Chapter 12: Reproducing Kernel Hilbert Spaces:

(Definition 12.1) Hilbert Space Operations

$\langle f, g \rangle = \langle g, v \rangle$ for all $f, g \in \mathbb{V}$

$\langle f, f \rangle \geq 0$ for all $f \in \mathbb{V}$

$\langle f + \alpha g, h \rangle = \langle f, h \rangle + \alpha \langle g, h \rangle$ for all $f, g, h \in \mathbb{V}$ and $\alpha \in \mathbb{R}$

(Definition 12.1)

A Hibert space is an inner product space $(\langle \circ, \circ \rangle_{\mathbb{H}}, \mathbb{H})$ in which every Cauchy Sequence $(f_n)_{n=0}^{\infty}$ converges to some element $f^* \in \mathbb{H}$.

12.1 $\text{null}(L) = \{ f \in \mathbb{H} \mid L(f) = 0 \}$

$\text{null}(L^c) = \{ f^* \in \mathbb{H} \mid L^c(f) \neq 0 \}$

12.2. $\mathbb{G}$ is a closed convex set of $\mathbb{H}$

$f \in \mathbb{H}$ and $\hat{g} \in \mathbb{G}$ such that $\|\hat{g} - f\|_{\mathbb{H}} = \underbrace{\inf_{g \in \mathbb{G}} \|\hat{g} - f\|_{\mathbb{H}}}_{p^*}$

"$\hat{g}$" is a projection of f onto $\mathbb{G}$.

a) $\|g_n - f\|_{\mathbb{H}} \to p^*$ when $(g_n)_{n=1}^{\infty}$ is in $\mathbb{G}$

This is a Cauchy sequence:

$$\|\hat{g} - f\|^2 = \|f - \hat{g}\|^2$$

$$= \|f - g_n\|^2 + \|f - g_m\|^2 + 2\langle f - g_n, f - g_m \rangle$$

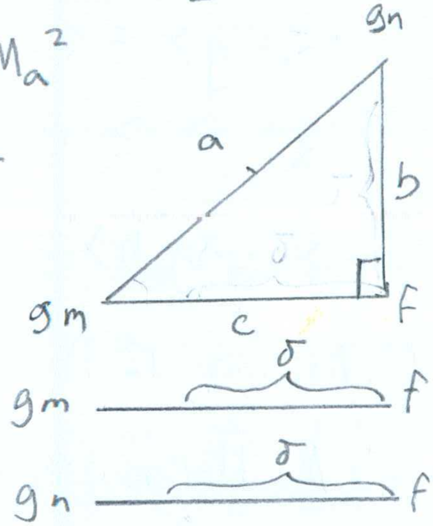
$$4\|f - \tfrac{g_n + g_m}{2}\|^2 = \|f - g_n\|^2 + \|f - g_m\|^2 - 2\langle f - g_n, f - g_m \rangle$$

Equation addition:

$$\|f-\hat{g}\|^2 + 4\|f - \tfrac{g_n+g_m}{2}\|^2 = 2\|f-g_m\|^2 + 2\|f-g_n\|^2$$

$$\|f-\hat{g}\|^2 = 2\|f-g_n\|^2 + 2\|f-g_m\|^2 - 4\|f - \tfrac{g_n+g_m}{2}\|^2$$

$$a^2 = 2b^2 + 2c^2 - 4M_a^2$$



$$\|f-\hat{g}\|^2 = 2(\delta^2+\tfrac{1}{n}) + 2(\delta^2+\tfrac{1}{m}) - 4\delta^2$$

$$= 2\left(\tfrac{1}{n}+\tfrac{1}{m}\right)$$

$$= \text{Cauchy Sequence}$$

"As a signal decays, the projected signal onto an axis decays to zero too."

b) The lower limit in a projected vector:

$\hat{g}$ is existant;

$$S := \{\|g\| ; g \in G\}$$

$$p^* = \inf_{g \in G} \|g\| = \inf \|S\|$$

When $n \geq 1$, $p^* \leq \inf \|S\| \leq p^* + \frac{1}{n}$

$$\lim_{n\to\infty} p^* \leq \lim_{n\to\infty} \inf \|S\| \leq \lim_{n\to\infty} p^* + \frac{1}{n}$$

