Kernels

Fernando Lozano

Universidad de los Andes

October 12, 2022

Kernels

- Suponga que existe un kernel:

$$
\begin{aligned}
k \; : \; & \mathcal{X} \times \mathcal{X} \to \mathbb{R} \\
& (\mathrm{x}_1, \mathrm{x}_2) \mapsto k(\mathrm{x}_1, \mathrm{x}_2) = \langle \phi(\mathrm{x}_i), \phi(\mathrm{x}_j) \rangle_{\mathcal{H}}
\end{aligned}
$$

Kernels

- Suponga que existe un kernel:

$$
\begin{aligned}
k \;:\; & \mathcal{X} \times \mathcal{X} \to \mathbb{R} \\
& (\mathsf{x}_1, \mathsf{x}_2) \mapsto k(\mathsf{x}_1, \mathsf{x}_2) = \langle \phi(\mathsf{x}_i), \phi(\mathsf{x}_j) \rangle_{\mathcal{H}}
\end{aligned}
$$

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

Kernels

- Suponga que existe un kernel:

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

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

Kernels

- Suponga que existe un kernel:

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

- Podemos calcular el producto punto $\langle \phi(\mathrm{x}_i), \phi(\mathrm{x}_j) \rangle_{\mathcal{H}}$ 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.

Kernels

- Suponga que existe un kernel:

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

- Podemos calcular el producto punto $\langle \phi(\mathrm{x}_i), \phi(\mathrm{x}_j) \rangle_{\mathcal{H}}$ 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

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

El truco del Kernel

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

El truco del Kernel

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

El truco del Kernel

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i\alpha_j y_i y_j {\color{red} k(\mathrm{x}_i, \mathrm{x}_j)} \\ \text{sujeto a} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0 \\ & 0 \le \alpha_i \le C \end{aligned}$$

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

El truco del Kernel

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

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

El truco del Kernel

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

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

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

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

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

El truco del Kernel

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

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

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

El truco del Kernel

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

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

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

Ejemplo

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

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

Ejemplo

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

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

Comprobando:

Ejemplo

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

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

Comprobando:

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

Ejemplo

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

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

Comprobando:

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

Ejemplo

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

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

Comprobando:

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

Ejemplo

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

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

Comprobando:

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

- 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 $k(\mathsf{x}_i, \mathsf{x}_j) = \langle \mathsf{x}_i, \mathsf{x}_j \rangle^2$, 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 $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle^2$, podemos tener:
  - $\mathcal{H} = \mathbb{R}^4$ y

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

- 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 $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle^2$, podemos tener:
  - $\mathcal{H} = \mathbb{R}^4$ y

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

- 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 $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle^2$, podemos tener:
  - $\mathcal{H} = \mathbb{R}^4$ y

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

✍

  - $\mathcal{H} = \mathbb{R}^2$ y

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

- 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 $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle^2$, podemos tener:
  - $\mathcal{H} = \mathbb{R}^4$ y

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

✍

  - $\mathcal{H} = \mathbb{R}^2$ y

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

✍

Ejemplos de kernels

Ejemplos de kernels

- Polinomial:

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

Ejemplos de kernels

- Polinomial:

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

- Polinomial no homogéneo:

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

Ejemplos de kernels

- Polinomial:

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

- Polinomial no homogéneo:

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

- Gaussiano (RBF):

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

Ejemplos de kernels

- Polinomial:

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

- Polinomial no homogéneo:

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

- Gaussiano (RBF):

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

- Sigmoidal:

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

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).

RKHS

Definición

Dada una función simétrica $k : \mathcal{X}^2 \to \mathbb{R}$ *(*$k(x_1, x_2) = k(x_2, x_1)$*) y* $x_1, x_2, \ldots, x_n \in \mathcal{X}$*, 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* $x_1, x_2, \ldots, x_n$*.*

RKHS

Definición

Dada una función simétrica $k : \mathcal{X}^2 \to \mathbb{R}$ $(k(x_1, x_2) = k(x_2, x_1))$ *y* $x_1, x_2, \ldots, x_n \in \mathcal{X}$, *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* $x_1, x_2, \ldots, x_n$.

