

Kernels

Fernando Lozano

Universidad de los Andes

7 de octubre de 2014



- Suponga que existe un **kernel**:

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (\mathbf{x}_1, \mathbf{x}_2) &\mapsto k(\mathbf{x}_1, \mathbf{x}_2) = \langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle_{\mathcal{H}} \end{aligned}$$

- Suponga que existe un **kernel**:

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (\mathbf{x}_1, \mathbf{x}_2) &\mapsto k(\mathbf{x}_1, \mathbf{x}_2) = \langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle_{\mathcal{H}} \end{aligned}$$

- Podemos calcular el producto punto $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$ operando en el **conjunto original** \mathcal{X} .

- Suponga que existe un **kernel**:

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (\mathbf{x}_1, \mathbf{x}_2) &\mapsto k(\mathbf{x}_1, \mathbf{x}_2) = \langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle_{\mathcal{H}} \end{aligned}$$

- Podemos calcular el producto punto $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$ operando en el **conjunto original** \mathcal{X} .
- Más aún, no requerimos conocer el mapeo ϕ , o el espacio \mathcal{H} .

- Suponga que existe un **kernel**:

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (\mathbf{x}_1, \mathbf{x}_2) &\mapsto k(\mathbf{x}_1, \mathbf{x}_2) = \langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle_{\mathcal{H}} \end{aligned}$$

- Podemos calcular el producto punto $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$ operando en el **conjunto original** \mathcal{X} .
- Más aún, no requerimos conocer el mapeo ϕ , o el espacio \mathcal{H} .
- De hecho \mathcal{H} puede tener dimensión infinita.

- Suponga que existe un **kernel**:

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (\mathbf{x}_1, \mathbf{x}_2) &\mapsto k(\mathbf{x}_1, \mathbf{x}_2) = \langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle_{\mathcal{H}} \end{aligned}$$

- Podemos calcular el producto punto $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$ operando en el **conjunto original** \mathcal{X} .
- Más aún, no requerimos conocer el mapeo ϕ , o el espacio \mathcal{H} .
- De hecho \mathcal{H} puede tener dimensión infinita.
- Cómo sabemos si una función $k(.,.)$ es un **kernel**?

El truco del Kernel

$$\begin{aligned} \text{máx} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \text{sujeto a} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

El truco del Kernel

$$\begin{aligned} \text{máx} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \\ \text{sujeto a} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

El truco del Kernel

$$\begin{aligned} \text{máx} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{sujeto a} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

El truco del Kernel

$$\text{máx} \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{sujeto a} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

- Solución: $\mathbf{w} = \sum_{i=1}^n (\alpha_i y_i) \phi(\mathbf{x}_i)$

El truco del Kernel

$$\text{máx} \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{sujeto a} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

- Solución: $\mathbf{w} = \sum_{i=1}^n (\alpha_i y_i) \phi(\mathbf{x}_i)$
- Evaluación:

El truco del Kernel

$$\text{máx} \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{sujeto a} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

- Solución: $\mathbf{w} = \sum_{i=1}^n (\alpha_i y_i) \phi(\mathbf{x}_i)$
- Evaluación:

$$\text{sign}(\langle w, \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b)$$

El truco del Kernel

$$\text{máx} \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{sujeto a} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

- Solución: $\mathbf{w} = \sum_{i=1}^n (\alpha_i y_i) \phi(\mathbf{x}_i)$
- Evaluación:

$$\text{sign}(\langle w, \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b) = \text{sign} \left(\sum_{i=1}^n (\alpha_i y_i) \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b \right)$$

El truco del Kernel

$$\begin{aligned} \text{máx} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{sujeto a} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

- Solución: $\mathbf{w} = \sum_{i=1}^n (\alpha_i y_i) \phi(\mathbf{x}_i)$
- Evaluación:

$$\begin{aligned} \text{sign}(\langle w, \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b) &= \text{sign} \left(\sum_{i=1}^n (\alpha_i y_i) \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b \right) \\ &= \text{sign} \left(\sum_{i=1}^n (\alpha_i y_i) k(\mathbf{x}_i, \mathbf{x}) + b \right) \end{aligned}$$