At the limit, $p^* = \inf_{g\in G} \|S\|$

$$= \inf_{g \in G} \|g\|$$

c.) The projection is unique:

For m_1 and m_2 points, that are unique,

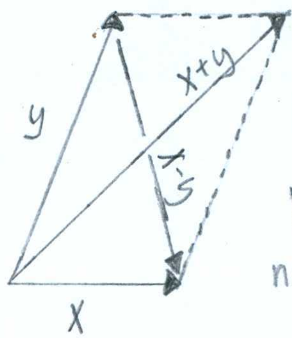
$$\|g_n - g_m\|^2 = 2\|g_n - x\|^2 + 2\|g_m - x\|^2 - 4\left\|\frac{g_n+g_m}{2} - x\right\|^2$$

$$\leq 2\delta^2 + 2\delta^2 - 4\delta^2$$

$$\leq 0$$

$g_n = g_m$ which show uniqueness.

d) The same holds true for an arbitrary convex set:



$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

$$\|g_n - g_m\|^2 \leq 2\|g_n\|^2 + 2\|g_m\|^2 - 4\|g_n - g_m\|^2$$

$$\lim_{n\to\infty} \|g_n\| = p^*$$

12.3 $\mathbb{H}$: Hilbert Space

G: Closed Linear Subspace in $\mathbb{H}$

Proof $f \in \mathbb{H}$, such that $g + g^\perp \in \mathbb{H}$:

$$\|f - \hat{g}\|^2 = \|f-g\|^2 + \|f-g^\perp\|^2 + 2\langle f-g, f-g^\perp\rangle$$

$$4\left\|f - \frac{g+g^\perp}{2}\right\|^2 = \|f-g\|^2 + \|f-g^\perp\|^2 - 2\langle f-g, f-g^\perp\rangle$$

Equation addition:

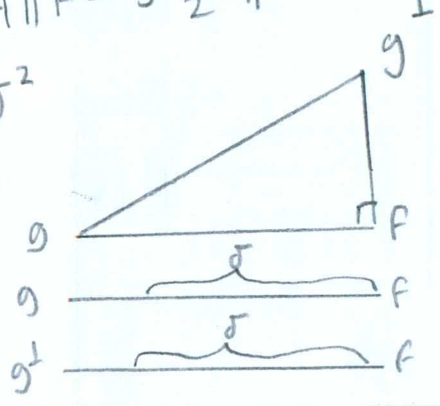
$$\|f-\hat{g}\|^2 + 4\left\|f - \frac{g+g^\perp}{2}\right\|^2 = 2\|f-g\|^2 + 2\|f-g^\perp\|^2$$

$$\|f-\hat{g}\|^2 = 2\|f-g\|^2 + 2\|f-g^\perp\|^2 - 4\left\|f - \frac{g+g^\perp}{2}\right\|^2$$

$$= 2\left(\delta^2 + \frac{1}{n}\right) + 2\left(\delta^2 + \frac{1}{m}\right) - 4\delta^2$$

$$= 2\left(\frac{1}{n} + \frac{1}{m}\right)$$

= Cauchy sequence in $\mathbb{H}$



12.4 (Kernel Function) $c^T K c = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k_{ij}$

(Reproducing Kernel Function) $c^T K(x_i, x_j) c = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j)$

A unique reproducing kernel has $K(X_i) = K(X_j) = 0$:

$$c^T K(X_i, X_j) c = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(X_i, X_j)$$

$$= \sum_{i=1}^{N} c_i \langle K(X_i), \sum_{j=1}^{N} c_i k(X_j) \rangle$$

$$= \langle \sum_{i=1}^{N} c_i k(X_i), \sum_{i=1}^{N} c_j K(X_j) \rangle$$

$$= 0$$

12.5.

