ACM118: Stochastic processes and regression

Instructor: Houman Owhadi

TA: Pau Batlle Franch

Use Piazza for all Q&A and discussions

ACM118 should not be your 1st course on prob. theory

Focus: Gaussian Processes & Kernel Methods

Objective: Learn to reason, model and organize computational knowledge with GPs & kernel methods

Grading: 4 pb sets, 25% each

Interactive course

- Questions are welcomed
- Pace can be adjusted
- Digressions are okay

Characteristic functions

X: rand. vect. on $\mathbb{R}^d$

ϕ_X: characteristic function of X

$$\mathbb{R}^d \longrightarrow \mathbb{C}$$
$$\lambda \longrightarrow \mathbb{E}[e^{i\lambda^T X}]$$

$\mathbb{E}[e^{\lambda X}]$

<u>Thm</u> $\phi_X(\lambda) = \phi_Y(\lambda) \quad \forall \lambda \in \mathbb{R}^d$

$$\Leftrightarrow \quad X \overset{\text{distribution}}{=} Y$$

$$\Leftrightarrow \quad \mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$$

$\forall f: \mathbb{R}^d \to \mathbb{R}$

bounded & measurable

continuous

- $X_1, \dots, X_m$ are indep. iif $\forall \lambda_1, \dots, \lambda_m$

$$\mathbb{E}[e^{i \sum_{j=1}^m \lambda_j^T X_j}] = \prod_{j=1}^m \mathbb{E}[e^{i \lambda_j^T X_j}]$$

- $(d=1)$ If $\mathbb{E}[|X|^k] < \infty$ then

$$\phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$$

If $\phi_X^{(k)}(0)$ well defined then

$$\mathbb{E}[|X|^k] < \infty$$

. $X_n \xrightarrow{dist} X \iff \mathbb{E}[f(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[f(X)]$

$$\forall f \in C_b(\mathbb{R}^d)$$

$$\iff \phi_{X_n}(\lambda) \xrightarrow[n\to\infty]{} \phi_X(\lambda)$$

$$\forall \lambda \in \mathbb{R}^d$$

$$\phi_X(\lambda) := \mathbb{E}[e^{i\lambda^T X}]$$

Thm 1. $\phi_X(0) = 1$ & $|\phi_X(\lambda)| \leq 1$

2. $\phi_X(\lambda)$ is uniformly continuous $\quad |\phi_X(\lambda)| \leq \mathbb{E}[|e^{i\lambda^T X}|] = 1$

3. $\forall \lambda_1, \ldots, \lambda_m \in \mathbb{R}^d, \ \forall z_1, \ldots, z_m \in \mathbb{C}$

$$\sum_{j,k=1}^{m} \phi_X(\lambda_j - \lambda_k) z_j \bar{z}_k \geq 0$$

$$\mathbb{E}\left[\left|\sum_j z_j e^{i\lambda_j^T X}\right|^2\right] \geq 0$$

4. Bochner's thm

$\phi : \mathbb{R}^d \to \mathbb{C}$ is the characteristic function of some rand. vect. X iif it satisfies 1, 2 & 3

<u>Thm</u> X_n: sequ. of $\mathbb{R}^d$ valued rand. var.

If $\lim_{n\to\infty} \phi_{X_n}(\lambda)$ exists $\forall \lambda \in \mathbb{R}^d$

and $\phi(\lambda) := \lim_{n\to\infty} \phi_{X_n}(\lambda)$ is continuous at $\lambda = 0$, then ϕ is the characteristic function of some r.v. X & $X_n \xrightarrow[n\to\infty]{\text{distrib}} X$

<u>Ex</u> $\mathbb{E}[e^{i\lambda X}]$

$\overset{\shortparallel}{\phi_X(\lambda)}$

$\phi_X(0) = 1$

$\lim_{\lambda\to 0} \mathbb{E}[e^{i\lambda X}] = \mathbb{E}[\lim_{\lambda\to 0} e^{i\lambda X}]$

$= 1$

$$\nabla_\lambda \phi_X(\lambda) \overset{?}{=} \mathbb{E}[e^{i\lambda X}\, iX]$$

$$\nabla_\lambda \phi_X(0) = i\,\mathbb{E}[X]$$

$$\phi_X'(\lambda) = \lim_{\varepsilon \to 0} \frac{\phi_X(\lambda+\varepsilon) - \phi_X(\lambda)}{\varepsilon}$$

$$= \lim_{\varepsilon \to 0} \mathbb{E}\left[\frac{e^{i(\lambda+\varepsilon)X} - e^{i\lambda X}}{\varepsilon}\right]$$

$$= \lim_{\varepsilon \to 0} \mathbb{E}\left[e^{i\lambda X}\, \frac{e^{i\varepsilon X} - 1}{\varepsilon}\right]$$