Ejemplo

- Si $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \mathbb{R}^3$, $\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, tenemos el kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Ejemplo

- Si $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \mathbb{R}^3$, $\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, tenemos el kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Comprobando:

Ejemplo

- Si $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \mathbb{R}^3$, $\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, tenemos el kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Comprobando:

$$\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} = \left\langle (x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2), (x_{j1}^2, \sqrt{2}x_{j1}x_{j2}, x_{j2}^2) \right\rangle$$

Ejemplo

- Si $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \mathbb{R}^3$, $\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, tenemos el kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Comprobando:

$$\begin{aligned}\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} &= \left\langle (x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2), (x_{j1}^2, \sqrt{2}x_{j1}x_{j2}, x_{j2}^2) \right\rangle \\ &= x_{i1}^2x_{j1}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + x_{i2}^2x_{j2}^2\end{aligned}$$

Ejemplo

- Si $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \mathbb{R}^3$, $\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, tenemos el kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Comprobando:

$$\begin{aligned}\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} &= \left\langle (x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2), (x_{j1}^2, \sqrt{2}x_{j1}x_{j2}, x_{j2}^2) \right\rangle \\ &= x_{i1}^2x_{j1}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + x_{i2}^2x_{j2}^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2\end{aligned}$$

Ejemplo

- Si $\mathcal{X} = \mathbb{R}^2$, $\mathcal{H} = \mathbb{R}^3$, $\phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$, tenemos el kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Comprobando:

$$\begin{aligned}\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} &= \left\langle (x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2), (x_{j1}^2, \sqrt{2}x_{j1}x_{j2}, x_{j2}^2) \right\rangle \\ &= x_{i1}^2x_{j1}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + x_{i2}^2x_{j2}^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2\end{aligned}$$

- El mapeo ϕ y el espacio \mathcal{H} correspondientes a un kernel **no son únicos**.

- El mapeo ϕ y el espacio \mathcal{H} correspondientes a un kernel **no son únicos**.
- Por ejemplo, para $\mathcal{X} = \mathbb{R}^2$ y el kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$, podemos tener:

- El mapeo ϕ y el espacio \mathcal{H} correspondientes a un kernel **no son únicos**.
- Por ejemplo, para $\mathcal{X} = \mathbb{R}^2$ y el kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$, podemos tener:
 - ▶ $\mathcal{H} = \mathbb{R}^4$ y

$$\phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}$$

- El mapeo ϕ y el espacio \mathcal{H} correspondientes a un kernel **no son únicos**.
- Por ejemplo, para $\mathcal{X} = \mathbb{R}^2$ y el kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$, podemos tener:
 - ▶ $\mathcal{H} = \mathbb{R}^4$ y

$$\phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_1x_2 \\ x_2^2 \end{pmatrix}$$



- El mapeo ϕ y el espacio \mathcal{H} correspondientes a un kernel **no son únicos**.
- Por ejemplo, para $\mathcal{X} = \mathbb{R}^2$ y el kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$, podemos tener:
 - ▶ $\mathcal{H} = \mathbb{R}^4$ y

$$\phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}$$



- ▶ $\mathcal{H} = \mathbb{R}^2$ y

$$\phi(\mathbf{x}) = \begin{pmatrix} \frac{x_1^2 - x_2^2}{2x_1 x_2} \\ x_1^2 + x_2^2 \end{pmatrix}$$

