what we have discussed in kernel method?

① feature map : nonlinear → linear

② kernel function, kernel matrix.

③ Distance and angle in feature space.

④ a simple classifier.

⑤ Dual form, duality theorem, Langrange multiplier, KKT.

⑥ PCA and kernel based PCA.

⑦ LDA and GDA

⑧ SVM : hard margin, soft margin, Lagrangrian

⑨ Linear regression, kernel based linear regression.

⑩ theoretical aspeat of kernel method.

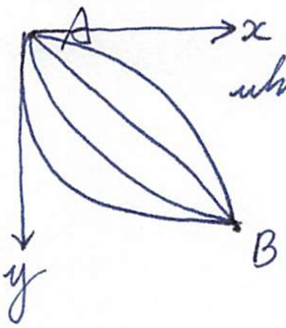
10.1 functional analysis.

10.2 位相、距離、ノルム、内積 and its space (Hilbert. Banach)

10.3 kernel property.

10.4 basic idea 内積 in High D space = f (内積 in low D space)

Brachistochrone curve 最速降下曲線 さいそくこうかきょくせん



which track

Potential Energy

$E_p = mgh$

位置(いち)エネルギー

kinetic energy

$E_k = \frac{1}{2} m v^2$

運動(うんどう)エネルギー

$E_p = E_k$

$mgy = \frac{1}{2} m v^2$

$v = \sqrt{2gh}$

$\because v = \frac{ds}{dt} \Rightarrow dt = \frac{ds}{v}$

$= \frac{\sqrt{dx^2 + dy^2}}{\sqrt{2gy}}$

$= \frac{\sqrt{1+y'^2}}{\sqrt{2gy}} dx$

$ds = \sqrt{dx^2 + dy^2}$

$\sqrt{dx^2 + dy^2} = dx\sqrt{1 + (\frac{dy}{dx})^2}$

$= dx\sqrt{1+y'^2}$

$\therefore T = \int dt = \int \frac{\sqrt{1+y'^2}}{\sqrt{2gy}} dx$

we want to find a y that let T to be minine value.

y is a set of many functions, it is a functions' function.

Distance 距離 きょり

X is ~~an~~ a non empty set. $\forall x, y \in X$, $\exists d(x, y) \in R$
and $d(x, y)$ satisfies that:

(1) $d(x, y) \geq 0$. $d(x, y) = 0 \Leftrightarrow x = y$; non-negativity 非負性.

(2) $d(x, y) = d(y, x)$ symmetry 对称性

(3) $d(x, y) \leq d(x, z) + d(z, y)$ triangle inequality 三角不等式

then $d(x, y)$ is a distance of x and y.

linear space 線型空間.

norm ノルム

$\|x\|$ is a norm of R^n, it satisfies:

(1) $\|x\| \geq 0 \quad \forall x \in R^n$; $\|x\| = 0 \Leftrightarrow x = 0$;

(2) $\|ax\| = |a| \|x\| \quad \forall a \in R \quad x \in R^n$

(3) $\|x+y\| \leq \|x\| + \|y\| \quad \forall x, y \in R^n$

norm can be considered as a distance from x to zero point.

From definitions of distance and norm, we can find:

① we can use norm to define distance:

$$d(x, y) = \|x - y\|$$

② however, we cannot use distance to define norm:

$$\|x\| = d(0, x)$$

but $\|ax\| = d(0, ax) \neq |a| \|x\|$

§2. 線 型 空 間

定 義. 集合 V がつぎの二条件 (I), (II) を充すとき，V を**複素線型空間**あるいは**複素ベクトル空間**と言う.

(I)　V の二元 $\boldsymbol{x}, \boldsymbol{y}$ に対して**和**と呼ばれる第三の元（これを $\boldsymbol{x}+\boldsymbol{y}$ で表わす）が定まり，つぎの法則が成立つ：

アーベル群

(1)　$(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z} = \boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})$　　（結合法則），

(2)　$\boldsymbol{x}+\boldsymbol{y} = \boldsymbol{y}+\boldsymbol{x}$　　（交換法則），

(3)　**零ベクトル**と呼ばれる特別な元（これを $\boldsymbol{o}$ で表わす）がただ一つ存在し，V のすべての元 $\boldsymbol{x}$ に対して，$\boldsymbol{o}+\boldsymbol{x} = \boldsymbol{x}$ が成立つ，

(4)　V の任意の元 $\boldsymbol{x}$ に対し，$\boldsymbol{x}+\boldsymbol{x}' = \boldsymbol{o}$ となる V の元 $\boldsymbol{x}'$ がただ一つ存在する．これを $\boldsymbol{x}$ の**逆ベクトル**と言い，$-\boldsymbol{x}$ で表わす.

