Espacio Pre-Hilbert: espacio dotado de producto interno con bases ortonormales $\{x_k \in \mathcal{H}\}_{k=1}^{\infty}$ tal que $\langle x_k, x_\ell \rangle_{\mathcal{H}} = \begin{cases} 1 & k=\ell \\ 0 & k\neq \ell \end{cases}$

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

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

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

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:

$$\|y_n - y_m\|_{\mathcal{H}}^2 = \left\| \sum_{k=1}^{n} a_k x_k - \sum_{k=1}^{m} a_k x_k \right\|_{\mathcal{H}}^2$$

$$= \left\| \sum_{k=1}^{m} a_k x_k + \sum_{k=m+1}^{n} a_k x_k - \sum_{k=1}^{m} a_k x_k \right\|_{\mathcal{H}}^2$$

$$= \left\| \sum_{k=m+1}^{n} a_k x_k \right\|_{\mathcal{H}}^2 = \left\langle \sum_{k=m+1}^{n} a_k x_k, \sum_{\ell=m+1}^{n} a_\ell x_\ell \right\rangle_{\mathcal{H}}$$

$$\|y_n - y_m\|_{\mathcal{H}}^2 = \sum_{k=m+1}^{n} \sum_{\ell=m+1}^{n} a_k a_\ell^* \langle x_k, x_\ell \rangle_{\mathcal{H}} = \sum_{k=m+1}^{n} |a_k|^2$$

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

1. $\sum_{k=m+1}^{n} a_k^2 \to 0 \quad (n,m \to \infty)$ $\Longrightarrow$ $\{y_n = \sum_{k=1}^{n} a_k x_k \in \mathcal{H}\}_{n=1}^{\infty}$

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

Secuencia de Cauchy.

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

Pre-Hilbert $\mathcal{H}$ ⟶ Hilbert $\mathcal{H}$

Pre-Hilbert $\mathcal{H}$	Hilbert $\mathcal{H}$
normado	normado
producto punto	producto punto
	Completo (secuencia de Cauchy converge)

Mercer Kernel (Núcleo de Mercer):

→ función continua, simétrica y definida positiva

$\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. $\mathcal{X}$: domino de entrada

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

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

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

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

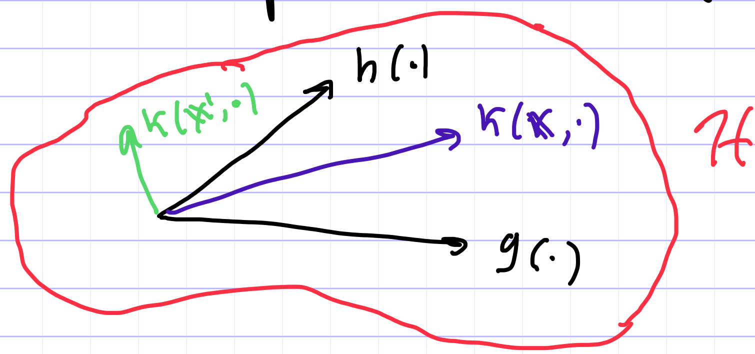
POLINOMIAL.

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

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 $\mathcal{X}$ generadas por el kernel (núcleo) $\kappa(\mathbf{x},\cdot)$.



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

$$h = \sum_{i=1}^{\ell} a_i \kappa(\mathbf{x}_i, \cdot)$$

$$g = \sum_{j=1}^{m} b_j \kappa(\mathbf{x}_j, \cdot) \quad ; \quad a_i, b_j \in \mathbb{R}$$

$$\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$$

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

$$\langle h, g \rangle_{\mathcal{H}} = \left\langle \sum_{i=1}^{\ell} a_i \kappa(\mathbf{x}_i, \cdot), \sum_{j=1}^{m} b_j \kappa(\mathbf{x}_j, \cdot) \right\rangle_{\mathcal{H}}$$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{m} a_i b_j \kappa(\mathbf{x}_i, \mathbf{x}_j)$$

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 L_2

$$\| f \|^2 = \langle f, f \rangle_{\mathcal{H}} \geq 0$$

Por ende, la forma bilineal es un producto interno.

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

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

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

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

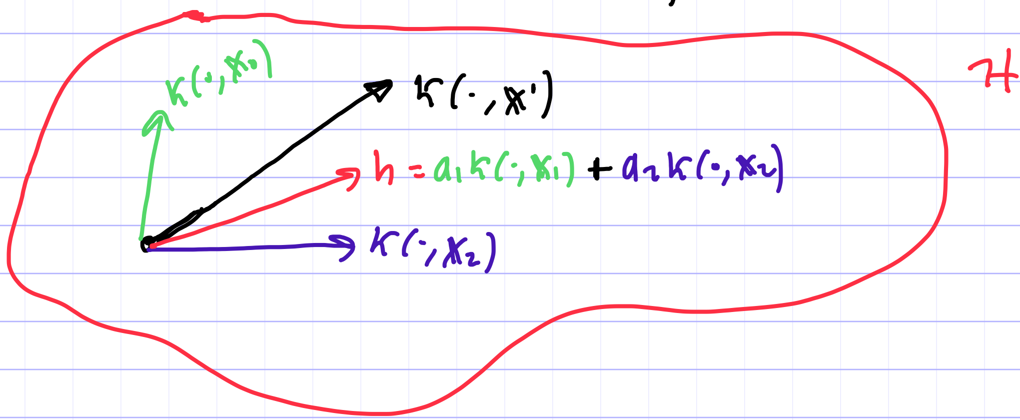
$$\langle h, g \rangle_{\mathcal{H}} = \sum_{i=1}^{\ell} a_i \langle k(x_i, \cdot), k(x, \cdot) \rangle_{\mathcal{H}} = \sum_{i=1}^{\ell} a_i k(x, x_i)$$