- El mapeo ϕ y el espacio \mathcal{H} correspondientes a un kernel **no son únicos**.
- Por ejemplo, para $\mathcal{X} = \mathbb{R}^2$ y el kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$, podemos tener:
 - ▶ $\mathcal{H} = \mathbb{R}^4$ y

$$\phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}$$



- ▶ $\mathcal{H} = \mathbb{R}^2$ y

$$\phi(\mathbf{x}) = \begin{pmatrix} \frac{x_1^2 - x_2^2}{2x_1 x_2} \\ x_1^2 + x_2^2 \end{pmatrix}$$



Ejemplos de kernels

Ejemplos de kernels

- Polinomial:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d$$

Ejemplos de kernels

- Polinomial:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d$$

- Polinomial no homogéneo:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + c)^d$$

Ejemplos de kernels

- Polinomial:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d$$

- Polinomial no homogéneo:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + c)^d$$

- Gaussiano (RBF):

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

Ejemplos de kernels

- Polinomial:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^d$$

- Polinomial no homogéneo:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + c)^d$$

- Gaussiano (RBF):

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

- Sigmoidal:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a \langle \mathbf{x}_i, \mathbf{x}_j \rangle + b)$$

Cómo sabemos si una función $k(.,.)$ es un **kernel**?

Cómo sabemos si una función $k(.,.)$ es un **kernel**?

Dos aproximaciones:

Cómo sabemos si una función $k(.,.)$ es un **kernel**?

Dos aproximaciones:

- Reproducing Kernel Hilbert Spaces (RKHS)

Cómo sabemos si una función $k(.,.)$ es un **kernel**?

Dos aproximaciones:

- Reproducing Kernel Hilbert Spaces (RKHS) (Aronszajn, 1950)

Cómo sabemos si una función $k(.,.)$ es un **kernel**?

Dos aproximaciones:

- Reproducing Kernel Hilbert Spaces (RKHS) (Aronszajn, 1950)
- Teorema de Mercer

Cómo sabemos si una función $k(.,.)$ es un **kernel**?

Dos aproximaciones:

- Reproducing Kernel Hilbert Spaces (RKHS) (Aronszajn, 1950)
- Teorema de Mercer (Mercer, 1911).

Definición

Dada una función simétrica $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ ($k(x_1, x_2) = k(x_2, x_1)$) y $x_1, x_2, \dots, x_n \in \mathcal{X}$, la matriz $n \times n$ con entradas

$$K_{ij} = k(x_i, x_j)$$

se llama la *matriz de Gram* de k con respecto a x_1, x_2, \dots, x_n .

Definición

Dada una función simétrica $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ ($k(x_1, x_2) = k(x_2, x_1)$) y $x_1, x_2, \dots, x_n \in \mathcal{X}$, la matriz $n \times n$ con entradas

$$K_{ij} = k(x_i, x_j)$$

se llama la *matriz de Gram* de k con respecto a x_1, x_2, \dots, x_n .

Definición

Una función $k : \mathcal{X}^2 \rightarrow \mathbb{R}$ para la cual para todo $n \in \mathbb{N}$, y todo $x_1, x_2, \dots, x_n \in \mathcal{X}$ resulta en una matriz de Gram *positiva semidefinida* es un *kernel positivo definido* (o simplemente un kernel).

El mapa del kernel reproductor

- Kernel positivo definido k .

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\phi : \mathcal{X} \rightarrow \mathbb{R}^{\mathcal{X}}$$

.

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(., x).\end{aligned}$$

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(., x).\end{aligned}$$

- Receta:

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(., x).\end{aligned}$$

- Receta:
 - 1 Imagen de $\phi \longrightarrow$ espacio vectorial.

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(., x).\end{aligned}$$

- Receta:
 - 1 Imagen de $\phi \longrightarrow$ espacio vectorial.
 - 2 Producto punto.

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(., x).\end{aligned}$$

- Receta:
 - 1 Imagen de $\phi \longrightarrow$ espacio vectorial.
 - 2 Producto punto.
 - 3 $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$

El mapa del kernel reproductor

- Kernel positivo definido k .
- Mapeo:

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(., x).\end{aligned}$$

- Receta:
 - 1 Imagen de $\phi \longrightarrow$ espacio vectorial.
 - 2 Producto punto.
 - 3 $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$
 - 4 Completar espacio.