(II)　V の任意の元と任意の複素数 a に対し，$\boldsymbol{x}$ の a 倍と呼ばれるもう一つの V の元（これを $a\boldsymbol{x}$ で表わす）が定まり，つぎの法則が成立つ：

(5)　$(a+b)\boldsymbol{x} = a\boldsymbol{x}+b\boldsymbol{x}$,

(6)　$a(\boldsymbol{x}+\boldsymbol{y}) = a\boldsymbol{x}+a\boldsymbol{y}$,

(7)　$(ab)\boldsymbol{x} = a(b\boldsymbol{x})$,

(8)　$1\boldsymbol{x} = \boldsymbol{x}$.

以上の二条件 (I), (II) を 複素線型空間の公理と言う．混同のおそれのないときは，V の元を単に**ベクトル**と言う．ベクトルと対照的に，複素数を**スカラー**と言うこともある.

上の定義に現われる言葉「複素」をすべて「実」で置換えれば，**実線型空間**が定義される．われわれはこの両方を必要とする．線型空間の一般論は，「複素」「実」の両方について，まったく並行に進む．一々ものごとを二重に表現するのを避けるため，複素数全体の集合 $\boldsymbol{C}$ および実数全体の集合 $\boldsymbol{R}$ の両方を，統一的に記号 $\boldsymbol{K}$ で表わすことにする．そして，複素線型空間，実線型空間のかわりに，$\boldsymbol{K}$ 上の線型空間という言葉を使う．もちろん，一続きの議論

metric space 距離空間 きょりくうかん

set of $d(x \cdot y)$ is a metric space.

norm space ノルム空間 ノルムくうかん

set of $\|x\|$ is a norm space.

if the elements of metric space satisfies linear structure.

this metric space is a linear metric space.

if the elements of norm space satisfies linear structure

this norm space is a linear norm space.

vector space: distance, angle, (norm space is wichout angle definition)

$\langle \cdot , \cdot \rangle \quad V \times V \to F$ note: $F = R$ or $F = C$.

$x \cdot y \cdot z \in V, \quad a \in F$

① $\langle x \cdot y \rangle = \overline{\langle y \cdot x \rangle}$ 共軛对称性

if $F = R$ $\langle x \cdot y \rangle = \langle y \cdot x \rangle$.

if $F = C$ $\langle x \cdot y \rangle =$ complex conjugate. 複素共役.

② Linearity in the first argument. 第一引数に対する線型性

$\langle ax \cdot y \rangle = a \langle x \cdot y \rangle$

$\langle x + y \cdot z \rangle = \langle x \cdot z \rangle + \langle y \cdot z \rangle$

③ positive-definiteness 正定値性

$\langle x \cdot x \rangle \geq 0$

$\langle x \cdot x \rangle = 0 \Rightarrow x = 0$.

now, we have vector space, note that vector space is a linear space. i.e. the elements of vector space statisfy 8 equations of defination of linear space.

In vector space, if the dimension of element is limited, then this vector space is a Euclidean space.

位相 $\xleftarrow{\text{汎化 generalize}}$ 距離 $\xrightarrow{\text{specify}}$ ノルム $\xrightarrow{\text{specify}}$ 内積

位相空間 ← 距離空間 ← ノルム空間 ← 内積空間

位相幾何学 { 代数的, 幾何的, 微分

ノルム空間 ↓ 1922, banach + complete

banach space (norm space + complete).

内積空間 {
- n = limited: Euclidean space
- n = ∞ + complete 完備化: Hilbert space (1904-1910, Hilbert).

the study on function in these two spaces, we call it as functional analysis

関数解析学. { linear, non-linear.

