^T)} \underset{Q \times 1}{}$$

$$= (X_n - X_m) A (A^T X_n^T - A^T X_m^T)$$

$$d_z^2(z_n, z_m) = (X_n - X_m) A A^T (X_n - X_m)^T = d_x^2(x_n, x_m) \underset{A = A^T}{\bar{z}}$$

\downarrow
Euclidia
en z

\downarrow
Mahalanobis en
 x ,

$$A^* = \arg \max_{(A)} \mathcal{F}(K_z(A), K_y)$$

$$\text{s.t. } \|A\|_q < \lambda,$$