dominated convergence

Gaussian random variables

$d = 1$

Def X is a Gaussian rand. var. with mean 0 and variance 1

$$X \sim \mathcal{N}(0,1)$$

if $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

$\hookrightarrow$ pdf of X

Thm $\mathbb{E}[e^{i\lambda X}] = e^{-\lambda^2/2} \quad \forall \lambda \in \mathbb{R}$

$$f(z) = \mathbb{E}[e^{zX}] = e^{z^2/2} \quad \forall z \in \mathbb{R}$$

↳ is analytic on $\mathbb{C}$

$$\Rightarrow \mathbb{E}[e^{zX}] = e^{z^2/2} \quad \forall z \in \mathbb{C}$$

f analytic $\Leftrightarrow$ $f(z) = \sum_{n=0}^{\infty} a_n z^n \quad \forall z \in \mathbb{R}$

$$\sum_{n=0}^{\infty} |a_n| |z|^n < \infty \quad \forall z \in \mathbb{R}$$

$$\tilde{f}(z) = \sum_{n=0}^{\infty} a_n z^n \quad \forall z \in \mathbb{C}$$

$$\sum_{n=0}^{\infty} |a_n| |z|^n < \infty \quad \forall z \in \mathbb{C}$$

$$\tilde{f}(z) = f(z) \quad \forall z \in \mathbb{R}$$

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

$$f = e^{z^2/2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{2^n n!}$$

$$\mathbb{E}[e^{i\lambda X}] = 1 + i\lambda\,\mathbb{E}[X] + \dots + \frac{(i\lambda)^n}{n!}\mathbb{E}[X^n] + O(|\lambda|^{n+1})$$

$$= e^{-\lambda^2/2}$$

$X \sim \mathcal{N}(0,1)$

$$\mathbb{E}[X] = 0$$

$$\mathbb{E}[X^2] = 1$$

$$\mathbb{E}[X^{2n+1}] = 0$$

$$\mathbb{E}[X^{2n}] = \frac{(2n)!}{2^n n!}$$

<u>Def</u> Y is Gaussian with mean m and variance σ^2, $Y \sim \mathcal{N}(m, \sigma^2)$ iif one of the equivalent properties is satisfied

(i) $Y = \sigma X + m$ where $X \sim \mathcal{N}(0,1)$

(ii) $f_Y(y) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(y-m)^2}{2\sigma^2}\right)$

(iii) $\mathbb{E}[e^{i\lambda Y}] = e^{im\lambda - \sigma^2\lambda^2/2}$

$$\mathbb{E}[e^{i\lambda(\sigma X + m)}] \quad \text{(}= \text{above)}$$

$$e^{i\lambda m}\,\mathbb{E}[e^{i\lambda\sigma X}] = e^{i\lambda m}\, e^{-(\lambda\sigma)^2/2}$$

Ex $X \sim N(\mu_1, \sigma_1^2)$

$Y \sim N(\mu_2, \sigma_2^2)$

X & Y indep

$X + Y \overset{?}{\sim} N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$$\mathbb{E}[e^{i\lambda(X+Y)}] \underset{X,Y \text{ indep}}{=} \mathbb{E}[e^{i\lambda X}]\,\mathbb{E}[e^{i\lambda Y}]$$

$$= e^{i\lambda\mu_1 - \sigma_1^2\frac{\lambda^2}{2}}\, e^{i\lambda\mu_2 - \sigma_2^2\frac{\lambda^2}{2}}$$

$$= e^{i\lambda(\mu_1+\mu_2) - (\sigma_1^2 + \sigma_2^2)\lambda^2/2}$$

The Gaussian distribution is a stable distribution

<u>Def</u> distrib. (X) is stable iif $\forall a, b > 0$

X_1, X_2 indep copies of X

$(*)$ $aX_1 + bX_2 \sim cX + d$ for $c > 0$, d

The distr. of X is said to be strictly stable if $(*)$ holds with $d = 0$

<u>Rk</u> X is Cauchy $\quad$ pdf$(X) \sim \dfrac{1}{\alpha^2 + x^2}$

$\Updownarrow$ stable

X heavy-tailed $\Leftrightarrow$ $\lim\limits_{x \to \infty} e^{tx}\, \mathbb{P}[X \geq x] = \infty$ $\quad \forall t > 0$

Gaussian r.v. are the only stable r.v. that do not have a heavy tail

$X \sim \text{Cauchy}(\alpha)$ $\qquad \mathbb{E}[X]$ not defined

$X_1, \dots, X_n$ iid $\text{Cauchy}(\alpha)$

$$\frac{X_1 + \dots + X_n}{n} \sim \text{Cauchy}(\alpha)$$

Ex $X \sim \mathcal{N}(0,1)$

$Y \sim \mathcal{N}(0,1)$

Is $X+Y$ Gaussian ?