Reproducing property $\Rightarrow$. $= h(x)$

$K(x,x')$ → kernel reproductivo

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

→ $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 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:

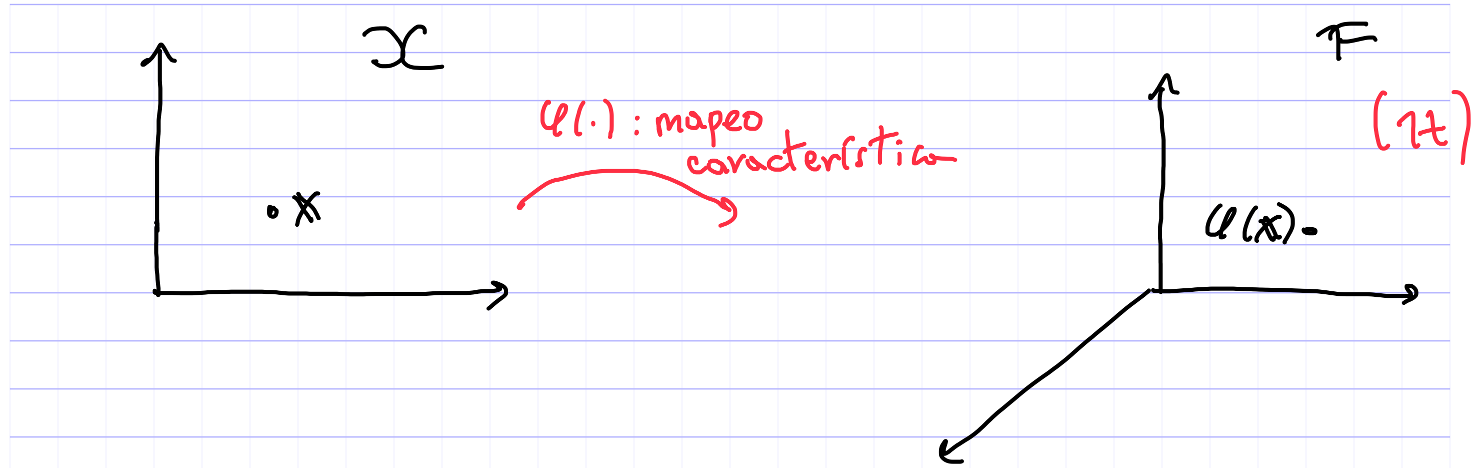
$$K(x,x') = \sum_{i=1}^{\infty} \lambda_i \, \phi_i(x) \, \phi_i(x')$$

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

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

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

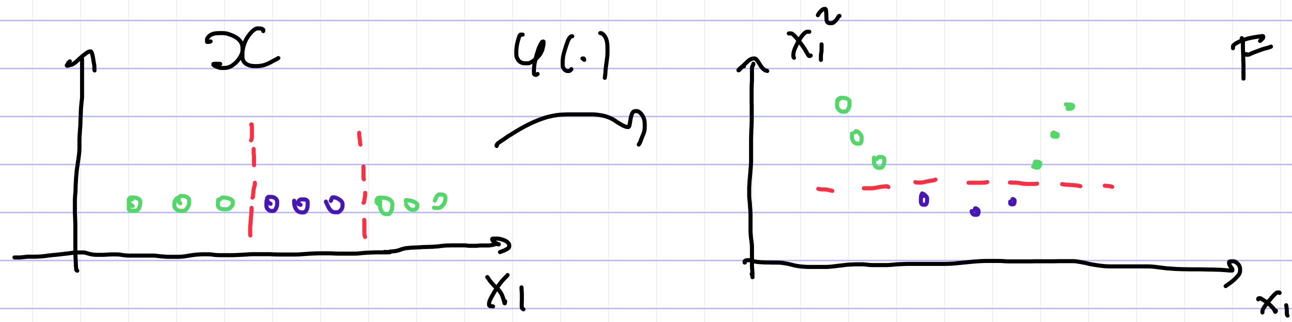
→ La dimensión de $\mathcal{F}$ se determina por el número de valores propios positivos → pueden ser ∞ para el kernel Gaussiano.



→ F es el RKHS inducido por k, con

$\varphi(\mathbf{x}) = k(\mathbf{x}, \cdot)$ → las bases de F y $\mathcal{H}$

Ejemplos:

① 

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

$$k(\mathbf{x},\mathbf{c}) = \left(1 + [x_1 \; x_2]\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}\right)^2 = \left(1 + c_1x_1 + c_2x_2\right)^2$$

$$= (1+c_1x_1)^2 + 2(1+c_1x_1)(c_2x_2) + (c_2x_2)^2 = 1 + 2c_1x_1 + c_1^2x_1^2 + c_2^2x_2^2 + 2(c_2x_2 + c_1c_2x_1x_2)$$