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto:

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto: Si $g(.) = \sum_{j=1}^m \beta_j k(., x'_j)$ definimos:

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto: Si $g(.) = \sum_{j=1}^m \beta_j k(., x'_j)$ definimos:

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

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto: Si $g(.) = \sum_{j=1}^m \beta_j k(., x'_j)$ definimos:

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

- Note que:

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

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto: Si $g(.) = \sum_{j=1}^m \beta_j k(., x'_j)$ definimos:

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

- Note que:

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

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto: Si $g(.) = \sum_{j=1}^m \beta_j k(., x'_j)$ definimos:

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

- Note que:

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

- Vectores:

$$f(.) = \sum_{i=1}^n \alpha_i k(., x_i) \quad n \in \mathbb{N}, \alpha_i \in \mathbb{R}, x_1, \dots \in \mathcal{X}$$

- Producto punto: Si $g(.) = \sum_{j=1}^m \beta_j k(., x'_j)$ definimos:

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

- Note que:

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

- En particular:

$$\langle k(., x), f \rangle =$$

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle =$$

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

Esta es la **propiedad del Kernel Reprodutor**.

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

Esta es la **propiedad del Kernel Reprodutor**.

- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una fila de esta matriz.

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

Esta es la **propiedad del Kernel Reprodutor**.

- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una fila de esta matriz.
- Sea $Q \in \mathbb{R}^{n \times n}$ una matriz positiva definida. Podemos definir un producto punto en \mathbb{R}^n :

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_{Q^{-1}} = \mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_j$$

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

Esta es la **propiedad del Kernel Reprodutor**.

- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una fila de esta matriz.
- Sea $Q \in \mathbb{R}^{n \times n}$ una matriz positiva definida. Podemos definir un producto punto en \mathbb{R}^n :

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_{Q^{-1}} = \mathbf{x}_i^T \mathbf{Q}^{-1} \mathbf{x}_j$$

Si $\mathbf{u}_i, \mathbf{u}_j$ son columnas de \mathbf{Q} :

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

Esta es la **propiedad del Kernel Reprodutor**.

- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una fila de esta matriz.
- Sea $Q \in \mathbb{R}^{n \times n}$ una matriz positiva definida. Podemos definir un producto punto en \mathbb{R}^n :

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_{Q^{-1}} = \mathbf{x}_i^T Q^{-1} \mathbf{x}_j$$

Si $\mathbf{u}_i, \mathbf{u}_j$ son columnas de Q :

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle_{Q^{-1}} =$$

- En particular:

$$\langle k(., x), f \rangle = f(x)$$

y

$$\langle k(., x), k(., x') \rangle = k(x, x')$$

Esta es la **propiedad del Kernel Reprodutor**.

- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una fila de esta matriz.
- Sea $Q \in \mathbb{R}^{n \times n}$ una matriz positiva definida. Podemos definir un producto punto en \mathbb{R}^n :

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle_{Q^{-1}} = \mathbf{x}_i^T Q^{-1} \mathbf{x}_j$$

Si $\mathbf{u}_i, \mathbf{u}_j$ son columnas de Q :

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle_{Q^{-1}} = q_{ij}$$

Propiedades

Propiedades

1 Simetría:

Propiedades

1 Simetría:

$$\langle f, g \rangle = \langle g, f \rangle$$

Propiedades

1 Simetría:

$$\langle f, g \rangle = \langle g, f \rangle$$

(porque $k(x_i, x_j) = k(x_j, x_i)$)

Propiedades

1 Simetría:

$$\langle f, g \rangle = \langle g, f \rangle$$

(porque $k(x_i, x_j) = k(x_j, x_i)$)