Definición

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

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle$

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathsf{x}_i, \mathsf{x}_j) = \langle \mathsf{x}_i, \mathsf{x}_j \rangle$
- Matriz de Gram para $\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n$:

$$\mathsf{K} = \begin{bmatrix} \langle \mathsf{x}_1, \mathsf{x}_1 \rangle & \langle \mathsf{x}_1, \mathsf{x}_2 \rangle & \cdots & \langle \mathsf{x}_1, \mathsf{x}_n \rangle \\ \langle \mathsf{x}_2, \mathsf{x}_1 \rangle & \langle \mathsf{x}_2, \mathsf{x}_2 \rangle & \cdots & \langle \mathsf{x}_2, \mathsf{x}_n \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \mathsf{x}_n, \mathsf{x}_1 \rangle & \langle \mathsf{x}_n, \mathsf{x}_2 \rangle & \cdots & \langle \mathsf{x}_n, \mathsf{x}_n \rangle \end{bmatrix}$$

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle$
- Matriz de Gram para $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$:

$$\mathrm{K} = \begin{bmatrix} \langle \mathrm{x}_1, \mathrm{x}_1 \rangle & \langle \mathrm{x}_1, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_1, \mathrm{x}_n \rangle \\ \langle \mathrm{x}_2, \mathrm{x}_1 \rangle & \langle \mathrm{x}_2, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_2, \mathrm{x}_n \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \mathrm{x}_n, \mathrm{x}_1 \rangle & \langle \mathrm{x}_n, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_n, \mathrm{x}_n \rangle \end{bmatrix} = \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix}$$

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle$
- Matriz de Gram para $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$:

$$\mathrm{K} = \begin{bmatrix} \langle \mathrm{x}_1, \mathrm{x}_1 \rangle & \langle \mathrm{x}_1, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_1, \mathrm{x}_n \rangle \\ \langle \mathrm{x}_2, \mathrm{x}_1 \rangle & \langle \mathrm{x}_2, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_2, \mathrm{x}_n \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \mathrm{x}_n, \mathrm{x}_1 \rangle & \langle \mathrm{x}_n, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_n, \mathrm{x}_n \rangle \end{bmatrix} = \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix}$$

- Para un vector arbitrario $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_n \end{bmatrix}^T$,

$$\boldsymbol{\alpha}^T \mathrm{K} \boldsymbol{\alpha} =$$

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle$
- Matriz de Gram para $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$:

$$\mathrm{K} = \begin{bmatrix} \langle \mathrm{x}_1, \mathrm{x}_1 \rangle & \langle \mathrm{x}_1, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_1, \mathrm{x}_n \rangle \\ \langle \mathrm{x}_2, \mathrm{x}_1 \rangle & \langle \mathrm{x}_2, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_2, \mathrm{x}_n \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \mathrm{x}_n, \mathrm{x}_1 \rangle & \langle \mathrm{x}_n, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_n, \mathrm{x}_n \rangle \end{bmatrix} = \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix}$$

- Para un vector arbitrario $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_n \end{bmatrix}^T$,

$$\boldsymbol{\alpha}^T \mathrm{K} \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix} \boldsymbol{\alpha}$$

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle$
- Matriz de Gram para $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$:

$$\mathrm{K} = \begin{bmatrix} \langle \mathrm{x}_1, \mathrm{x}_1 \rangle & \langle \mathrm{x}_1, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_1, \mathrm{x}_n \rangle \\ \langle \mathrm{x}_2, \mathrm{x}_1 \rangle & \langle \mathrm{x}_2, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_2, \mathrm{x}_n \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \mathrm{x}_n, \mathrm{x}_1 \rangle & \langle \mathrm{x}_n, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_n, \mathrm{x}_n \rangle \end{bmatrix} = \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix}$$