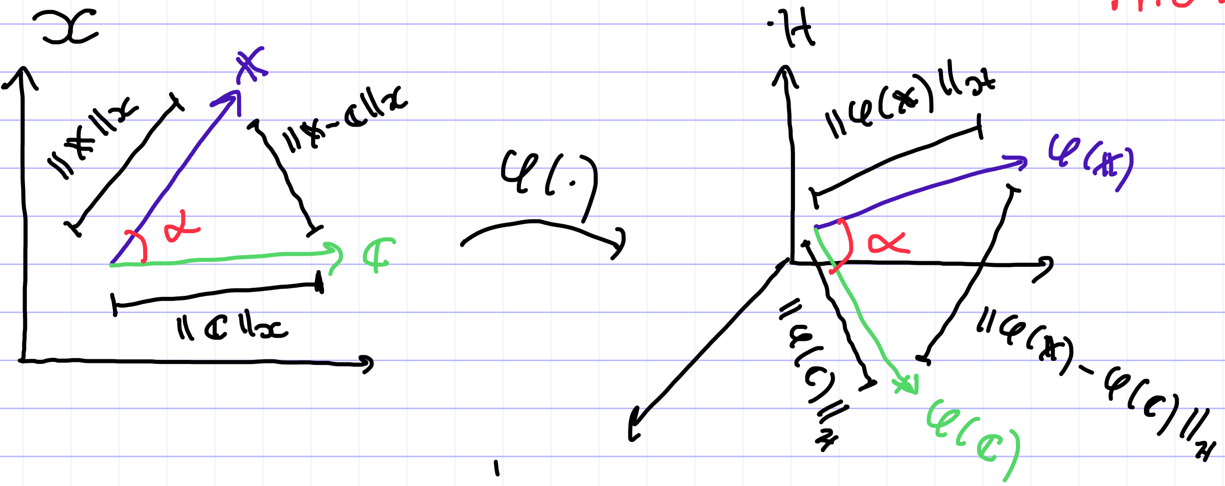
$$= c_1^2x_1^2 + c_2^2x_2^2 + 2c_1x_1 + 2c_2x_2 + 2c_1c_2x_1x_2 + 1$$

Así: $\varphi(\mathbf{x}) = [1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2]^T$

$\varphi(\mathbf{c}) = [1, c_1^2, \sqrt{2}c_1c_2, c_2^2, \sqrt{2}c_1, \sqrt{2}c_2]^T$

Por ende: $k(\mathbf{x},\mathbf{c}) = (1+\mathbf{x}^T\mathbf{c})^2 = \langle \varphi(\mathbf{x}), \varphi(\mathbf{c}) \rangle_{\mathcal{H}} = \varphi^T(\mathbf{x})\varphi(\mathbf{c})$.

Ilustración geométrica del RKHS. "Kernel Trick"



$\|\mathbf{x}\|^2_{\mathcal{X}} = \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{X}}$

$\|\mathbf{x}-\mathbf{c}\|_{\mathcal{X}} = \langle \mathbf{x}-\mathbf{c}, \mathbf{x}-\mathbf{c} \rangle_{\mathcal{X}}$

$= \langle \mathbf{x},\mathbf{x}\rangle - 2\langle \mathbf{x},\mathbf{c}\rangle + \langle \mathbf{c},\mathbf{c}\rangle$

$= \|\mathbf{x}\|^2_{\mathcal{X}} - 2\langle \mathbf{x},\mathbf{c}\rangle_{\mathcal{X}} + \|\mathbf{c}\|^2_{\mathcal{X}}$

$$\cos(\alpha) = \frac{\langle \mathbf{x},\mathbf{c}\rangle_{\mathcal{X}}}{\|\mathbf{x}\|_{\mathcal{X}}\|\mathbf{c}\|_{\mathcal{X}}}$$

$\|\varphi(\mathbf{x})\|^2_{\mathcal{H}} = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}) \rangle_{\mathcal{H}} = k(\mathbf{x},\mathbf{x})$

$\|\varphi(\mathbf{x}) - \varphi(\mathbf{c})\|^2_{\mathcal{H}} = \langle \varphi(\mathbf{x})-\varphi(\mathbf{c}), \varphi(\mathbf{x})-\varphi(\mathbf{c}) \rangle_{\mathcal{H}}$

$= \|\varphi(\mathbf{x})\|^2_{\mathcal{H}} - 2\langle \varphi(\mathbf{x}), \varphi(\mathbf{c}) \rangle_{\mathcal{H}} + \|\varphi(\mathbf{c})\|^2_{\mathcal{H}}$

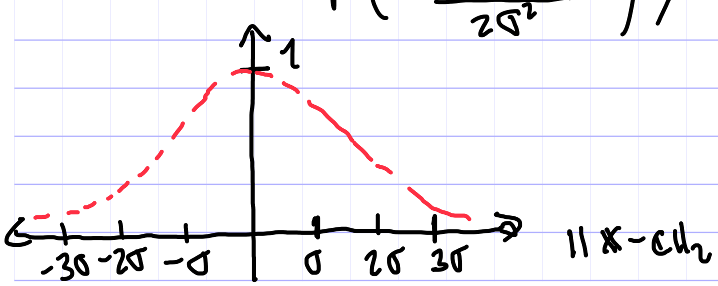
$= k(\mathbf{x},\mathbf{x}) - 2k(\mathbf{x},\mathbf{c}) + k(\mathbf{c},\mathbf{c})$

$$\cos(\alpha) = \frac{\langle \varphi(\mathbf{x}), \varphi(\mathbf{c}) \rangle_{\mathcal{H}}}{\|\varphi(\mathbf{x})\|_{\mathcal{H}}\|\varphi(\mathbf{c})\|_{\mathcal{H}}}$$

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

Ej: Kernel Gaussiano normalizado

$$k(\mathbf{x},\mathbf{c}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^2_2}{2\sigma^2}\right);$$



$\|\varphi(\mathbf{x}) - \varphi(\mathbf{c})\|^2_{\mathcal{H}} = 1 - 2k(\mathbf{x},\mathbf{c}) + 1$

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

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

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

Aplicaciones del RKHS en métodos clásicos basados en riesgo empírico.