抽象代数学.

Functional analysis, topology, abstract algebra are three frontier research area in math, now.

Reproducing kernel Hilbert Space:

① Given a kernel function, how to construct a feature space, i.e. a feature map, such that the kernel function value between two samples in the original space is the inner product between the corresponding samples in the feature space.

② Given a feature map, how to construct the corresponding kernel function.

kernel function $\quad\quad\quad\quad$ feature map.

result $= k(x, z) \longleftrightarrow \varphi(x)$

result = result of inner product of two training samples in feature space.

E.g.

$$\varphi : R^2 \to R^3$$

$$(x_1, x_2) \to (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2).$$

$$\langle \varphi(x_1, x_2) . \varphi(x_1' . x_2') \rangle$$

$$= \langle (z_1 . z_2 . z_3) , (z_1' , z_2' , z^{3'}) \rangle$$

$$= \langle (x_1^2 , \sqrt{2}x_1x_2 , x_2^2) \ (x_1'^2 , \sqrt{2}x_1'x_2' , x_2^{2'}) \rangle$$

$$= x_1^2 x_1'^2 + 2x_1x_1'x_2x_2' + x_2^2 x_2'^2$$

$$= (x_1x_1' + x_2x_2')$$

$$= (\langle x . x' \rangle)^2$$

$$= k(x . x') \leftarrow \text{kernel function.}$$

Gram Matrix or kernel Matrix.

Given a funtion $k: R^d \to R$ and the samples $x_1, \ldots x_m \in R^d$, then an $m \times m$ matrix K with elements

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \cdots & k(x_m, x_m) \end{bmatrix}$$

is called the Gram matrix (or kernel matrix) of k with respect to $x_1, \ldots x_m$.

E.g.

$x, z \in R$, $k(x, z) = \exp\left(-\frac{(x-z)^2}{2}\right)$, $x_1 = 2$, $x_2 = 0$.

$$k = \begin{bmatrix} k(x_1 x_1) & k(x_1 x_2) \\ k(x_2 x_1) & k(x_2 x_2) \end{bmatrix} = \begin{bmatrix} \exp(0) & \exp(-1) \\ \exp(-1). & \exp(0) \end{bmatrix} = \begin{bmatrix} 1 & 0.1353 \\ 0.1353 & 1 \end{bmatrix}$$

E.g.

$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in R^2$, $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in R^2$, $k(x. z) = (x_1 z_1 + x_2 z_2)^2.$
$= (x^T z)^2$

$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$k = \begin{bmatrix} k(x_1 x_1) & k(x_1 x_2) \\ k(x_2 x_1) & k(x_2 x_2) \end{bmatrix} = \begin{bmatrix} (0\times0+0\times0)^2 & (0\times1+0\times1)^2 \\ (1\times0+1\times0)^2 & (1\times1+1\times1)^2 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

Positive Semi-definite Matrix.

definition: a real symmetric $m \times m$ matrix k satisfying

$$a^T k a \geq 0$$

for all $a \in R^m$ is called a positive semi-definite matrix.

E.g. $k = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ is positive semi-definite

$k = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ is positive semi-definite.

how to proof?

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

$$a^T k a = \cancel{\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}^T \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} [a_1, a_2] =}$$

$$= [a_1\ a_2] \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = [2a_1, 0] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 2a_1^2 \geq 0.$$

$\forall\ a_1, a_2 \in R.\ \exists\ 2a_1^2 \geq 0.$ i.e. $a^T k a \geq 0$

$\therefore\ k = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ is positive semi-definite.

Theorems

1. A real symmetric matrix is diagonalizable.

2. A real symmetric matrix is positive semi-definite if and only if all its eigenvalues are non-negative.

E.g. $k = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ is positive semi-definite.

$k = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ is positive.

how to proof it using this Theorems.

$\det(k - \lambda I) = 0.$

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & -\lambda \end{vmatrix} = (2-\lambda)(-\lambda) = 0. \Rightarrow \underset{\geq 0}{\lambda = 0} \text{ or } \underset{\geq 0.}{\lambda = 2.}$$

$\therefore k = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ is positive semi-definite.

Positive Semi-Definite kernel.

