Reproducing Kernel Hilbert Space

JhuoW

03-25 2022

1 Hilbert Space

Definition 1 (Norm)**.** Let $\mathcal{F}$ be a vector space over $\mathbb{R}$ (For example $\mathcal{F} = \mathbb{R}^n$ is a vector space). A function $||\cdot||_{\mathcal{F}} : \mathcal{F} \to [0, \inf)$ is said to be a *norm* on $\mathcal{F}$ if ($||\cdot||_{\mathcal{F}}$ 是一个有效 norm 算子要满足以下条件)

1. For $f \in \mathcal{F}$, $||f||_{\mathcal{F}} = 0$ if and only if $f = 0$. (*norm separates points*)
2. $\|\lambda f\|_{\mathcal{F}} = |\lambda|\|f\|_{\mathcal{F}}$, $\forall \lambda \in \mathbb{R}, \forall f \in \mathcal{F}$ (positive homogeneity).
3. $\|f + g\|_{\mathcal{F}} \leq \|f\|_{\mathcal{F}} + \|g\|_{\mathcal{F}}, \forall f, g \in \mathcal{F}$ (triangle inequality).

向 $||\cdot||_{\mathcal{F}}$ 中输入任意一个向量，只要满足以上条件，那么 $||\cdot||_{\mathcal{F}}$ 是一个 valid norm operator.

1.1 Inner Product

An *inner product* takes two elements of a vector space $\mathcal{X}$ and outputs a number. An inner product could be a usual dot product: $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}'\mathbf{v} = \sum_i u^{(i)} v^{(i)}$ (Inner Product can be Dot Product). Or the inner product could be something fancier（即内积不一定表示为点积的形式）. If an Inner Product $\langle \cdot, \cdot \rangle$ is valid, it **MUST** satisfy the following conditions:

1. Symmetry
$$\langle u, v \rangle = \langle v, u \rangle \quad \forall u, v \in \mathcal{X}$$
2. Bilinearity
$$\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle \quad \forall u, v, w \in \mathcal{X}, \forall \alpha, \beta \in \mathbf{R}$$
3. Strict Positive Definiteness
$$\langle u, u \rangle \geq 0 \forall x \in \mathcal{X}$$
$$\langle u, u \rangle = 0 \Longleftrightarrow u = 0$$

An *inner product space* (or pre-Hilbert space) is a vector space together with an inner product. (包含内积运算的向量空间称为内积空间，即可以定义内积运算的向量空间)。

Kernel is a kind of Inner Product. For example, the Gaussian kernel is defined as:

$$\langle u, v\rangle = k(u, v) = \exp -\frac{||u-v||^2}{2\sigma} \tag{1}$$

1.2 Hilbert space

Definition 2 (Hilbert Space)**.** A Hilbert Space is an **Inner Product space** that is complete and separable with respect to the norm defined by the inner product.

'Complete' means sequences converge to elements of the space - there aren't any "holes" in the space.

2 Finite States

Given finite input space $\{x_1, x_2, \cdots x_m\}$. I want to be able to take inner products between any two of them using my function k as the inner product (k is customized and satisfy three conditions. For example, k is a Gaussian inner product as Eq.(1)). Inner products by definition are symmetric, so $k(x_i, x_j) = k(x_j, x_i)$, which yields a symmetric matrix $\mathbf{K}$.

Since $\mathbf{K}$ is real symmetric, and this means we can diagonalize it （实对称阵可以对角化，即特征分解）, and the eigendecomposition takes this form:

$$\begin{aligned}
\mathbf{K} &= \mathbf{V}\Lambda\mathbf{V}^T \\
&= \mathbf{V}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_m \end{bmatrix}\mathbf{V}^T \\
&= \begin{bmatrix} v_1 & v_2 & \cdots v_m \end{bmatrix}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_m \end{bmatrix}\begin{bmatrix} v_1^T \\ v_2^T \\ \cdots \\ v_m^T \end{bmatrix} \\
&= v_1\lambda_1 v_1^T + \cdots + v_m\lambda_m v_m^T = \sum_{t=1}^{m} v_t\lambda_t v_t^T
\end{aligned} \tag{2}$$

Let the i-th element of vector v as $v^{(i)}$, then

$$\begin{aligned}\mathbf{K}_{ij} = k(x_i, x_j) &= [\sum_{t=1}^{m} v_t \lambda_t v_t^T]_{ij} \\ &= \sum_{t=1}^{m} v_t^{(i)} \lambda_t v_t^{(j)}\end{aligned} \tag{3}$$

If $\mathbf{K}$ is a **positive semi-definite** (**PSD**) matrix, then $\lambda_1, \cdots \lambda_m \geq 0$.