1. Regresión rígida Kernel:

$$\mathbf{y}\in\mathbb{R}^{N};\quad \mathbf{X}\in\mathbb{R}^{N\times P};\qquad \varphi:\mathbb{R}^{P}\rightarrow\mathcal{H};\quad \mathcal{H}\subseteq\mathbb{R}^{Q};\quad Q\rightarrow\infty.$$

$$\hat{\mathbf{y}}=f(\varphi(\mathbf{X})\mid\mathbf{w})=\Phi(\mathbf{X})\mathbf{w};\quad f:\mathbb{R}^{Q}\rightarrow\mathbb{R};\quad \mathbf{w}\in\mathcal{H}\subseteq\mathbb{R}^{Q}.$$

$$\mathbf{w}^{*}=\underset{\mathbf{w}}{\arg\min}\ \mathcal{L}\big(\mathbf{y},f(\varphi(\mathbf{X})\mid\mathbf{w})\big)+\mathcal{R}(f\mid\lambda)$$

$$=\underset{\mathbf{w}}{\arg\min}\ \frac{1}{N}\|\mathbf{y}-\Phi(\mathbf{X})\mathbf{w}\|_2^2+\lambda\|\mathbf{w}\|_{\mathcal{H}}^2$$

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

$$\|\mathbf{y}-\Phi(\mathbf{X})\mathbf{w}\|_2^2=\langle\mathbf{y}-\Phi(\mathbf{X})\mathbf{w},\mathbf{y}-\Phi(\mathbf{X})\mathbf{w}\rangle$$

$$=\mathbf{y}^{T}\mathbf{y}-2\mathbf{y}^{T}\Phi(\mathbf{X})\mathbf{w}+(\Phi(\mathbf{X})\mathbf{w})^{T}(\Phi(\mathbf{X})\mathbf{w})$$

$$=\mathbf{y}^{T}\mathbf{y}-2\mathbf{y}^{T}\Phi(\mathbf{X})\mathbf{w}+\mathbf{w}^{T}\Phi^{T}(\mathbf{X})\Phi(\mathbf{X})\mathbf{w}$$

NOTA:
$$\Phi(\mathbf{X})=\begin{bmatrix}\varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \cdots & \varphi_Q(\mathbf{x}_1)\\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \cdots & \varphi_Q(\mathbf{x}_2)\\ \vdots & \vdots & \ddots & \vdots\\ \varphi_1(\mathbf{x}_N) & \varphi_2(\mathbf{x}_N) & & \varphi_Q(\mathbf{x}_N)\end{bmatrix}_{N\times Q}$$

$$\frac{1}{N}\Phi^{T}(\mathbf{X})\Phi(\mathbf{X})\in\mathbb{R}^{Q\times Q}\simeq\mathbb{E}\{\Phi(\mathbf{X})\Phi(\mathbf{X})\}=R_{\varphi(x)\varphi(x)}$$

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

NO SIEMPRE SE PUEDE CALCULAR.

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

$$\nabla_{\mathbb{W}} = \frac{2}{N}\left[\left(\phi^T(\mathbb{X})\phi(\mathbb{X})+N\lambda\right)\mathbb{W} - \phi^T(\mathbb{X})y\right]$$

Solución analítica:

$$\frac{2}{N}\left[\left(\phi^T(\mathbb{X})\phi(\mathbb{X})+N\lambda I\right)\mathbb{W}\right] = \frac{2}{N}\phi^T(\mathbb{X})Y.$$

$$\mathbb{W}^* = \left[\phi^T(\mathbb{X})\phi(\mathbb{X})+N\lambda I\right]^{-1}\phi^T(\mathbb{X})Y$$

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

Propiedades matriz inversa:

$$(I+AB)^{-1}A = A(I+BA)^{-1}$$

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

$$= \frac{1}{N\lambda}\Big(I+\underbrace{\phi^T}_{A}\underbrace{\frac{\phi}{N\lambda}}_{B}\Big)^{-1}\underbrace{\phi^T}_{A} = \phi^T\frac{1}{N\lambda}\left(I+\frac{\phi}{N\lambda}\phi^T\right)^{-1}$$

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

$$\boxed{W^* = \phi^T(\mathbb{X})\left[N\lambda I+\phi(\mathbb{X})\phi^T(\mathbb{X})\right]^{-1}Y}$$

$$\hat{y} = f(\varphi(\mathbf{x}_{new}) \mid \mathbf{w}) = \varphi^T(\mathbf{x}_{new})\mathbf{w}$$

$$= \varphi^T(\mathbf{x}_{new})\,\Phi^T(\mathbf{X})\left[N\lambda I + \Phi(\mathbf{X})\Phi^T(\mathbf{X})\right]^{-1}\mathbf{Y}$$

$\varphi(\mathbf{x}_{new}) \in \mathbb{R}^Q$.

$$\Phi(\mathbf{X}) = \left[\varphi(\mathbf{x}_1)\ \ \varphi(\mathbf{x}_2)\ \cdots\ \varphi(\mathbf{x}_N)\right]^T_{N\times Q}$$

$$\Phi(\mathbf{X})\Phi^T(\mathbf{X}) = \begin{bmatrix} \varphi^T(\mathbf{x}_1) \\ \varphi^T(\mathbf{x}_2) \\ \vdots \\ \varphi^T(\mathbf{x}_N) \end{bmatrix}_{N\times Q} \left[\varphi(\mathbf{x}_1)\ \varphi(\mathbf{x}_2)\ \cdots\ \varphi(\mathbf{x}_N)\right]_{Q\times N}$$

$$\mathbb{K} \in \mathbb{R}^{N\times N} = \begin{bmatrix} \varphi^T(\mathbf{x}_1)\varphi(\mathbf{x}_1) & \varphi^T(\mathbf{x}_1)\varphi(\mathbf{x}_2) & \cdots & \varphi^T(\mathbf{x}_1)\varphi(\mathbf{x}_N) \\ \varphi^T(\mathbf{x}_2)\varphi(\mathbf{x}_1) & \varphi^T(\mathbf{x}_2)\varphi(\mathbf{x}_2) & \cdots & \varphi^T(\mathbf{x}_2)\varphi(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi^T(\mathbf{x}_N)\varphi(\mathbf{x}_1) & \varphi^T(\mathbf{x}_N)\varphi(\mathbf{x}_2) & & \varphi^T(\mathbf{x}_N)\varphi(\mathbf{x}_N) \end{bmatrix}$$

$$\mathbb{K} = \left[\kappa(\mathbf{x}_n, \mathbf{x}_m)\right]_{n,m=1}^{N} \Rightarrow$$ POR KERNE TRICK.