$X = -Y$

$X+Y = 0 \sim \mathcal{N}(0,0)$

$$Y = \varepsilon X \qquad \varepsilon = \begin{cases} +1 & \text{with prob } 1/2 \\ -1 & 1/2 \end{cases}$$

ε indep. from X

$X+Y = (1+\varepsilon)X$

$\mathrm{Var}(X+Y) > 0$

$\mathbb{P}[X+Y=0] = 1/2$

$\Rightarrow X+Y$ is not a Gaussian r.v.

<u>Rk</u> $$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right)$$

maximize the entropy $-\int f(x) \ln f(x)\, dx$

s.t. $\int_{\mathbb{R}} x f(x)\, dx = m$

$\int_{\mathbb{R}} (x-m)^2 f(x)\, dx = \sigma^2$

Gaussian vectors in $\mathbb{R}^d$

<u>Def</u> If X is an $\mathbb{R}^d$-valued rand. var. s.t. $\mathbb{E}[|X|^2] < \infty$ then $\mathrm{Cov}(X)$, the covariance matrix of X is the $d \times d$ matrix defined by

$$\mathrm{Cov}[X]_{i,j} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$$

$$\mathrm{Cov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$$

<u>Ex</u> $v \in \mathbb{R}^d$

$$\mathrm{Var}[v^T X] = v^T \mathrm{Cov}(X) v$$

$$= \mathbb{E}[(v^T(X - \mathbb{E}[X]))^2]$$

$$= \mathbb{E}[v^T (X - \mathbb{E}[X])(X - \mathbb{E}(X))^T v]$$

Gaussian rand. vectors

X: a $\mathbb{R}^d$ valued rand. vect.

X is a Gaussian vector with mean μ and covariance matrix K

$$X \sim \mathcal{N}(\mu, K)$$

iif

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det K}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)}$$

$\mu \in \mathbb{R}^d$

K: symmetric positive definite $d \times d$ matrix

$K_{i,j} = K_{j,i}$

$v^T K v \geq \alpha\, v^T v$ for some $\alpha > 0$

$v^T K v > 0 \quad \forall v \neq 0$

Properties $\mathbb{E}[X] = \mu$

$\mathrm{Cov}[X] = K$

$\forall \lambda \in \mathbb{R}^d$

$$\mathbb{E}[e^{i\lambda^T X}] = e^{-\frac{1}{2}\lambda^T K \lambda + i\lambda^T \mu}$$

$$= e^{-\frac{1}{2}\lambda^T \mathrm{cov}(X)\lambda + i\lambda^T \mathbb{E}[X]}$$

Ex $X \sim \mathcal{N}(\mu, K)$

$v^T X \sim ?$

$v \in \mathbb{R}^d$ $\qquad \lambda = t v$

$$\phi(t) = \mathbb{E}[e^{it\, v^T X}] = \mathbb{E}[e^{i\lambda^T X}]$$

$$= \exp\left(-\frac{t^2}{2} v^T K v + i t v^T \mu\right)$$

characteristic function of

$v^T X \sim \mathcal{N}(v^T \mu, v^T K v)$

Ex $X = (X_1, X_2, X_3)$ $\qquad X \sim \mathcal{N}(0, K)$

$$K = \begin{pmatrix} 30 & 0 & 2 \\ 0 & 20 & 1 \\ 2 & 1 & 20 \end{pmatrix}$$

$$\underbrace{X_2 + X_3}_{\overset{\shortparallel}{v^T X}} \sim \mathcal{N}(0, \underbrace{v^T K v}_{42}) \qquad v = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

<u>Ex</u> X: $\mathbb{R}^d$ valued rand. vect.

s.t. $v^T X$ is Gaussian $\forall v \in \mathbb{R}^d$

Is X a Gaussian vector ?

<u>Sol</u> $\mathbb{E}[e^{i v^T X}] = e^{-\frac{1}{2} \mathrm{Var}[v^T X] + i \mathbb{E}[v^T X]}$

$v^T X$ Gaussian

Y is Gaussian $\Rightarrow$ $\mathbb{E}[e^{iY}] = e^{-\frac{1}{2} \mathrm{Var}[Y] + i \mathbb{E}[Y]}$

$= \exp\left(-\frac{1}{2} v^T \mathrm{Cov}(X) v + i v^T \mathbb{E}[X]\right)$

$\downarrow$ characteristic function of a Gaussian vector

$\Rightarrow X$ is Gaussian

Def We say that X is a Gaussian vector in $\mathbb{R}^d$ with mean μ and covariance matrix K (symm. positive) iif

$$\forall v \in \mathbb{R}^d, \quad v^T X \sim \mathcal{N}(v^T\mu, v^T K v)$$

Ex $X \sim \mathcal{N}(\mu, K)$ $\quad$ X is $\mathbb{R}^d$ valued

M: $d \times d$ matrix

$$MX \sim ? \ \mathcal{N}(M\mu, MKM^T)$$

$$\Uparrow$$

$$v^T(MX) = (M^T v)^T X$$

$$\sim \mathcal{N}(v^T M\mu, v^T MKM^T v)$$

Ex $Z_1, \dots, Z_d$ iid $\mathcal{N}(0,1)$

$Z = (Z_1, \dots, Z_d)$

M: $d \times d$ matrix

$MZ \sim \mathcal{N}(0, MM^T)$

Ex $X \sim \mathcal{N}(\mu, K)$

How to sample from X?