Assume 1. All λ_t are nonnegative.

We consider this feature map:

$$\Phi(x_i) = \left[\sqrt{\lambda_1} v_1^{(i)}, \ldots, \sqrt{\lambda_t} v_t^{(i)}, \ldots, \sqrt{\lambda_m} v_m^{(i)}\right] \in \mathbb{R}^m \tag{4}$$

(writing it for x_j too):

$$\boldsymbol{\Phi}(x_j) = \left[\sqrt{\lambda_1} v_1^{(j)}, \ldots, \sqrt{\lambda_t} v_t^{(j)}, \ldots, \sqrt{\lambda_m} v_m^{(j)}\right] \in \mathbb{R}^m \tag{5}$$

即 $\Phi : \mathcal{X} \to \mathbb{R}^m$ 将 $x \in \mathcal{X}$ 映射到 m 维向量空间 $\mathbb{R}^m$ 中的一个点。

With this choice, the inner product k is just defined as a dot product in $\mathbb{R}^m$:

$$\langle \Phi(x_i), \Phi(x_j) \rangle_{\mathbf{R}^m} = \sum_{t=1}^{m} \lambda_t v_t^{(i)} v_t^{(j)} = \left(\mathbf{V} \Lambda \mathbf{V}'\right)_{ij} = K_{ij} = k(x_i, x_j) \tag{6}$$

If there exists an eigenvalue $\lambda_s < 0$ (即 $\sqrt{\lambda_s} = \sqrt{|\lambda_s|} i$). λ_s 对应的特征向量 v_s。用 $v_s \in \mathbb{R}^m$ 的 m 个元素 $v_s = [v_s^{(1)}, \cdots, v_s^{(m)}]$, 来对 $\Phi(x_1), \cdots, \Phi(x_m)$ 做线性组合:

$$\mathbf{z} = \sum_{i=1}^{m} v_s^{(i)} \boldsymbol{\Phi}(x_i) \tag{7}$$

It is obvious that $\mathbf{z} \in \mathbb{R}^m$. Then calculate

$$\begin{aligned}\|\mathbf{z}\|_2^2 = \langle \mathbf{z}, \mathbf{z} \rangle_{\mathbf{R}^m} &= \sum_i \sum_j v_s^{(i)} \boldsymbol{\Phi}(x_i)^T \boldsymbol{\Phi}(x_j) v_s^{(j)} = \sum_i \sum_j v_s^{(i)} K_{ij} v_s^{(j)} \\ &= \mathbf{v}_s^T \mathbf{K} \mathbf{v}_s = \lambda_s < 0\end{aligned} \tag{8}$$

which conflicts with the geometry of the feature space.

如果 $\mathbf{K}$ 不是半正定，那么 feature space $\mathbb{R}^m$ 存在小于 0 的值。所以假设 Assumption 不成立。即，若 k 表示有限集的内积，那么它的 Gram Matrix 一定半正定 (PSD)，否则无法保证该空间中的 norm 大于 0。

有效的内积对应的 Gram Matrix 必定 PSD.

3 Kernel

Definition 3 (Kernel)**.** A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a *kernel* if

1. k is symmetric: $k(x, y) = k(y, x)$.
2. k gives rise to a positive semi-definite "Gram matrix," i.e., for any $m \in \mathbb{N}$ and any $x_1, \cdots, x_m$ chosen from X, the Gram matrix $\mathbf{K}$ defined by $\mathbf{K}_{ij} = k(x_i, x_j)$ is positive semi-definite.

Another way to show that a matrix $\mathbf{K}$ is positive semi-definite is to show that

$$\forall \mathbf{c} \in \mathbf{R}^m, \mathbf{c}^T \mathbf{K} \mathbf{c} \geq 0 \tag{9}$$

Here are some nice properties of k:

- $k(u, u) \geq 0$ (Think about the Gram matrix of $m = 1$.)
- $k(u, v) \leq \sqrt{k(u, u)k(v, v)}$ (This is the Cauchy-Schwarz inequality.)