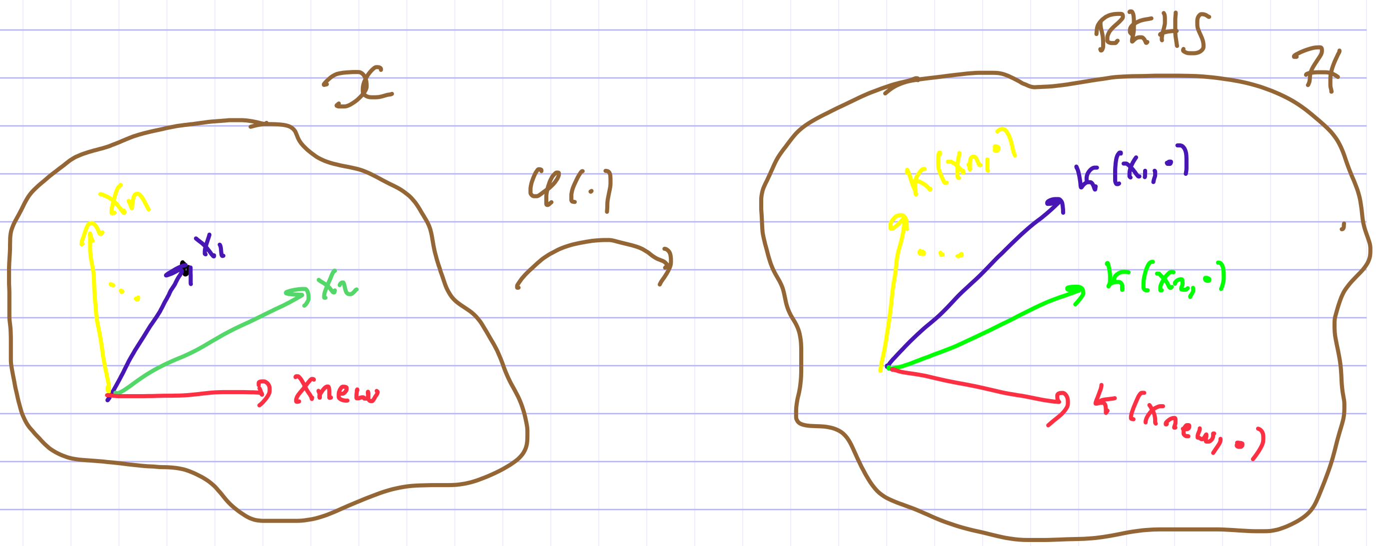
$$\varphi^T(\mathbf{x}_{new})\Phi^T(\mathbf{X}) = \left[\varphi^T(\mathbf{x}_{new})\right]_{1\times Q}\left[\varphi(\mathbf{x}_1)\ \varphi(\mathbf{x}_2)\ \cdots\ \varphi(\mathbf{x}_N)\right]_{Q\times N}$$

$$\mathbb{K}^T_{new} \in \mathbb{R}^{1\times N} = \left[\kappa(\mathbf{x}_{new}, \mathbf{x}_n)\right]_{n=1}^{N}$$

$$\hat{y} = f(x_{new}) = \mathbb{K}^T_{new} \left(N\lambda I + \mathbb{K} \right)^{-1} \mathbb{y}$$
$$\quad 1\times N \qquad\qquad N\times N \times N\times 1$$

$$\hat{y} = \mathbb{K}^T_{new}\, \alpha \;; \qquad \alpha \in \mathbb{R}^{N\times 1}$$

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



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

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

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

NOTA: $\alpha_n \in \mathbb{R}$ se encuentra mediante mín. cuadrador regularizador en RKHS.

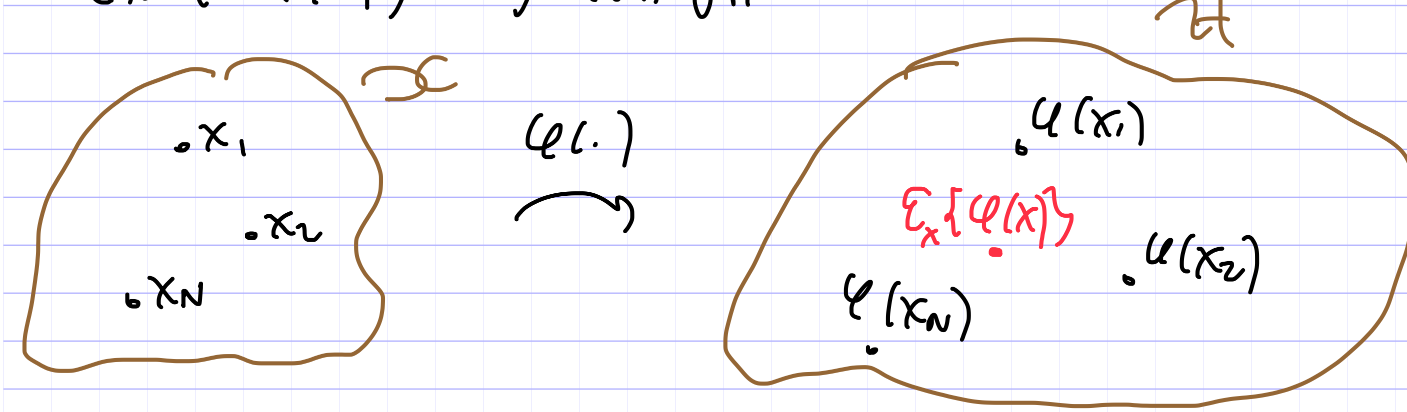
2. Centered Kernel Alignment - CKA

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

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



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

$$k_c(x,x') = \langle \varphi(x) - \mathbb{E}_x\{\varphi(x)\}, \varphi(x') - \mathbb{E}_{x'}\{\varphi(x')\} \rangle_{\mathcal{H}}$$