Assume that you can only sample from

$Z = (Z_1, \dots, Z_d)$ $\quad Z_i$ iid $\mathcal{N}(0,1)$

Sol find L: $d \times d$ | $K = LL^T$

$X = X - \mu + \mu$

$X - \mu \sim LZ$

$\Downarrow$ ex

$LZ \sim \mathcal{N}(0, \underbrace{LL^T}_{=K})$

$X \sim \mu + LZ$

K symm. positive $\Rightarrow$ | $\exists L$ lower triangular / $K = LL^T$

$K = LL^T$ Cholesky decomposition of K
factorization

$$\text{Cost} \sim \mathcal{O}(d^3) \rightarrow \text{naive}$$

$$\sim \mathcal{O}\left(d \operatorname{polylog}\left(\frac{d}{\varepsilon}\right)\right) \rightarrow \text{fast Cholesky fact.}$$

<u>Ex</u> $X \sim \mathcal{N}(0, K)$

λ_i, v_i : eigenpairs of K

$$K v_i = \lambda_i v_i$$

$v_1, \ldots, v_d$ forms an orthonormal basis of $\mathbb{R}^d$

$\Downarrow$

$$X = \sum_{j=1}^{d} \langle v_j, X \rangle v_j$$

$$\langle v_j, X \rangle = v_j^T X$$

K invertible $\Rightarrow \lambda_j > 0$

$$X = \sum_{j=1}^{d} \sqrt{\lambda_j}\, Z_j\, v_j$$

$$Z_j = \frac{\langle v_j, X\rangle}{\sqrt{\lambda_j}}$$

$Z = (Z_1, \dots, Z_d)$

$Z \sim ?$

$$\mathbb{E}\left[e^{i\langle t, Z\rangle}\right] = \mathbb{E}\left[e^{i\langle v, X\rangle}\right] = e^{-v^T K v}$$

$t = (t_1, \dots, t_d)$

$$v = \sum_{j=1}^{d} \frac{t_j\, e_j}{\sqrt{\lambda_j}}$$

$e_1, \dots, e_d$ orthon. basis of $\mathbb{R}^d$

$$= \exp\left(-\frac{1}{2}\sum_{j=1}^{d} t_j^2\right)$$

$$v^T K v = \sum_{j=1}^{d} t_j^2$$

$$\mathbb{E}\left[e^{i\langle t, Z\rangle}\right] = e^{-\frac{1}{2}\sum_{j=1}^{d} t_j^2}$$

$Z \sim \mathcal{N}(0, I_d)$

CLT in $\mathbb{R}^d$

Thm X_n : sequ. of $\mathbb{R}^d$ valued r.v. i.i.d.

$$\mathbb{E}[X_i] = \mu \qquad Cor(X_i) = K$$

$$S_n := X_1 + \cdots + X_n$$

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{distrib.}} \mathcal{N}(0,K)$$

Rk $K = LL^T$ $\quad X_i \sim \mathcal{N}(0,K)$

$$(L^T)^{-1} X_i \sim \mathcal{N}(0, \overbrace{L^{-1} K (L^T)^{-1}}^{Id})$$

$$(L^T)^{-1} \frac{S_n - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{distr.}} \mathcal{N}(0, Id)$$

proof Assume $\mu = 0$ $\quad X_j$ indep.

$$\mathbb{E}\left[e^{i\lambda^T \frac{S_n}{\sqrt{n}}}\right] = \prod_{j=1}^{n} \mathbb{E}\left[e^{i\lambda^T X_j/\sqrt{n}}\right]$$

$$= \left(\mathbb{E}\left[e^{i\lambda^T X_1/\sqrt{n}}\right]\right)^n$$

X_j are id. dist.