If a function $k: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_m) \\ \vdots & \ddots & \vdots \\ k(x_m, x_1) & \cdots & k(x_m, x_m) \end{bmatrix}$$

is positive semi-definite matrix for any set $\{x_1 \cdots x_m\} \subset \mathbb{R}^d$, the k is called a positive semi-definite kernel function.

E.g. proof $k(x, z) = x^T z$ is positive semi-definite.

given any $x_1 \cdots x_m$.

$$K = \begin{bmatrix} k(x_1, x_1) \cdots k(x_1, x_m) \\ k(x_m, x_1) \cdots k(x_m, x_m) \end{bmatrix} = \begin{bmatrix} x_1^T x_1 & \cdots & x_1^T x_m \\ \vdots & \ddots & \vdots \\ x_m^T x_1 & \cdots & x_m^T x_m \end{bmatrix}$$

$$= \begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix} [x_1 \cdots x_m]$$

$$a^T K a = [a_1 \cdots a_m] \begin{bmatrix} x_1^T \\ \vdots \\ x_m^T \end{bmatrix} [x_1 \cdots x_m] \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$$

$$= (a_1 x_1^T + \cdots + a_m x_m^T) \cdot (a_1 x_1 + \cdots + a_m x_m)$$

$$= (a_1 x_1 + \cdots + a_m x_m)^T \cdot (a_1 x_1 + \cdots + a_m x_m)$$

$$= \| a_1 x_1 + \cdots + a_m x_m \| \geq 0.$$

solve it!

Proposition: Cauchy-Schwarz Inequality for kernel.

if $k: R^d \times R^d \to R$ is a kernel and $x, z \in R^d$, then

$$|k(x,z)|^2 \le k(x,x) \cdot k(z,z)$$

inner product $\le$ distance · distance.

proof it:

$$K = \begin{bmatrix} k(x,x) & k(x,z) \\ k(z,x) & k(z,z) \end{bmatrix} = \begin{bmatrix} k(x,x) & k(x,z) \\ k(x,z) & k(z,z) \end{bmatrix}$$

$\because$ k is a kernel, i.e. $\det(K) \geqslant 0$.

$$\det(K) = \begin{vmatrix} k(x,x) & k(x,z) \\ k(x,z) & k(z,z) \end{vmatrix} = k(x,x) \cdot k(z,z) - k(x,z)^2 \geqslant 0.$$

i.e. $k(x,x) \cdot k(z,z) \geqslant |k(x,z)|^2$

The Reproducing kernel Map.

Definition : The space of Function.

R^{R^d} is the set of all functions from R^d to R.

That is $R^{R^d} = \{f: R^d \to R\}$.

E.g. Given $k(x.z) = \exp(-\frac{(x-z)^2}{2})$

$\varphi_0(x) = k(x.0) = \exp(-\frac{x^2}{2}) \in R^R$ $\quad R \to R$

$\varphi_1(x) = k(x.1) = \exp(-\frac{(x-1)^2}{2}) \in R^R$ $\quad R \to R$

Given $k(x\,z) = (x_1 z_1 + x_2 z_2)^2$

$\varphi_{[0\,0]^T}(x) = k(x.[0.0]^T) = 0 \in R^{R^d}.\ d=2.\ R^{R^2}$ $\quad R^2 \to R$

$\varphi_{[1.1]^T}(x) = k(x,[1.1]^T) = (x_1+x_2)^2 \in R^{R^2}$ $\quad R^2 \to R.$

$\varphi_{[1.-1]^T}(x) = ?$

$\varphi_{[1.15]^T}(x) = ?$

The Reproducing Kernel Map.

Suppose $k: R^d \times R^d \to R$ is a kernel.

The reproducing kernel map is a map.

$$\underset{\text{vector}}{\Phi : R^d} \to \underset{\text{function}}{R^{R^d}} \text{ such that } \Phi(z) = \varphi_z = k(\cdot, z)$$

that is $\Phi(z)(x) = \varphi_z(x) = k(x.z)$.