2 Linealidad-

Propiedades

1 Simetría:

$$\langle f, g \rangle = \langle g, f \rangle$$

(porque $k(x_i, x_j) = k(x_j, x_i)$)

2 Linealidad-

3 Positividad:

$$\langle f, f \rangle = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j)$$

Propiedades

1 Simetría:

$$\langle f, g \rangle = \langle g, f \rangle$$

(porque $k(x_i, x_j) = k(x_j, x_i)$)

2 Linealidad-

3 Positividad:

$$\langle f, f \rangle = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j) \geq 0$$

y $\langle f, f \rangle = 0$ sólo si $f = 0$

Propiedades

1 Simetría:

$$\langle f, g \rangle = \langle g, f \rangle$$

(porque $k(x_i, x_j) = k(x_j, x_i)$)

2 Linealidad-

3 Positividad:

$$\langle f, f \rangle = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i, x_j) \geq 0$$

y $\langle f, f \rangle = 0$ sólo si $f = 0$ 

1 Imagen de ϕ

- 1 Imagen de $\phi \longrightarrow$ espacio vectorial. ✓

- ① Imagen de $\phi \longrightarrow$ espacio vectorial. ✓
- ② Producto punto ✓

RKHS

- ① Imagen de $\phi \rightarrow$ espacio vectorial. ✓
- ② Producto punto ✓
- ③ $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$ ✓

- ① Imagen de $\phi \longrightarrow$ espacio vectorial. ✓
- ② Producto punto ✓
- ③ $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$ ✓
 - ▶ Hasta ahora el espacio \mathcal{H} es un espacio **pre-Hilbert**.

- ① Imagen de $\phi \rightarrow$ espacio vectorial. ✓
- ② Producto punto ✓
- ③ $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$ ✓
 - ▶ Hasta ahora el espacio \mathcal{H} es un espacio **pre-Hilbert**.
- ④ Completar espacio: añadir puntos límite de secuencias convergentes en la norma $\|f\| = \sqrt{\langle f, f \rangle}$.

- ① Imagen de $\phi \rightarrow$ espacio vectorial. ✓
- ② Producto punto ✓
- ③ $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$ ✓
 - ▶ Hasta ahora el espacio \mathcal{H} es un espacio **pre-Hilbert**.
- ④ Completar espacio: añadir puntos límite de secuencias convergentes en la norma $\|f\| = \sqrt{\langle f, f \rangle}$.
 - ▶ Obtenemos espacio de Hilbert llamado **Espacio de kernel reproductor**.

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$.

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{H} es un RKHS si existe una función $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ con las siguientes propiedades:

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{H} es un RKHS si existe una función $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ con las siguientes propiedades:

- 1 k tiene la propiedad de reproducción:

$$\langle k(., x), f \rangle_{\mathcal{H}} =$$

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{H} es un RKHS si existe una función $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ con las siguientes propiedades:

- 1 k tiene la propiedad de reproducción:

$$\langle k(\cdot, x), f \rangle_{\mathcal{H}} = f(x) \quad \forall f \in \mathcal{H}$$

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{H} es un RKHS si existe una función $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ con las siguientes propiedades:

- 1 k tiene la propiedad de reproducción:

$$\langle k(\cdot, x), f \rangle_{\mathcal{H}} = f(x) \quad \forall f \in \mathcal{H}$$

en particular,

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

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{H} es un RKHS si existe una función $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ con las siguientes propiedades:

- 1 k tiene la propiedad de reproducción:

$$\langle k(\cdot, x), f \rangle_{\mathcal{H}} = f(x) \quad \forall f \in \mathcal{H}$$

en particular,

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

RKHS (definición general)

Definición

Sea \mathcal{X} un conjunto no vacío, y \mathcal{H} un espacio de Hilbert de funciones $f : \mathcal{X} \rightarrow \mathbb{R}$. \mathcal{H} es un RKHS si existe una función $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ con las siguientes propiedades:

- 1 k tiene la propiedad de reproducción:

$$\langle k(., x), f \rangle_{\mathcal{H}} = f(x) \quad \forall f \in \mathcal{H}$$

en particular,

$$\langle k(., x), k(., x') \rangle_{\mathcal{H}} = k(x, x')$$

- 2 k **expande** \mathcal{H} , es decir $\mathcal{H} = \overline{\text{span} \{k(x, .) : x \in \mathcal{X}\}}$.