Taylor expansion

$$\mathbb{E}\left[e^{iv^T X}\right] = 1 + i\,\mathbb{E}[v^T X] - \frac{1}{2}\mathbb{E}\left[|v^T X|^2\right] + o(|v|^2)$$

$$\mathbb{E}[X]=0 \qquad = 1 - \frac{1}{2} v^T \underbrace{\mathbb{E}[XX^T]}_{\mathrm{Cov}(X)} v + o(|v|^2)$$

$$= \left(1 - \frac{\lambda^T K \lambda}{2n} + o\left(\frac{|\lambda|^2}{n}\right)\right)^n$$

$$\xrightarrow[n\to\infty]{} \exp\left(-\frac{\lambda^T K \lambda}{2}\right)$$

charact. funct. of $\mathcal{N}(0,K)$

conv. of charact. funct. $\Rightarrow$ conv. in distr.

Ex Proof of the Chi-Square test

An experiment with d possible outcomes

outcome j occurs with prob. p_j

n repetitions of experiment are done

$$X_{kj} = \begin{cases} 1 & \text{outcome of } k\text{th experiment is } j \\ 0 & \text{else} \end{cases}$$

$$X_1 = (X_{11}, \dots, X_{1d})$$

$$\vdots$$

$$X_n = (X_{n1}, \dots, X_{nd})$$

X_i are iid vectors of $\mathbb{R}^d$

N_j : # of times outcome j appears in n repititions of exper.

$$N_j = X_{1j} + X_{2j} + \dots + X_{nj}$$

$$D^2(n) = \sum_{j=1}^{d} \frac{(N_j - np_j)^2}{np_j}$$

Thm (Karl Pearson)

$$D^2(n) \xrightarrow[n\to\infty]{\text{distr.}} \chi^2(d-1)$$

$$\sim Z_1^2 + \dots + Z_{d-1}^2$$

Z_i iid $\mathcal{N}(0,1)$

proof $Y_i = \left(\frac{X_{i1} - p_1}{\sqrt{p_1}}, \dots, \frac{X_{id} - p_d}{\sqrt{p_d}} \right)$

Y_i iid $\mathbb{E}[Y_i] = 0$

$$\underbrace{\left(\frac{N_1 - np_1}{\sqrt{np_1}}, \dots, \frac{N_d - np_d}{\sqrt{np_d}} \right)}_{V} = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$$

$D^2(n) = \|V\|^2$

$$\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{CLT}} \mathcal{N}(0, C)$$

$$C_{ij} = \mathbb{E}[Y_{1,i} Y_{1,j}]$$

$$= \mathbb{E}\left[\frac{X_{1i} - p_i}{\sqrt{p_i}} \frac{X_{1j} - p_j}{\sqrt{p_j}}\right]$$

$$= \sqrt{p_i p_j} + \frac{\mathbb{E}[X_{1i} X_{1j}]}{\sqrt{p_i p_j}}$$

$$- \frac{\mathbb{E}[X_{1i}] p_j}{\sqrt{p_i p_j}} - \frac{\mathbb{E}[X_{1j}] p_i}{\sqrt{p_i p_j}}$$

$$= \sqrt{p_i p_j} + \delta_{ij} - \sqrt{p_i p_j} - \sqrt{p_i p_j}$$

$$= \delta_{ij} - \sqrt{p_i p_j}$$

$$\sqrt{p} = (\sqrt{p_1}, \ldots, \sqrt{p_d})$$

$$C = I - \sqrt{p}\sqrt{p}^T$$

$C\sqrt{p} = 0$

$Cv = v$ for $v \perp \sqrt{p}$ $\Big| \Rightarrow$ $\lambda_1, \ldots, \lambda_d$ eig of C

$\lambda_1 = \ldots = \lambda_{d-1} = 1$

$\lambda_d = 0$

$$D^2(n) \xrightarrow[n\to\infty]{\text{distr.}} \|Y\|^2$$

$$Y \sim \mathcal{N}(0, C)$$

previous ex $\quad Y = \sum_{j=1}^{d-1} Z_j v_j$

Z_j: iid $\mathcal{N}(0,1)$; v_j: eigenvectors of C

$$\Rightarrow \|Y\|^2 \sim \sum_{j=1}^{d-1} Z_j^2 \sim \chi^2(d-1)$$

Application

6 faced die

100 rolls:

$30 \to 6 \quad\quad 22 \to 2$

$12 \to 5 \quad\quad 8 \to 1$

$10 \to 4$

$18 \to 3$

Is the die fair or not?

$$X_i = (X_{i1}, \dots, X_{i6})$$

$$X_{ij} = \begin{cases} 1 & \text{if outcome roll } i \text{ is } j \\ 0 & \text{otherwise} \end{cases}$$

$$N_j = X_{1j} + X_{2j} + \dots + X_{nj}$$

$$D^2(n) = \sum_{j=1}^{d} \frac{(N_j - n p_j)^2}{n p_j}$$

$$p_j = \frac{1}{6}$$

$$\mathbb{P}[\underbrace{Z_1^2 + \dots + Z_5^2}_{\chi^2(5)} \geq D^2(n)]$$

$= 0.000001$
$\rightarrow$ not fair

$= 0.7 \rightarrow$ fair

Gaussian spaces and Gaussian condition.