E.g. Given $k(x.z) = \exp\left(-\frac{(x-z)^2}{2}\right)$

$\varphi_0(x) = k(x.0) = \exp\left(-\frac{x^2}{2}\right) \in R^R$

$\Phi(0) = \varphi_0$

$\Phi(0)(x) = \varphi_0(x) = \exp\left(-\frac{x^2}{2}\right)$

$\Phi(0)(0) = \varphi_0(0) = \exp\left(-\frac{0^2}{2}\right) = 1 = k(0.0)$.

$\varphi_1(x) = ?$

$\Phi(1) = ?$

$\Phi(1)(x) = ?$

$\Phi(1)(1) = ?$

$\varphi_1(x) = k(x.1) = \exp\left(-\frac{(x-1)^2}{2}\right) \in R^R$

$\Phi(1) = \varphi_1$

$\Phi(1)(x) = \varphi_1(x) = \exp\left(-\frac{(x-1)^2}{2}\right) = k(x.1)$

$\Phi(1)(1) = \varphi_1(1) = \exp\left(-\frac{0}{2}\right) = 1 = k(1.1)$.

Vector Space Definition.

A vector space is a set V over R together with two binary operators, $+$ and $\cdot$, that satisfy 8 axioms listed below:

$$u \cdot v \cdot w \in V$$

1. Associativity of addition: $u+(v+w)=(u+v)+w$

2. Commutatitivity of addtion: $v+w=w+v$.

3. Identity element of addition:
   $\exists\, 0 \in V$, the zero vector, such that $v+0=v$ for all $v \in V$

4. Inverse elements of addition:
   $\forall\, v \in V$. $\exists\, w \in V$. the additive inverse of v. such that $v+w=0$.

$$a \cdot b \cdot 1 \in R,\ v \cdot w \in V.$$

5. Distributivity of scalor multiplication with respect to vector addtion: $a(v+w)=av+aw$.

6. Distributivity of scalor multiplication with respect to R addtion: $(a+b)v=av+bv$.

7. Compatibility of scalor muliplication with R muliplicatio

$$a(bv)=(ab)\cdot v$$

8 Identiey element of scalor multiplication:

$$1v=v.$$

Define a Vector Space.

Suppose $k: R^d \times R^d \to R$ is a kernel and Φ is the corresponding reproducing kernel map. Let

$$\mathcal{F}_k = \left\{ f = \sum_{i=1}^{m} \alpha_i \Phi(x_i) \;\middle|\; m \in N,\ x_1 \cdots x_m \in R^d,\ \alpha_1 \cdots \alpha_m \in R \right\}$$

The $\mathcal{F}_k$ is a vector space over R.

E.g. <u>giving a kernel function to create a vector space:</u>

Given $k(x, z) = \exp\left(-\frac{(x-z)^2}{2}\right)$

$\Phi(0)(x) = \varphi_0(x) = k(x, 0) = \exp\left(-\frac{x^2}{2}\right) \in R^R$

$\Phi(1)(x) = \varphi_1(x) = k(x, 1) = \exp\left(-\frac{(x-1)^2}{2}\right) \in R^R$

$f(x) = 1 \cdot \Phi(0)(x) + 1 \cdot \Phi(1)(x) = \exp\left(-\frac{x^2}{2}\right) + \exp\left(-\frac{(x-1)^2}{2}\right) \in \mathcal{F}_k$.

now we got the definition of vector space.

Then, let create inner product in this vector space.

$$\mathcal{F} = \left\{ f = \sum_{i=1}^{m} \alpha_i \Phi(x_i) \right\} \quad : \text{vector space.}$$

Definition of inner product:

An inner product space over R is a vector space V over R together with an inner product. i.e. with a map

$$\langle \cdot , \cdot \rangle : V \times V \to R$$

that satisfies the following three axioms for all vectors $x, y, z \in V$ and all scalors $a \in R$:

1. Symmetry $\langle x . y \rangle = \langle y . x \rangle$

2. Linearity in the first argument:

$$\langle ax . y \rangle = a \langle x . y \rangle$$

$$\langle x + y . z \rangle = \langle x . z \rangle + \langle y . z \rangle.$$

3. Positive - definiteness:

$$\langle x . x \rangle > 0 \text{ if } x \neq 0$$

$$\langle x . x \rangle = 0 \Leftrightarrow x = 0$$

Then let's define inner product in vector space.