$$= \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}} - \langle \varphi(x), \mathbb{E}_{x'}\{\varphi(x')\} \rangle_{\mathcal{H}} - \langle \mathbb{E}_x\{\varphi(x)\}, \varphi(x') \rangle_{\mathcal{H}} + \langle \mathbb{E}_x\{\varphi(x)\}, \mathbb{E}_{x'}\{\varphi(x')\} \rangle_{\mathcal{H}}$$

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

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

NOTA: $\mathcal{E}_{x,x'}\{k(x,x')\} = \int\limits_{x\in X}\int\limits_{x'\in X} k(x,x')\, dP_x\, dP_{x'}$

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

$$\mathcal{E}_{x,x'}\{k(x,x')\} = \int\limits_{x\in X}\int\limits_{x'\in X} k(x,x')\,p(x)\,p(x')\,dx\,dx'$$

Estimación del CKA:

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

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

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

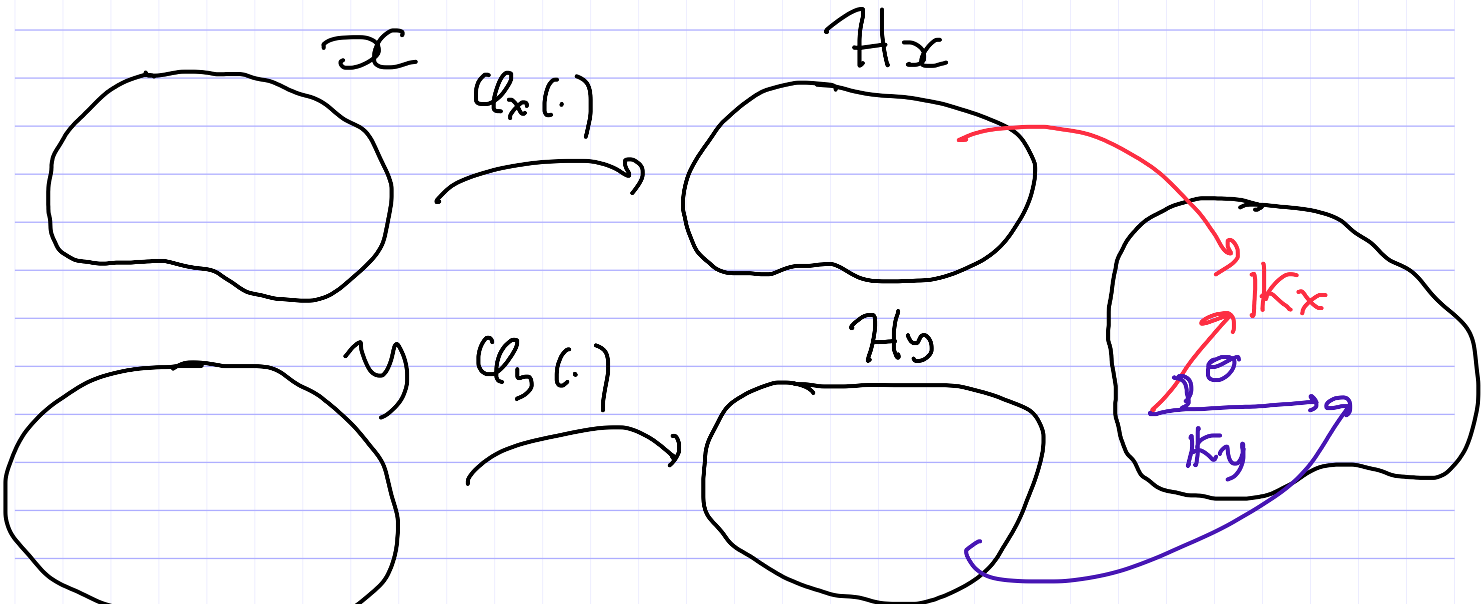
$$k_c(x_n, x_m) = k(x_n, x_m) - \frac{1}{N}\sum_{n=1}^{N} k(x_n, x_m) - \frac{1}{N}\sum_{m=1}^{N} k(x_n, x_m)$$
$$+ \frac{1}{N^2}\sum_{n,m=1}^{N} k(x_n, x_m).$$

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

Así: $\mathbb{K}_c = \mathbb{H}\,\mathbb{K}\,\mathbb{H}$; $\mathbb{H} = \mathbb{I} - \frac{\mathbf{1}\mathbf{1}^T}{N}$

TAREA: DEMOSTRAR

Función de alineamiento



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

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

$$\rho(K_x,K_y) = \frac{\mathbb{E}\{\bar{k}_x \bar{k}_y\}}{\sqrt{\mathbb{E}\{\bar{k}_x^2\}\,\mathbb{E}\{\bar{k}_y^2\}}}$$