a) $K : X \times X \to \mathbb{R}$

$$K(X, Z) = \begin{bmatrix} \langle XX \rangle & \langle XZ \rangle \\ \langle ZX \rangle & \langle ZZ \rangle \end{bmatrix}$$

$$= \langle XX \rangle \langle ZZ \rangle - \langle ZX \rangle \langle XZ \rangle$$

$$= \sqrt{\langle XX \rangle^2 \langle ZZ \rangle^2} - \langle ZX \rangle \langle XZ \rangle$$

$$= \sqrt{K(X,X) K(Z,Z)} - \langle ZX \rangle \langle XZ \rangle$$

$$K(X,Z) \leq \sqrt{K(X,X) K(Z,Z)}$$

b) Proof with a Cauchy-Shwarz inequality:

$$\|K(x,x) + K(y,y)\|^2$$

$$= \|K(X,X)\| + 2\|K(x) \circ K(z)\| + \|K(z,z)\|$$

$$= (\|K(X,X)\| + \|K(y,y)\|)^2$$

$$\|K(X,X)+K(Z,Z)\|^2 = (\|K(X,X)\| + \|K(y,y)\|)^2$$

$$\|K(X,X)+K(Z,Z)\| = \|K(X,X)\| + \|K(Z,Z)\|$$

"The absolute distance in a sum is ~~the~~ the sum of each absolute component"

(Theorem 12.20) "Mercer's Theorem"

When conditions are met : 1) Positive Semidefinite

$$K \geq 0$$

2) Finite in Hilbert-Shmidt

$$\int_{X \times X} K^2(X,Z)\, d\mathbb{P}(X)\, d\mathbb{P}(Z) < \infty$$

3) Continuous

A sequence is also a sum of non-negative components. In many cases, a sequence is a sum of eigenvalues.

$$K(X,Z) = \sum_j \mu_j \, \phi_j(X) \, \phi_j(Z)$$

Kernel

Eigenvalue

Eigenvector

Eigenvector

12.6. Proof by Deductive Logic:

$$\text{If } \|K(X)\|^2 = \sum_{j=1}^{\infty} \mu_j K(X) K(X)$$

$$= K(X,X)$$

$$< \infty$$

Then $\langle K(X), k(z)\rangle = \sum_{i=1}^{\infty} \mu_i X_i z_i$

$= K(X,z)$

12.7. $m \geq 1$

$K_1(x,z) = (1+xz)^m \qquad K_2(x,z) = \sum_{l=0}^{m} \frac{x^l}{l!} \frac{z^l}{l!}$

Proof Positive semi-definite by Induction:

Base Case $(n=1)$: $K_1(x,z) = 1 + xz$

≥ 0 when $x,z \in \mathbb{N}$

Next Step $(n=m)$: $K_1(x,z) = (1+xz)^m$

$= 1 + m\,xz + \frac{m(m-1)x^2z^2}{2!} + \cdots$

≥ 0

Inductive Step $(n=m+1)$: $K_1(x,z) = (1+xz)^{m+1}$

$= 1 + (m+1)xz + \frac{(m+1)m\,x^2z^2}{2!} + \cdots$

≥ 0

Proof Positive semi-definite by Induction:

Base Case $(n=1)$: $K_2(x,z) = \sum_{l=0}^{1} \frac{x^l}{l!} \frac{z^l}{l!}$

$= 1 + xz$

≥ 0 when $x,z \in \mathbb{N}$

Next Step $(n=m)$: $K_2(x,z) = \sum_{l=0}^{m} \frac{x^l}{l!} \frac{z^l}{l!}$

$= 1 + \sum_{l=1}^{m} \frac{x^l}{l!} \frac{z^l}{l!}$

$$\geq 0$$

Inductive Step$(n=m+1)$: $K_2(X,Z) = \sum_{l=0}^{m+1} \frac{X^l}{l!}\frac{Z^l}{l!}$

$$= 1 + \sum_{l=1}^{m+1} \frac{X^l}{l!}\frac{Z^l}{l!}$$

$$\geq 0$$

(Reproducing Kernel Hilbert Space Polynomial)