3.1 Reproducing Kernel Hilbert Space (RKHS)

给定一个 kernel $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. 定义一个函数空间（space of functions）$\mathbf{R}^{\mathcal{X}} := \{f : \mathcal{X} \to \mathbb{R}\}$. $\mathbf{R}^{\mathcal{X}}$ 是一个 Hilbert Space，该空间中的每个元素是一个 $\mathcal{X}$ 映射到 $\mathbb{R}$ 的函数。

令 $k_x(\cdot) = k(x, \cdot)$, 假设 x 是一个定值（Constant），自变量（输入）用 $\cdot$ 表示。那么 $k(x, \cdot)$ 也是 $\mathbf{R}^{\mathcal{X}}$ 空间中的一个函数。

每个函数 $k_x(\cdot)$ 都与一个特定的 $x \in \mathcal{X}$ 有关，即每个 x 对应于一个函数 $k_x(\cdot) = k(\cdot, x)$. 这种对应关系表示为 $\Phi(x) = k_x(\cdot) = k(x, \cdot)$, 即:

$$\Phi : x \longmapsto k(\cdot, x) \tag{10}$$

即 Φ 的输入为 $x \in \mathcal{X}$, 输出一个函数, 输出的函数属于 $\mathbf{R}^{\mathcal{X}}$ 空间。

在连续空间 $\mathcal{X}$ 中，$x \in \mathcal{X}$ 有无穷多种情况，那么 $\Phi(x) = k_x(\cdot) = k(x, \cdot)$ 也有无穷多种情况，即无穷多种函数。这些函数可以 span 一个 Hilbert Space:

$$\mathcal{H}_k = \operatorname{span}(\{\Phi(x) : x \in \mathcal{X}\}) = \left\{ f(\cdot) = \sum_{i=1}^{m} \alpha_i k\left(\cdot, x_i\right) : m \in \mathbf{N}, x_i \in \mathcal{X}, \alpha_i \in \mathbf{R} \right\} \tag{11}$$

其中 $k(x, \cdot) = \Phi(x)$ 可以理解为将 x 映射为一个函数（or vector）。上述 Hilbert Space 是由任意 $k(x, \cdot)$ 线性组合而成的函数空间，该空间中的每个元素可以表示为:

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i k\left(\cdot, x_i\right) \tag{12}$$

所以 $\mathcal{H}$ 可以看作是 kernel k 对应的一个 Hilbert Space。

给定 $\mathcal{H}$ 中的任意两个函数 $f(\cdot) = \sum_{i=1}^{m} \alpha_i k\left(\cdot, x_i\right)$, $g(\cdot) = \sum_{j=1}^{m'} \beta_j k\left(\cdot, x_j'\right)$。注意 $f(\cdot)$ 和 $g(\cdot)$ 可以表示 $\mathcal{H}$ 中任意两个元素。我们将 $\mathcal{H}$ 上的内积定义为:

$$\langle f, g \rangle_{\mathcal{H}_k} = \sum_{i=1}^{m} \sum_{j=1}^{m'} \alpha_i \beta_j k\left(x_i, x_j'\right) \tag{13}$$

由Proof证明了该内积符合三个条件，顾上式是 $\mathcal{H}$ 空间中一个有效的内积算子。注：$\mathcal{H}_k$ 表示该 Hilbert Space 是由函数 $k(x, \cdot)$ span 而成的，与 Kernel k 有关.

$k(x, \cdot)$ 也是 $\mathcal{H}_k$ 中的一个函数，那么它与 f 的内积为:

$$\langle k(\cdot, x), f \rangle_{\mathcal{H}_k} = \sum_{i=1}^{m} \alpha_i k(x, x_i) = f(x) \tag{14}$$

Theorem 1. $k(\cdot, \cdot)$ is a reproducing kernel of a Hilbert space $\mathcal{H}_k$ if $f(x) = \langle k(x, \cdot), f(\cdot) \rangle$.

$\mathcal{H}_k$ 为 $k(\cdot, \cdot)$ 的再生核希尔伯特空间。

同理，$k(\cdot, x_i)$, $k(\cdot, x_j)$ 都为 $\mathcal{H}_k$ 中的函数，计算他们的内积:

$$\langle k(\cdot, x_i), k\left(\cdot, x_j\right)\rangle_{\mathcal{H}_k} = k\left(x_i, x_j\right) \tag{15}$$

因为 $k(\cdot, x_i) = \Phi(x_i)$, $k(\cdot, x_j) = \Phi(x_j)$, 所以

$$k\left(x_i, x_j\right) = \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{H}_k} \tag{16}$$

表示将 x_i 和 x_j 映射成 $\mathcal{H}_k$ 中的函数（向量）后再做内积。