Demostrar: $\mathbb{K}_c = \mathbb{H} \mathbb{K} \mathbb{H}$; $\mathbb{H} = \mathbb{I} - \frac{\mathbb{1}\mathbb{1}^T}{N}$

$$\left(I - \frac{11^T}{N}\right) K \left(I - \frac{11^T}{N}\right) = \left(K - \frac{11^T}{N}K\right)\left(I - \frac{11^T}{N}\right)$$

$$= K - K\frac{11^T}{N} - \frac{11^T}{N}K + \frac{11^T}{N}K\frac{11^T}{N}.$$

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

$$\begin{bmatrix} k_c(x_1,x_1) & \cdots & k_c(x_1,x_N) \\ \vdots & \ddots & \vdots \\ k_c(x_N,x_1) & \cdots & K_c(x_N,x_N) \end{bmatrix} = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_N) \\ \vdots & \ddots & \vdots \\ k(x_N,x_1) & \cdots & k(x_N,x_N) \end{bmatrix} - \frac{1}{N}\begin{bmatrix} \sum_{n=1}^{N} k(x_1,x_n) & \cdots & \sum_{n=1}^{N} k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ \sum_{n=1}^{N} k(x_N,x_n) & \cdots & \sum_{n=1}^{N} k(x_N,x_n) \end{bmatrix}$$

$$- \frac{1}{N}\begin{bmatrix} \sum_{m=1}^{N} k(x_m,x_1) & \cdots & \sum_{m=1}^{N} k(x_m,x_N) \\ \vdots & \ddots & \vdots \\ \sum_{m=1}^{N} k(x_m,x_1) & \cdots & \sum_{m=1}^{N} k(x_m,x_N) \end{bmatrix} + \frac{1}{N^2}\begin{bmatrix} \sum_{m=1}^{N}\sum_{n=1}^{N} k(x_m,x_n) & \cdots & \sum_{m=1}^{N}\sum_{n=1}^{N} k(x_m,x_n) \\ \vdots & \ddots & \vdots \\ \sum_{m=1}^{N}\sum_{n=1}^{N} k(x_m,x_n) & \cdots & \sum_{m=1}^{N}\sum_{n=1}^{N} k(x_m,x_n) \end{bmatrix}$$

Por desigualdad de Cauchy-Schwarz:

$$\left|\mathbb{E}\{\bar{k}_x \bar{k}_y\}\right| \le \sqrt{\mathbb{E}\{\bar{k}_x^2\}\,\mathbb{E}\{\bar{k}_y^2\}}.$$

$$\hat{\rho}(\mathbb{K}_x,\mathbb{K}_y)\in[0,1]; \qquad \|\mathbb{K}\|_F^2 = tr(\mathbb{K}\mathbb{K}^*) = tr(\mathbb{K}^*\mathbb{K})$$

Algoritmos desde CKA:

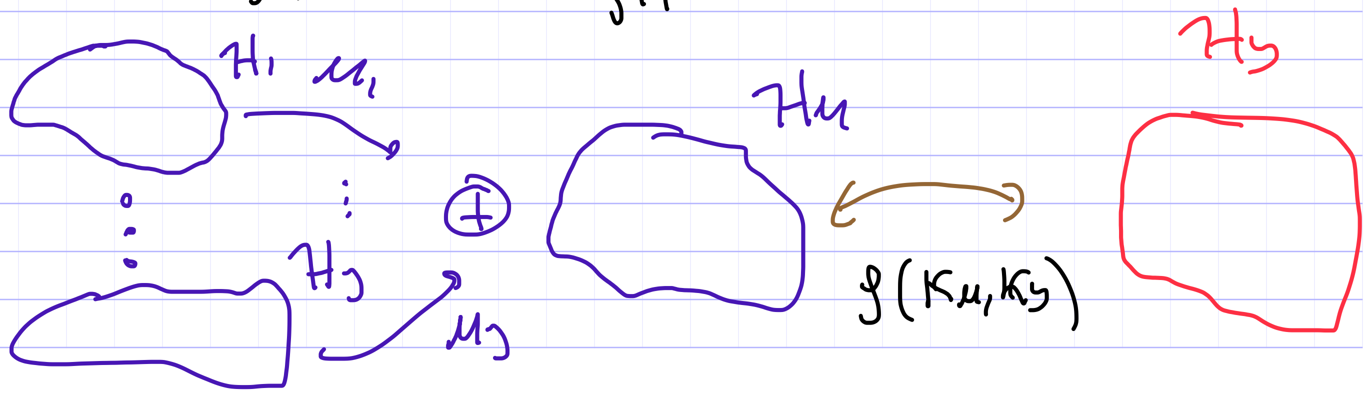
a. Alineamiento independiente:

$$\{\mathbb{K}_j\in\mathbb{R}^{N\times N}\}_{j=1}^{J} \qquad \mathbb{K}_Y\in\mathbb{R}^{N\times N} \rightarrow \text{kernel objetivo.}$$

$$\mathbb{K}_\mu \propto \sum_{j=1}^{J}\mu_j\mathbb{K}_j; \qquad \mu_j = \hat{\rho}(\mathbb{K}_j,\mathbb{K}_y)$$

$$\mathbb{K}_\mu = \sum_{j=1}^{J}\hat{\rho}(\mathbb{K}_j,\mathbb{K}_y)\,\mathbb{K}_j = \sum_{j=1}^{J}\frac{\langle\bar{\mathbb{K}}_j,\bar{\mathbb{K}}_y\rangle_F}{\|\bar{\mathbb{K}}_j\|_F\,\|\bar{\mathbb{K}}_y\|_F}\,\mathbb{K}_j$$