- Para un vector arbitrario $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_n \end{bmatrix}^T$,

$$\boldsymbol{\alpha}^T \mathrm{K} \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix} \boldsymbol{\alpha} = \left\| \sum_{i=1}^{n} \alpha_i \mathrm{x}_i \right\|^2$$

Ejemplo

- $\mathcal{X} = \mathbb{R}^n$, $k(\mathrm{x}_i, \mathrm{x}_j) = \langle \mathrm{x}_i, \mathrm{x}_j \rangle$
- Matriz de Gram para $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$:

$$\mathrm{K} = \begin{bmatrix} \langle \mathrm{x}_1, \mathrm{x}_1 \rangle & \langle \mathrm{x}_1, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_1, \mathrm{x}_n \rangle \\ \langle \mathrm{x}_2, \mathrm{x}_1 \rangle & \langle \mathrm{x}_2, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_2, \mathrm{x}_n \rangle \\ \vdots & \ddots & \ddots & \vdots \\ \langle \mathrm{x}_n, \mathrm{x}_1 \rangle & \langle \mathrm{x}_n, \mathrm{x}_2 \rangle & \cdots & \langle \mathrm{x}_n, \mathrm{x}_n \rangle \end{bmatrix} = \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix}$$

- Para un vector arbitrario $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_n \end{bmatrix}^T$,

$$\boldsymbol{\alpha}^T \mathrm{K} \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \begin{bmatrix} \mathrm{x}_1^T \\ \mathrm{x}_2^T \\ \vdots \\ \mathrm{x}_n^T \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \cdots & \mathrm{x}_n \end{bmatrix} \boldsymbol{\alpha} = \left\| \sum_{i=1}^{n} \alpha_i \mathrm{x}_i \right\|^2 \geq 0$$

Desigualdad de Cauchy-Schwartz para kernels

Proposición

Si k es un kernel positivo definido, y $x_1, x_2 \in \mathcal{X}$ entonces:

$$|k(x_1, x_2)|^2 \leq k(x_1, x_1)k(x_2, x_2)$$

Desigualdad de Cauchy-Schwartz para kernels

Proposición

Si k es un kernel positivo definido, y $x_1, x_2 \in \mathcal{X}$ entonces:

$$|k(x_1, x_2)|^2 \leq k(x_1, x_1)k(x_2, x_2)$$

Proof.

Desigualdad de Cauchy-Schwartz para kernels

Proposición

Si k es un kernel positivo definido, y $x_1, x_2 \in \mathcal{X}$ entonces:

$$|k(x_1, x_2)|^2 \leq k(x_1, x_1)k(x_2, x_2)$$

Proof.

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) \\ k(x_1, x_2) & k(x_2, x_2) \end{bmatrix} \geq 0$$

Desigualdad de Cauchy-Schwartz para kernels

Proposición

Si k es un kernel positivo definido, y $x_1, x_2 \in \mathcal{X}$ entonces:

$$|k(x_1, x_2)|^2 \leq k(x_1, x_1)k(x_2, x_2)$$

Proof.

$$\mathsf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) \\ k(x_1, x_2) & k(x_2, x_2) \end{bmatrix} \geq 0 \Rightarrow \det \mathsf{K} \geq 0$$

□

El mapa del kernel reproductor

- Kernel positivo definido k.