$(\Omega, \mathcal{U}, \mathbb{P})$: probability space

Ω : sample space

$\mathcal{U}$: σ-algebra (set of subsets of Ω stable under countable $\cap, \cup, {}^c$)

$\mathbb{P} : \mathcal{U} \to [0,1]$
prob. measure

$L^2(\Omega, \mathcal{U}, \mathbb{P})$: set of r.v. $X : \Omega \to \mathbb{R}$
$\mathcal{U}$-measurable

$$\mathbb{E}[X^2] < \infty$$

Def A (centered) Gaussian space is a closed subspace of $L^2(\Omega, \mathcal{U}, \mathbb{P})$ whose elements are centered Gaussian r.v.

Ex $X_1, \dots, X_d$ iid $\mathcal{N}(0,1)$

$$H = \text{span}\{X_1, \ldots, X_d\}$$
$$= \left\{ \sum_{i=1}^{d} \lambda_i X_i \mid \lambda \in \mathbb{R}^d \right\}$$

H : Gaussian space

Ex X_1, X_2 Gaussian r.v

$\Downarrow$?

Is $\text{Span}\{X_1, X_2\}$ a Gaussian space ?

No $X_1 + X_2$ may not be Gaussian

Ex $(X_1, \ldots, X_d)$ is Gaussian vector

$\Downarrow$?

$\text{Span}\{X_1, \ldots, X_d\}$ is a Gaussian space

YES

Ex $(X_1, X_2, X_3) \sim \mathcal{N}(0, K)$

$$K = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

a) $X_2 + X_3 \sim \mathcal{N}(0, \overbrace{v^T K v}^{6})$

$$v = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

b) Are $X_2 + X_3$ & X_1 indep ?

$$\mathrm{Cov}(X_1, X_2 + X_3) = \underset{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}{\underset{\parallel}{w}}^T K \underset{\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}}{\underset{\parallel}{v}} = 0 + 1 = 1 \neq 0$$

No

$\mathrm{Cov}(Y, Z) \neq 0 \Rightarrow Y, Z$ not indep

Is it true that

$\text{Cor}(Y,Z)=0 \overset{?}{\Longrightarrow} Y, Z$ indep.

Y, Z are both Gaussian

NO

1. $Y \sim N(0,1)$

$Z \sim \varepsilon Y \qquad \varepsilon = \begin{cases} 1 & \text{prob. } 1/2 \\ -1 & 1/2 \end{cases}$

$\text{Cor}(Y,Z)=0$

but not indep. $|Z| = |Y|$

$Z, Y \text{ indep} \overset{?}{\underset{\text{YES}}{\Longrightarrow}} |Z| \text{ indep } |Y|$

Y, Z indep.

Y, Z Gaussian

$\overset{\text{YES}}{\Longrightarrow} Y, Z$ jointly Gaussian

$(v, w) \quad \begin{pmatrix} Y \\ Z \end{pmatrix} = \underbrace{v^T Y}_{\text{Gaussian}} + \underbrace{w^T Z}_{\text{Gaussian}} \;)$ Gaussian

indep

Is it true that

$$\left.\begin{array}{l} Cov(Y,Z)=0 \\ (Y,Z) \text{ is Gaussian} \end{array}\right\} \overset{?}{\Longrightarrow} Y,Z \text{ indep}$$

YES

$$\mathbb{E}\left[e^{i\lambda Y + i\mu Z}\right] \overset{?}{=} \mathbb{E}\left[e^{i\lambda Y}\right]\mathbb{E}\left[e^{i\mu Z}\right]$$

$$\shortparallel$$

$$e^{i(\lambda,\mu)\mathbb{E}\left[\binom{Y}{Z}\right] - \frac{1}{2}(\lambda,\mu)\,Cov\binom{Y}{Z}\binom{\lambda}{\mu}}$$

$$\shortparallel$$

$$e^{i\lambda\mathbb{E}[Y] + i\mu\mathbb{E}[Z] - \frac{1}{2}\lambda^2 Var[Y] - \frac{1}{2}\mu^2 Var(Z) - \lambda\mu \underbrace{\cancel{Cov(Y,Z)}}_{=0}}$$

Ex $(X_1, X_2, X_3) \sim \mathcal{N}(0, K)$

$$K = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

d Are $X_2 + X_3$ and $X_2 - X_3$ indep ?