En la otra dirección

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

- Podemos obtener un kernel positivo definido:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

- Podemos obtener un kernel positivo definido:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

Note que $\forall x_1, x_2, \dots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \dots, c_n \in \mathbb{R}$:

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

- Podemos obtener un kernel positivo definido:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

Note que $\forall x_1, x_2, \dots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \dots, c_n \in \mathbb{R}$:

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j)$$

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

- Podemos obtener un kernel positivo definido:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

Note que $\forall x_1, x_2, \dots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \dots, c_n \in \mathbb{R}$:

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) = \left\langle \sum_{i=1}^n c_i \phi(x_i), \sum_{j=1}^n c_j \phi(x_j) \right\rangle_{\mathcal{H}}$$

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

- Podemos obtener un kernel positivo definido:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

Note que $\forall x_1, x_2, \dots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \dots, c_n \in \mathbb{R}$:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) &= \left\langle \sum_{i=1}^n c_i \phi(x_i), \sum_{j=1}^n c_j \phi(x_j) \right\rangle_{\mathcal{H}} \\ &= \left\| \sum_{i=1}^n c_i \phi(x_i) \right\|_{\mathcal{H}}^2 \end{aligned}$$

En la otra dirección

- Suponga que tenemos un mapeo

$$\phi : \mathcal{X} \rightarrow \mathcal{H}$$

donde \mathcal{H} es un espacio con producto punto.

- Podemos obtener un kernel positivo definido:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

Note que $\forall x_1, x_2, \dots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \dots, c_n \in \mathbb{R}$:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) &= \left\langle \sum_{i=1}^n c_i \phi(x_i), \sum_{j=1}^n c_j \phi(x_j) \right\rangle_{\mathcal{H}} \\ &= \left\| \sum_{i=1}^n c_i \phi(x_i) \right\|_{\mathcal{H}}^2 \geq 0 \end{aligned}$$

Teorema

(Mercer, 1911) Sea (\mathcal{X}, μ) un espacio de medida finito.

Teorema

(Mercer, 1911) Sea (\mathcal{X}, μ) un espacio de medida finito. Suponga que $k \in L_\infty(\mathcal{X}^2)$ es una función real simétrica tal que el operador integral

$$T_k : L_2(\mathcal{X}) \rightarrow L_2(\mathcal{X})$$
$$(T_k f)(x) = \int_{\mathcal{X}} k(x, x') f(x') d\mu(x')$$

es positivo definido

Teorema

(Mercer, 1911) Sea (\mathcal{X}, μ) un espacio de medida finito. Suponga que $k \in L_\infty(\mathcal{X}^2)$ es una función real simétrica tal que el operador integral

$$T_k : L_2(\mathcal{X}) \rightarrow L_2(\mathcal{X})$$
$$(T_k f)(x) = \int_{\mathcal{X}} k(x, x') f(x') d\mu(x')$$

es positivo definido, es decir, para toda $f \in L_2(\mathcal{X})$ tenemos

$$\int_{\mathcal{X}^2} k(x, x') f(x) f(x') d\mu(x) d\mu(x') \geq 0$$

Teorema

(Mercer, 1911) Sea (\mathcal{X}, μ) un espacio de medida finito. Suponga que $k \in L_\infty(\mathcal{X}^2)$ es una función real simétrica tal que el operador integral

$$T_k : L_2(\mathcal{X}) \rightarrow L_2(\mathcal{X})$$
$$(T_k f)(x) = \int_{\mathcal{X}} k(x, x') f(x') d\mu(x')$$

es positivo definido, es decir, para toda $f \in L_2(\mathcal{X})$ tenemos

$$\int_{\mathcal{X}^2} k(x, x') f(x) f(x') d\mu(x) d\mu(x') \geq 0$$

Sean $\psi \in L_2(\mathcal{X})$ las funciones propias ortogonales normalizadas de T_k con valores propios $\lambda_j > 0$, ordenados de manera no creciente.