$$K(X,Z) = X^d Z^d + \binom{d}{1} c X^{d-1} Z^{d-1} + \binom{d}{2} c^2 X^{d-2} Z^{d-2} + \cdots + c^d$$

$$K_1(X,Z) = (1+XZ)^m$$

$$= (1+XZ)^{m-1} + (1+XZ)^{m-2} XZ + (1+XZ)^{m-3} X^2 Z^2 + \cdots + X^{m-1} Z^{m-1}$$

$$= X^{m-1} Z^{m-1} + (1+XZ) X^{m-2} Z^{m-2} + (1+XZ)^2 X^{m-3} Z^{m-3} + \cdots + (1+XZ)^{m-1}$$

where $d = m-1$; $c = (1+XZ)$

$$K_2(X,Z) = \sum_{l=0}^{m} \frac{X^l}{l!}\frac{Z^l}{l!}$$

$$= 1 + XZ + \frac{X^2}{2!}\frac{Z^2}{2!} + \frac{X^3}{3!}\frac{Z^3}{3!} + \cdots + \frac{X^m}{m!}\frac{Z^m}{m!}$$

$$= XZ + \frac{X^m}{m!}\frac{Z^m}{m!} + \frac{X^{m-1}}{(m-1)!}\frac{Z^{m-1}}{(m-1)!} + \frac{X^{m-2}}{(m-2)!}\frac{Z^{m-2}}{(m-2)!} + \cdots + 1$$

where $d = m+1$, $m = 0$, $c = \binom{d}{n}$ for the nth-term.

b) Why are these unique?

An arbitrary variable, such as, a new basis set reproduces the same Reproducing Kernel Hibert Space.

$$K(X\cdot) = (1+X(\circ))^m = (1+(\circ)\cdot X)^m = K(\circ, X)$$

12.8 True or False:

a) When K_1 and $K_2 \geq 0$, the bivariate function $K(X,z) = \min_{j=1,2} K_j(X,z)$ is a positive semidefinite kernel.

True

<u>Proof by Exhaustion:</u>

Case #1 : $K_1 > K_2$;

$$K(X,z) = \min_{j=1,2} K_j = K_2 \geq 0$$

Case #2 : $K_1 < K_2$;

$$K(X,z) = \min_{j=1,2} K_j = K_1 \geq 0$$

b) If $f: X \to \mathbb{H}$ where $X \in \mathbb{H}$ defines a positive semidefinite kernel on $X \times X$, then also $K(x,z)$.

True

$$K(x,z) = \frac{\langle f(x), f(z) \rangle_{\mathbb{H}}}{\|f(x)\|_{\mathbb{H}} \, \|f(z)\|_{\mathbb{H}}}$$

← positive in Hilbert space

← positive range in a square root

(Schur Product Theorem)

An element-wise product (Hadamard product) of two positive definite matrices is also positive.

12.9. $K: X \times X$

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

$$\begin{aligned} \bar{K}(x,z) &= f(x)\, K(x,z) \cdot f(z) \\ &= f(x)\, f(x) \cdot f(z) \cdot f(z) \\ &= f(x)^2 \cdot f(z)^2 \\ &\geq 0 \end{aligned}$$

(Power Set) A set of all subsets, including the empty set itself.

$$S=\{1,2\}, \quad P(S)=\{\{\emptyset\},\{1\},\{2\},\{1,2\}\}$$

12.10. $K: P(S)\times P(S) \to \mathbb{R}$ is $K(A,B)=2^{A\cap B}$

The smallest cardinality in a set is one $(n=1)$, so

$K(A,B)=2^{A\cap B}=2^1 \geq 0$ at a minimum size.

Other definitions define $K(A,B)=|2^{A\cap B}|$

$$=2^{|A\cap B|}$$

$$=2^n$$

Where $n=\#$ elements $(n\geq 1)$.