$(X_2 + X_3, X_2 - X_3)$ is Gaussian

$$\mathrm{Cov}(X_2 + X_3, X_2 - X_3)$$

$$\parallel$$

$$(0\ 1\ 1)\, K \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = 2 - 2 = 0$$

YES

$X = (X_1, \dots, X_d)$ Gaussian vector

$X_1, \dots, X_d$ indep $\Rightarrow$ $\mathrm{Cov}(X)$ diagonal

$\overset{?}{\Leftarrow}$ YES

because characteristic function is a product

$$\mathbb{E}\left[e^{i\sum_{j=1}^{d} \lambda_j X_j}\right] = \prod_{j=1}^{d} \mathbb{E}\left[e^{i\lambda_j X_j}\right]$$

<u>Thm</u> Let H be a Gaussian space and let $(H_i, i \in I)$ be a family of closed sub-vector spaces of H. Then the subspaces H_i for $i \in I$ are orthogonal in $\mathbb{L}^2$ ($\Leftrightarrow \forall X_i \in H_i, X_j \in H_j$ $\mathbb{E}[X_i X_j] = 0$)

i iff the σ-algebra $\sigma(H_i)$ for $i \in I$ are indep. ($\sigma(H_i)$ is the σ-algebra generated by rand. var. elements of H_i)

$\sigma(H_i)$ indep $\sigma(H_j)$

$\forall\, X_{i,1}, \dots, X_{i,m} \in H_i$

$X_{j,1}, \dots, X_{j,n} \in H_j$

$\forall\, f, g \in C_b$

$f(X_{i,1}, \dots, X_{i,m})$ indep $g(X_{j,1}, \dots, X_{j,n})$

$H_i \perp_{L^2} H_j$ $\Leftrightarrow$ All r.v. in H_i indep from all r.v. in H_j

$\forall i \neq j$ $\quad\quad$ $\forall i \neq j$

$H_i \subset H$

H Gaussian

easy

$\Leftarrow$

$\Rightarrow$

charact. funct.

Ex $(X_1, X_2, X_3) \sim \mathcal{N}(0, K)$

$$K = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

d $\mathbb{E}[X_1 | X_2, X_3]$?

Reminder $\mathbb{E}[X_1 | X_2, X_3] = f(X_2, X_3)$

where f minimizes

$$\mathbb{E}[|X_1 - f(X_2, X_3)|^2]$$

$$X_1 = X_1 - aX_2 - bX_3 + aX_2 + bX_3$$

$$\mathbb{E}[X_1 | X_2, X_3] = \mathbb{E}[X_1 - aX_2 - bX_3 | X_2, X_3]$$
$$+ \underbrace{\mathbb{E}[aX_2 + bX_3 \mid X_2, X_3]}_{aX_2 + bX_3}$$

if $X_1 - aX_2 - bX_3$ indep from X_2, X_3 $\Rightarrow$ $\mathbb{E}[X_1 - aX_2 - bX_3 \mid X_2, X_3] = \mathbb{E}[X_1 - aX_2 - bX_3] = 0$

$\Updownarrow$

$$\mathrm{Cov}(X_1 - aX_2 - bX_3, X_2) = 0$$
$$\mathrm{Cov}(X_1 - aX_2 - bX_3, X_3) = 0$$

$$\begin{cases} -2a - b = 0 \\ 1 - a - 2b = 0 \end{cases} \Leftrightarrow \quad a = -1/3, \quad b = 2/3$$

$$\mathbb{E}[X_1 \mid X_2, X_3] = -\frac{X_2}{3} + \frac{2}{3}X_3$$
$$= aX_2 + bX_3 \quad \text{where}$$

a, b minimize

$$\mathbb{E}[|X_1 - aX_2 - bX_3|^2]$$

e! $\mathcal{L}aw(X_1 / X_2, X_3)$?

$X = (X_1, X_2, X_3) \sim \mathcal{N}(0, K)$

$K = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ $\quad a = -1/3 \quad b = 2/3$

$$X_1 = \underbrace{\mathbb{E}[X_1 | X_2, X_3]}_{\underbrace{aX_2 + bX_3}_{\text{behaves like a cst if } X_2 \,\&\, X_3 \text{ known}}} + \underbrace{X_1 - \mathbb{E}[X_1 | X_2, X_3]}_{X_1 - aX_2 - bX_3}$$

? indep from X_2, X_3

$\mathcal{N}(0, \sigma^2)$

$\sigma^2 = Var(X_1 - aX_2 - bX_3) = \frac{7}{9}$

because $X_1 - aX_2 - bX_3$

indep from X_2, X_3

$\Downarrow$

$\mathcal{L}aw(X_1 - aX_2 - bX_3 | X_2, X_3) = \mathcal{L}aw(X_1 - aX_2 - bX_3$

Law $(X_1 | X_2, X_3) \sim \mathcal{N}\left(-\frac{X_2}{3} + \frac{2}{3} X_3, \frac{7}{9}\right)$