Teorema

(Mercer, 1911) Sea (\mathcal{X}, μ) un espacio de medida finito. Suponga que $k \in L_\infty(\mathcal{X}^2)$ es una función real simétrica tal que el operador integral

$$T_k : L_2(\mathcal{X}) \rightarrow L_2(\mathcal{X})$$
$$(T_k f)(x) = \int_{\mathcal{X}} k(x, x') f(x') d\mu(x')$$

es positivo definido, es decir, para toda $f \in L_2(\mathcal{X})$ tenemos

$$\int_{\mathcal{X}^2} k(x, x') f(x) f(x') d\mu(x) d\mu(x') \geq 0$$

Sean $\psi \in L_2(\mathcal{X})$ las funciones propias ortogonales normalizadas de T_k con valores propios $\lambda_j > 0$, ordenados de manera no creciente.

Entonces:

- 1 La secuencia $\{\lambda_j\}_j$ es absolutamente sumable.

Entonces:

- 1 La secuencia $\{\lambda_j\}_j$ es absolutamente sumable.
- 2 $k(x, x') = \sum_{j=1}^N \psi_j(x)\psi_j(x')$ es válida para casi todo (x, x') . $N \in \mathbb{N}$ o $N = \infty$.

Entonces:

- 1 La secuencia $\{\lambda_j\}_j$ es absolutamente sumable.
- 2 $k(x, x') = \sum_{j=1}^N \psi_j(x)\psi_j(x')$ es válida para casi todo (x, x') . $N \in \mathbb{N}$ o $N = \infty$. En este último caso la serie converge absolutamente y uniformemente para casi todo (x, x') .

Entonces:

- 1 La secuencia $\{\lambda_j\}_j$ es absolutamente sumable.
- 2 $k(x, x') = \sum_{j=1}^N \psi_j(x)\psi_j(x')$ es válida para casi todo (x, x') . $N \in \mathbb{N}$ o $N = \infty$. En este último caso la serie converge absolutamente y uniformemente para casi todo (x, x') .

Entonces:

- ① La secuencia $\{\lambda_j\}_j$ es absolutamente sumable.
 - ② $k(x, x') = \sum_{j=1}^N \psi_j(x)\psi_j(x')$ es válida para casi todo (x, x') . $N \in \mathbb{N}$ o $N = \infty$. En este último caso la serie converge absolutamente y uniformemente para casi todo (x, x') .
-
- Es decir, si el operador integral definido por k es positivo definido, $k(x, x')$ corresponde a un producto punto de secuencias (vectores) en l_2^N . Tenemos $\langle \phi(x), \phi(x') \rangle$ con

Entonces:

- ① La secuencia $\{\lambda_j\}_j$ es absolutamente sumable.
 - ② $k(x, x') = \sum_{j=1}^N \psi_j(x)\psi_j(x')$ es válida para casi todo (x, x') . $N \in \mathbb{N}$ o $N = \infty$. En este último caso la serie converge absolutamente y uniformemente para casi todo (x, x') .
-
- Es decir, si el operador integral definido por k es positivo definido, $k(x, x')$ corresponde a un producto punto de secuencias (vectores) en l_2^N . Tenemos $\langle \phi(x), \phi(x') \rangle$ con

$$\begin{aligned}\phi : \mathcal{X} &\rightarrow l_2^N \\ x &\mapsto [\sqrt{\lambda_1}\psi_1(x) \quad \sqrt{\lambda_2}\psi_2(x) \quad \dots]\end{aligned}$$