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



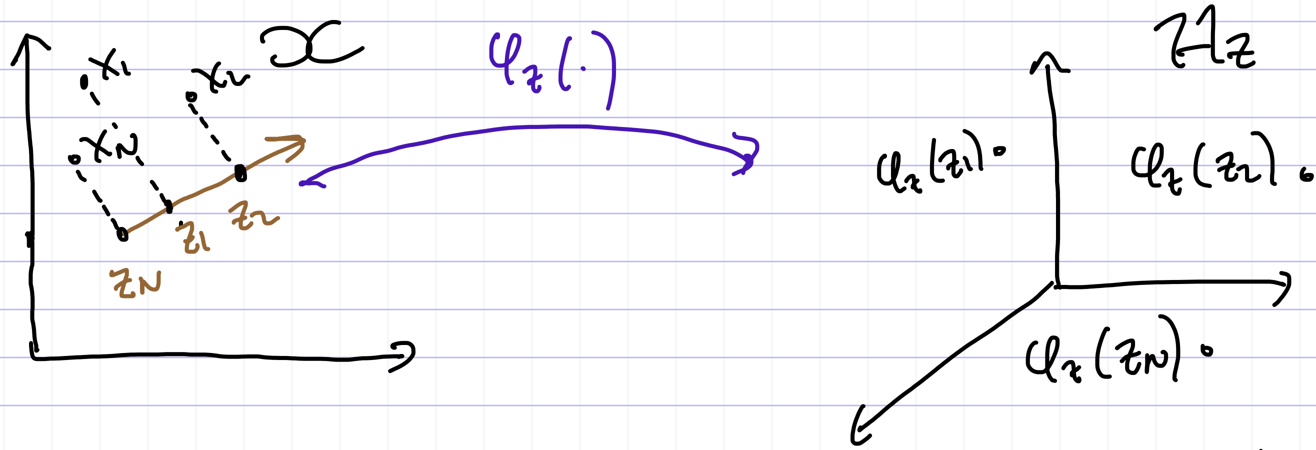
b. Alineamiento restringido

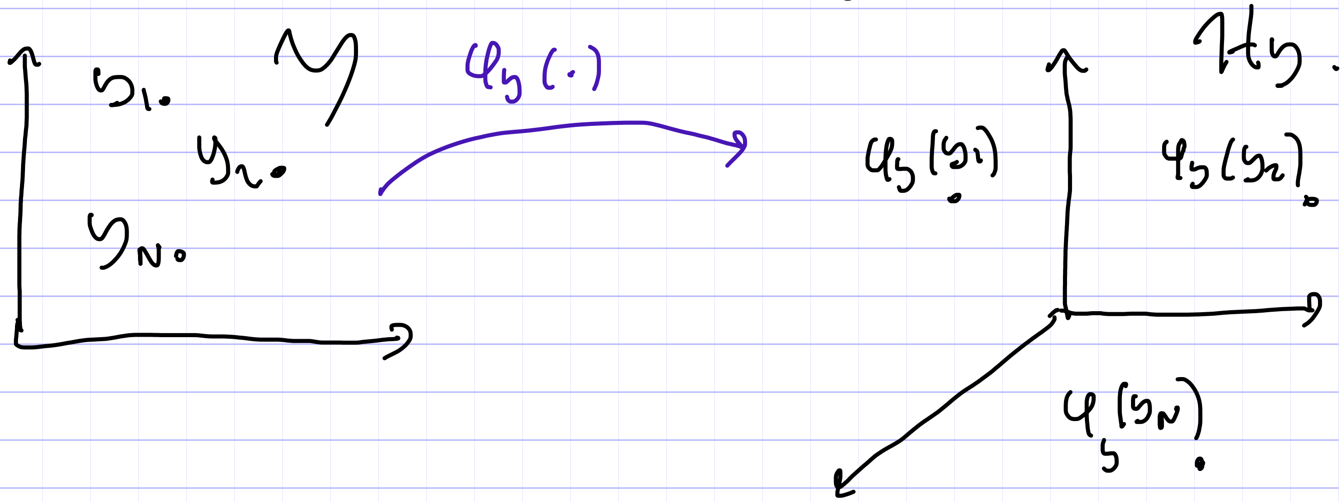
$$\max_{\mu} \; \hat{\rho}\left(\sum_{j=1}^{J} \mu_j \mathbb{K}_j , \mathbb{K}_y \right)$$

$$\text{s.t.} \quad \|\mu\|_q^q \leq \xi$$

TAREA: Resolver por Lagrange para $q=2$ y $\|\mu\|_2 = 1$, $\mu \geq 0$. (combinación convexa).

c. Aprendizaje de métrica. (Metric Learning)





$$\rho(\mathbb{K}_z, \mathbb{K}_y) = \frac{\langle \mathbb{K}_z, \mathbb{K}_y \rangle_F}{\|\mathbb{K}_z\|_F \, \|\mathbb{K}_y\|_F}$$

Si $z = XA$; $x \in \mathbb{R}^P$; $A \in \mathbb{R}^{P \times Q}$; $Q \leq P$.

$z \in \mathbb{R}^Q$; $y \in \mathcal{Y}$.

Si $\kappa_z(z_n, z_m) = \exp\left(-\gamma \|z_n - z_m\|_2^2\right)$

$$\kappa_z(x_n A, x_m A) = \exp\left(-\gamma \|x_n A - x_m A\|_2^2\right)$$

$$= \exp\left(-\gamma \langle x_n A - x_m A, x_n A - x_m A \rangle\right)$$

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

$$= \underset{1 \times Q}{(x_n A - x_m A)} \underset{Q \times 1}{(x_n A - x_m A)^T}$$

$$= \underset{1 \times P}{(x_n - x_m)} \underset{P \times Q}{A} \underset{Q \times 1}{\left((x_n A)^T - (x_m A)^T\right)}$$

$$= (x_n - x_m) A \left(A^T x_n^T - A^T x_m^T\right)$$

$$d_z^2(z_n, z_m) = \underset{1 \times P}{(x_n - x_m)} \underset{P \times P}{A A^T} \underset{P \times 1}{(x_n - x_m)^T} = d_x^2\left(x_n, x_m \mid \Sigma^{-1} = A A^T\right)$$

Euclídea en z

Mahalanobis en X.

$$A^* = \underset{A}{\arg\max} \; \rho\left(\mathbb{K}_z(A), \mathbb{K}_y\right)$$

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