Given $\mathcal{F}_k$ $\nearrow f = \sum_{i=1}^{m} \alpha_i \Phi(x_i)$, $\searrow g = \sum_{j=1}^{m'} \beta_j \Phi(x_j)$

Recall that $f = \sum_{i=1}^{m} \alpha_i k(\cdot, x_i)$ and $g = \sum_{j=1}^{m'} \beta_j k(\cdot, x_j')$

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

$$= \alpha_1 \beta_1 k(x_1 x_1') + \alpha_1 \beta_2 k(x_1, x_2') + \cdots + \alpha_1 \beta_{m'} k(x_1, x_{m'}')$$
$$+ \alpha_2 \beta_1 k(x_2 x_1') + \alpha_2 \beta_2 k(x_2, x_2') + \cdots + \alpha_2 \beta_{m'} k(x_2, x_{m'}')$$
$$\vdots$$
$$+ \alpha_m \beta_1 k(x_m x_1') + \alpha_m \beta_2 k(x_m, x_2') + \cdots + \alpha_m \beta_{m'} k(x_m x_{m'}')$$

then $\langle f \cdot g \rangle$ is an inner product over $\mathcal{F}_k$.

E.g.: Given $k(x \cdot z) = \exp\left(-\frac{(x-z)^2}{2}\right)$

$$f(x) = k(x \cdot 0) - k(x \cdot 1) = \exp\left(-\frac{x^2}{2}\right) - \exp\left(-\frac{(x-1)^2}{2}\right)$$
$$g(x) = -k(x, 0) + k(x \cdot 1) = -\exp\left(-\frac{x^2}{2}\right) + \exp\left(-\frac{(x-1)^2}{2}\right)$$

$$\langle f \cdot g \rangle = 1 \cdot (-1) \cdot k(0 \cdot 0) + 1 \cdot 1 \cdot k(0 \cdot 1)$$
$$+ (-1) \cdot (-1) k(1 \cdot 0) + (-1) \cdot (1) k(1 \cdot 1).$$

Proposition

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

$= \sum_{i=1}^{m} \alpha_i \sum_{j=1}^{m'} \beta_j k(x_i \cdot x_j')$

$= \sum_{i=1}^{m} \alpha_i g(x_i)$ $\qquad g = \sum_{j=1}^{m'} \beta_j k(\cdot \cdot x_j')$.

② $\langle f \cdot g \rangle = \sum_{i=1}^{m} \sum_{j=1}^{m'} \alpha_i \beta_j k(x_i \cdot x_j')$

$= \sum_{j=1}^{m'} \beta_j \sum_{i=1}^{m} \alpha_i k(x_j' \cdot x_i)$

$= \sum_{j=1}^{m'} \beta_j f(x_j')$ $\qquad f = \sum_{i=1}^{m} \alpha_i k(\cdot x_i)$.

Homework: to proof $\langle f \cdot g \rangle$ is an inner product.

hint: satisfy three axioms of definition of inner product.

$\langle \underset{g}{\underset{\|}{k(\cdot x)}} \cdot f \rangle = f(x) = \sum_{i=1}^{m} \alpha_i k(x_i \cdot x)$.

$\langle k(\cdot x) k(\cdot z) \rangle = k(z \cdot x) = k(x \cdot z)$.

Let's explain its meaning

As a summary:

$$k(x \cdot z) = \langle k(\cdot x) \cdot k(\cdot z) \rangle$$
$$= \langle \Phi(x) \, \Phi(z) \rangle$$

since $\Phi(x) = \varphi_x = k(\cdot x)$ and $\Phi(z) = \varphi_z = k(\cdot z)$.

From now on, we show that any positive definite kernel can be through of as a dot product in another space.

① given a kernel funt k.

$\downarrow$ concept of function space.

② reproducting feature map. $\Phi(\cdot z) = k(\cdot z)$

$\downarrow$ vector space.

③ $\mathcal{F}_k$. (vetor space).

$\downarrow$

④ inner product over $\mathcal{F}_k$.

The result value of inner product over $\mathcal{F}_k$ can be expressed by the result of kernel function !!!