12.11. $K(X,Z)=(1+\langle X,Z\rangle)^m$

$$=\sum_{m=0}^{d+m}\binom{d+m}{m}\langle X,Z\rangle^m$$

$$=\sum_{m=0}^{d+m}\phi(X_i)\phi(Z)i$$

$=\Phi(X)\cdot\Phi(Z)$ where $\Phi(X)=[1,\sqrt{d+1}\cdot X,\sqrt{(d+1)(d+2)/2}\,X^2,\dots]$

$\Phi(Z)=[1,\sqrt{(d+1)}\cdot Z,\sqrt{(d+1)(d+2)/2}\,Z^2,\dots]$

12.12 $K(A,B)=P[A\cap B]-P[A]P[B]$

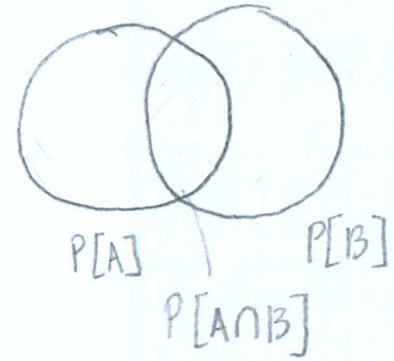
$$=\mathbb{E}[(1-\mathbb{E}[A])(1-\mathbb{E}[B])]$$

$$=\mathbb{E}[1-\mathbb{E}[A]-\mathbb{E}[B]-\mathbb{E}[A]\mathbb{E}[B]]$$

≥ 0, when $P[A\cup B]=1$.

<u>Probability Law:</u>

$P[A\cup B]=P[A]+P[B]-P[A\cap B]$



12.13. $K(A,B) = \sum_{\substack{x\in A \\ z\in B}} K(x,z) = \sum_{x\in A}\sum_{z\in B} \phi(x)\cdot\phi(z)$

$$= \begin{bmatrix} \phi(x_0)\phi(z_0) & \cdots & \phi(x_0)\phi(z_n) \\ \vdots & \ddots & \vdots \\ \phi(x_n)\phi(z_0) & \cdots & \phi(x_n)\phi(z_n) \end{bmatrix}$$

$$= \{\{\emptyset\}, \{\phi(x_0)\phi(z_0)\}, \ldots, \{\phi(x_0)\phi(z_0), \ldots, \phi(x_n)\phi(z_n)\}\}$$

"Power Set" when $\phi(x) \in \mathbb{H}$
$\phi(z) \in \mathbb{H}$

12.14. $K : X \times X \to \mathbb{R}$

$K(x,z) \le b^2$ for all $x, z \in X$

$\|f(x)\|_\infty = \|f\| \cdot \sqrt{K(x,z)}$

$\le \|f\| \cdot b$

$\le b$ when in a unit ball, $\|f\| = 1$

(Proposition 12.27)

When $\mathbb{H}_1$ and $\mathbb{H}_2$ are both reproducible kernels in Hilbert space as by K_1 and K_2, then

$\mathbb{H} = \mathbb{H}_1 + \mathbb{H}_2$

12.15. (Equation 12.20) $K(x,z) = \sum_{\ell=0}^{\alpha-1} \frac{x^\ell}{\ell!}\frac{z^\ell}{\ell!} + \int_0^1 \frac{(x-y)^{\alpha-1}}{(\alpha-1)!}\frac{(z-y)^{\alpha-1}}{(\alpha-1)!}\,dy$

$$= \sum_{\ell=0}^{\alpha-1} \left(\frac{x^\ell}{\ell!}\right)^2 + \int_0^1 \left(\frac{(x-y)^{(\alpha-1)}}{(\alpha-1)!}\right)^2 dy$$

when $x = z$

$$= \sum_{\ell=0}^{\alpha-1} (f^{(\ell)}(0))^2 + \int_0^1 f^{(\ell)}(x)\,dx$$

$$= \|f\|_H^2 \quad \text{(Equation 12.21)}$$

12.16. Γ and Σ are each $n \times n$ and symmetric.