El mapa del kernel reproductor

- Kernel positivo definido k.
- Mapeo:

$$\phi \, : \, \mathcal{X} \to \mathbb{R}^{\mathcal{X}}$$

El mapa del kernel reproductor

- Kernel positivo definido k.
- Mapeo:

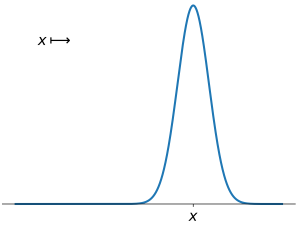
$$\begin{aligned} \phi \ : \ \mathcal{X} &\rightarrow \mathbb{R}^{\mathcal{X}} \\ x &\mapsto k(.,x) \end{aligned}$$

El mapa del kernel reproductor

- Kernel positivo definido k.
- Mapeo:

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

Ejemplo:



Receta:

Receta:

1. Imagen de ϕ

Receta:

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

Receta:

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

Receta:

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

Receta:

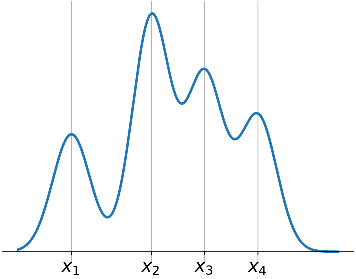
1. Imagen de $\phi \longrightarrow$ espacio vectorial.
2. Producto punto.
3. $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$
4. Completar espacio.

- Vectores:

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

- Vectores:

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



- Producto punto:

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

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

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

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

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

- Note que:

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

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

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

- Note que:

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

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

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

- Note que:

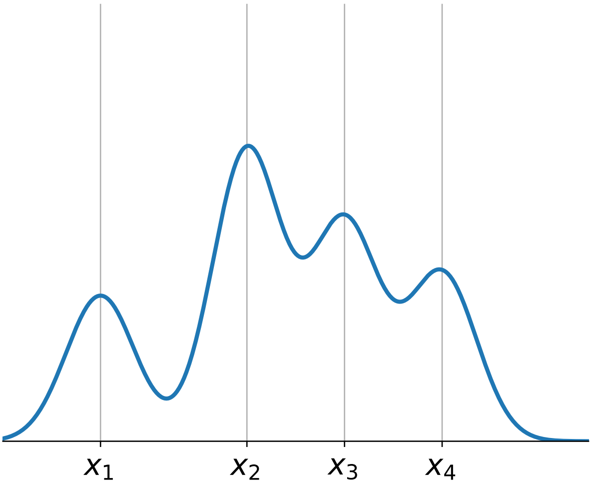
$$\langle f, g \rangle = \sum_{j=1}^{m} \beta_j \sum_{i=1}^{n} \alpha_i k(x_i, x_j') = \sum_{j=1}^{m} \beta_j f(x_j') = \sum_{i=1}^{n} \alpha_i g(x_i)$$

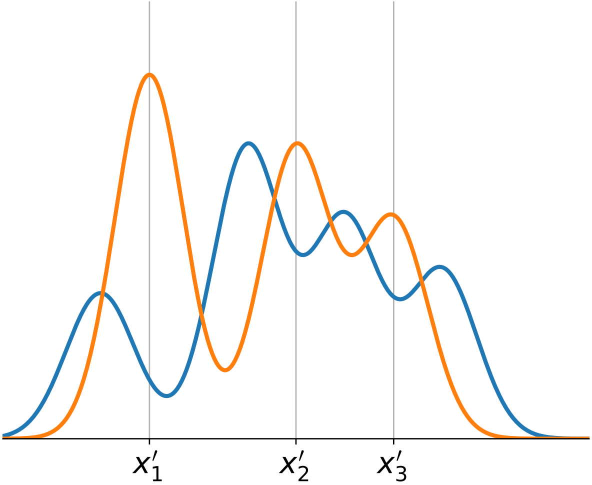
- Producto punto: Si $g(.) = \sum_{j=1}^{m} \beta_i k(., x_i')$ definimos:

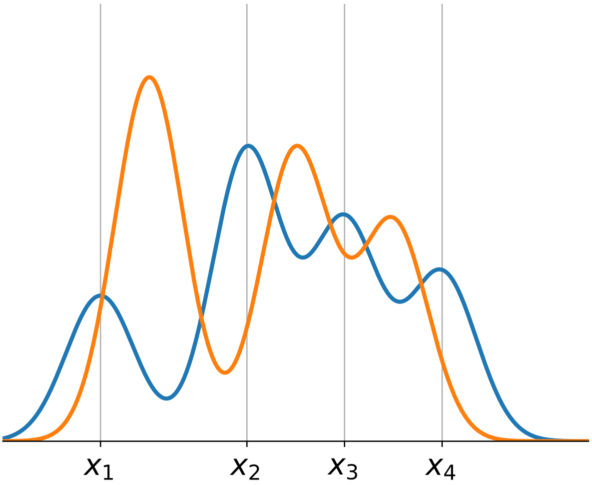
$$\langle f, g \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, x_j')$$

- Note que:

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


x_1
x_2
x_3
x_4


x_1'
x_2'
x_3'

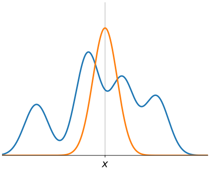


x_1
x_2
x_3
x_4

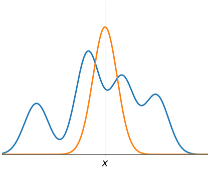
$$\langle k(., x), f\rangle =$$

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

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

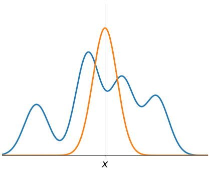


$$\langle k(.,x), f \rangle = f(x)$$



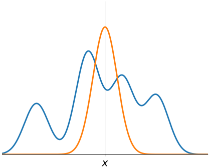
$$\left\langle k(.,x), k(.,x') \right\rangle =$$

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

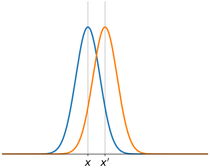


$$\left\langle k(.,x), k(.,x') \right\rangle = k(x,x')$$

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



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



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

- Esta es la propiedad del Kernel Reproductor.

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

- Esta es la propiedad del Kernel Reproductor.
- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una “fila“ de esta matriz.

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

- Esta es la propiedad del Kernel Reproductor.
- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(.,x)$ como una “fila“ de esta matriz.
- Sea $\mathrm{Q} \in \mathbb{R}^{n\times n}$ una matriz positiva definida. El producto punto en $\mathbb{R}^n$:

$$\langle \mathrm{x}_i, \mathrm{x}_j\rangle_{\mathrm{Q}} = \mathrm{x}_i^T \mathrm{Q} \mathrm{x}_j$$

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

- Esta es la propiedad del Kernel Reproductor.
- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(.,x)$ como una "fila" de esta matriz.
- Sea $\mathrm{Q} \in \mathbb{R}^{n \times n}$ una matriz positiva definida. El producto punto en $\mathbb{R}^n$:

$$\langle \mathrm{x}_i, \mathrm{x}_j \rangle_{\mathrm{Q}} = \mathrm{x}_i^T \mathrm{Q} \mathrm{x}_j$$

es un RKHS en el espacio de columnas de Q^{-1}. Si $\mathrm{u}_i, \mathrm{u}_j$ son columnas de Q^{-1}:

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

- Esta es la propiedad del Kernel Reproductor.
- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(.,x)$ como una "fila" de esta matriz.
- Sea $\mathrm{Q} \in \mathbb{R}^{n\times n}$ una matriz positiva definida. El producto punto en $\mathbb{R}^n$:

$$\langle \mathrm{x}_i, \mathrm{x}_j \rangle_{\mathrm{Q}} = \mathrm{x}_i^T \mathrm{Q} \mathrm{x}_j$$

es un RKHS en el espacio de columnas de Q^{-1}. Si $\mathrm{u}_i, \mathrm{u}_j$ son columnas de Q^{-1}:

$$\langle \mathrm{u}_i, \mathrm{u}_j \rangle_{\mathrm{Q}} =$$

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

- Esta es la propiedad del Kernel Reproductor.
- Podemos interpretar $k(x_i, x_j)$ como una matriz de infinitas dimensiones, y $k(., x)$ como una “fila“ de esta matriz.
- Sea $\mathrm{Q} \in \mathbb{R}^{n \times n}$ una matriz positiva definida. El producto punto en $\mathbb{R}^n$:

$$\langle \mathrm{x}_i, \mathrm{x}_j \rangle_{\mathrm{Q}} = \mathrm{x}_i^T \mathrm{Q} \mathrm{x}_j$$

  es un RKHS en el espacio de columnas de Q^{-1}. Si $\mathrm{u}_i, \mathrm{u}_j$ son columnas de Q^{-1}:

$$\langle \mathrm{u}_i, \mathrm{u}_j \rangle_{\mathrm{Q}} = q_{ij}$$

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 $\langle f, f \rangle = 0$ implica $f = 0$:

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 $\langle f, f \rangle = 0$ implica $f = 0$:
$$|f(x)|^2 = |\langle k(.,x), f \rangle|^2$$

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 $\langle f, f \rangle = 0$ implica $f = 0$:
$$|f(x)|^2 = |\langle k(., x), f \rangle|^2 \leq k(x, x) \langle f, f \rangle$$

RKHS

RKHS

1. Imagen de ϕ

RKHS

1. Imagen de $\phi \longrightarrow$ espacio vectorial. ✔

RKHS

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

RKHS

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

RKHS

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.

RKHS

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.
4. Completar espacio: añadir puntos límite de secuencias convergentes en la norma $\|f\| = \sqrt{\langle f, f \rangle}$.

RKHS

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.
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 $f : \mathcal{X} \to \mathbb{R}$.

RKHS (definición general)

Definición

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

RKHS (definición general)

Definición

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

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

RKHS (definición general)

Definición

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

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

RKHS (definición general)

Definición

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

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

RKHS (definición general)

Definición

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

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

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

En la otra dirección

En la otra dirección

- Suponga que tenemos un mapeo

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

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

En la otra dirección

- Suponga que tenemos un mapeo

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

  donde $\mathcal{H}$ es un espacio con producto punto.
- Podemos obtener un kernel positivo definido:

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

En la otra dirección

- Suponga que tenemos un mapeo

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

  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 $\forall x_i, x_2, \ldots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \ldots, c_n \in \mathbb{R}$:

En la otra dirección

- Suponga que tenemos un mapeo

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

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 $\forall x_i, x_2, \ldots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \ldots, c_n \in \mathbb{R}$:

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

En la otra dirección

- Suponga que tenemos un mapeo

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

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 $\forall x_i, x_2, \ldots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \ldots, c_n \in \mathbb{R}$:

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

En la otra dirección

- Suponga que tenemos un mapeo

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

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 $\forall x_i, x_2, \ldots, x_n \in \mathcal{X}$ y $\forall c_1, c_2, \ldots, c_n \in \mathbb{R}$:

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

$$= \left\| \sum_{i=1}^{n} c_i \phi(x_i) \right\|^2$$

En la otra dirección

- Suponga que tenemos un mapeo

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

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

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

$$= \left\| \sum_{i=1}^{n} c_i \phi(x_i) \right\|^2 \geq 0$$

Teorema

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

Teorema

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

$$\begin{aligned} T_k \; &: L_2(\mathcal{X}) \to L_2(\mathcal{X}) \\ (T_k f)(x) &= \int_{\mathcal{X}} k(x, x') f(x') d\mu(x') \end{aligned}$$

es positivo definido

Teorema

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

$$\begin{aligned} T_k &: L_2(\mathcal{X}) \to L_2(\mathcal{X}) \\ (T_k f)(x) &= \int_{\mathcal{X}} k(x, x') f(x') d\mu(x') \end{aligned}$$

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

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

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

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

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

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