a) Proof by Example:

$$\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\Gamma \odot \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \odot \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\geq 0$$

b) $K(x,z) = K_1(x,z) \cdot K_2(x,z)$

$$= \sum_{i=1}^{n} xz \sum_{j=1}^{n} xz$$

$$= (1+0+0+1)(3+1+1+1)$$

$$\geq 0$$

12.17.
$$\sup_{\|f\|_\infty \leq 1} \left| \int f(d\mathbb{P} - d\mathbb{Q}) \right| = \sup_{f \in [-1,1]} \left| \int f(d\mathbb{P} - d\mathbb{Q}) \right|$$

$$\leq \left| \int d\mathbb{P} - d\mathbb{Q} \right|$$

$$\leq 2 \cdot (\mathbb{P}(A) - \mathbb{Q}(A))$$

12.18
$$\sup_{\|f\|_H \leq 1} \left| \mathbb{E}_P[f(x)] - \mathbb{E}_Q[f(z)] \right|^2 = \mathbb{E}\left[f(x)^2 + f(z)^2 - 2f(x)f(z) \right]$$

$$= \mathbb{E}[K(X,X) + K(Z,Z) - 2K(X,Z)]$$

12.19 $K(X,Z) = e^{\frac{-\|X-Z\|_2^2}{2\sigma^2}}$

a) <u>Proof by Exhaustion</u>

Even Polynomial: $p(\bar{K}(X,Z)) = \bar{K}(X,Z)^2 + \bar{K}(X,Z) + c$

$$= e^{\frac{-\|X-Z\|^2}{\sigma^2}} + e^{\frac{-\|X-Z\|^2}{2\sigma^2}} + c$$

$$\geq 0$$

Odd Polynomial: $p(\bar{K}(X,Z)) = \bar{K}(X,Z) + c$

$$= e^{\frac{-\|X-Z\|^2}{2\sigma^2}} + c$$

$$\geq 0 \quad , \text{ when } c \geq 0$$

b) $K_1(X,Z) = e^{\langle X,Z\rangle/\sigma^2}$

$$= 1 + \frac{\langle X,Z\rangle}{\sigma^2} + \frac{1}{2}\left(\frac{\langle X,Z\rangle}{\sigma^2}\right)^2 + \frac{1}{3!}\left(\frac{\langle X,Z\rangle}{\sigma^2}\right)^3 + \cdots$$

$$\geq 0$$

c) $K(X,X) = e^{\frac{-\langle X,X\rangle}{2\sigma^2}}$

$$= 1 - \frac{\langle X,X\rangle}{2\sigma^2} + \frac{1}{2}\left(\frac{\langle X,X\rangle}{2\sigma^2}\right)^2 + \left(\frac{\langle X,X\rangle}{2\sigma^2}\right)^3 + \cdots$$

12.20 $(X_i, y_i)_{i=1}^n$ pairs

$$\hat{f} = \underset{f \in \mathbb{H}}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^n \max\{0, 1 - y_i f(X)\} + \frac{1}{2}\lambda_n \|f\|_{\mathbb{H}}^2 \right\}$$

a) $\hat{f} = \underset{f \in \mathbb{H}}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^{n} \max\{0, 1 - y_i f(x)\} + \frac{1}{2} \lambda_n \|f\|_{\mathbb{H}}^2 \right\}$

$= \frac{1}{n} \sum_{i=1}^{n} \max\{0, -y_i\} + \lambda_n$

$= \frac{1}{2n} \sum_{i=1}^{n} \max\{\|f\|, (\|f\| - 1)\} y_i$

$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\alpha} K(\cdot, x_i)$ from Example 12.33

b) $\hat{\alpha} \in \underset{\alpha \in \mathbb{R}^n}{\arg\max} \left\{ \frac{1}{n} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \alpha^T \bar{K} \alpha \right\}$

Subject to $\in \left[0, \frac{1}{\lambda_n \sqrt{n}}\right]$ for all $i = 1, \ldots, n$

$\hat{\alpha} = \hat{f} = \underset{f \in \mathbb{H}}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^{n} \max\{0, 1 - y_i f(x)\} + \frac{1}{2} \lambda_n \|f\|_{\mathbb{H}}^2 \right\}$

$= \underset{\alpha \in \mathbb{R}^n}{\arg\max} \left\{ \frac{1}{n} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \alpha^T \bar{K} \alpha \right\}$