<u>Ex</u> $X = (X_1, X_2, X_3) \sim \mathcal{N}(\underset{\substack{\downarrow \\ \text{not centered}}}{m}, K)$

$$K = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$\mathbb{E}[X_1 | X_2, X_3] = \; ?$$

$$X_1 = X_1 - m_1 + m_1$$

$$\mathbb{E}[X_1 | X_2, X_3] = m_1 + \mathbb{E}[X_1 - m_1 | X_2, X_3]$$

$$= m_1 + \mathbb{E}[\underbrace{X_1 - m_1}_{Y_1} | \underbrace{X_2 - m_2}_{Y_2}, \underbrace{X_3 - m_3}_{Y_3}]$$

$$Y \sim \mathcal{N}(0, K)$$

$$\downarrow$$

$$= \mathbb{E}[X_1] - \frac{1}{3}\left(X_2 - \mathbb{E}[X_2]\right) + \frac{2}{3}\left(X_3 - \mathbb{E}[X_3]\right)\Big)$$

<u>Thm</u> H: Gaussian space

V: closed sub.vector space of H

Pr: orthogonal projection on V

$$\mathrm{pr}: H \longrightarrow V$$

$$X \longrightarrow \mathrm{pr}(X) = \underset{Z \in V}{\operatorname{argmin}}\, \mathbb{E}[|X - Z|^2]$$

Let $X \in H$, then

(i) $\mathbb{E}[X \mid \sigma(V)] = \mathrm{pr}(X)$

(ii) $\sigma^2 := \mathbb{E}[(X - P_V(X))^2]$ then

$$X \,|\, \sigma(V) \sim \mathcal{N}(P_V(X), \sigma^2)$$

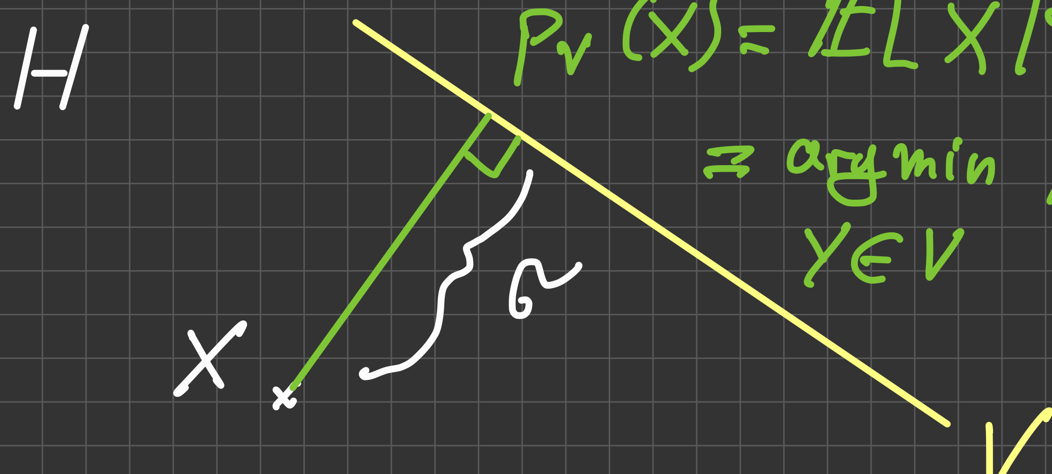
$\forall \Gamma \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}[X \in \Gamma \,|\, \sigma(V)] = \frac{1}{\sigma\sqrt{2\pi}} \int_\Gamma dy \, e^{\frac{(y - P_V(X))^2}{2\sigma^2}}$$

<u>Rk</u> important

in general $\mathbb{E}[X \,|\, \sigma(V)] = P_{L^2(\Omega, \sigma(V), \mathbb{P})}(X)$

here $= P_V(X)$



$$P_V(X) = \mathbb{E}[X \,|\, \sigma(V)] = \underset{Y \in V}{\operatorname{argmin}} \, \mathbb{E}[|X - Y|^2]$$

$$\sigma^2 = \mathrm{Var}[X \,|\, \sigma(V)] = \mathbb{E}[|X - P_V(X)|^2]$$

$$= \min_{Y \in V} \mathbb{E}[|X-Y|^2]$$

<u>proof</u>

$$X = (X - pr(X)) + pr(X)$$

$$X - pr(X) \perp_{L^2} pr(X)$$

$$\Rightarrow X - pr(X) \text{ indep. from } pr(X)$$

$$\mathbb{E}[X | \sigma(V)] = \mathbb{E}[pr(X) | \sigma(V)] = pr(X)$$

$$+ \underbrace{\mathbb{E}[X - pr(X) | \sigma(V)]}_{\mathbb{E}[X - pr(X)] = 0}$$

$$\mathbb{E}[X | \sigma(V)] = pr(X)$$

$$X | \sigma(V) = \underbrace{(X - pr(X))}_{X - pr(X) \sim \mathcal{N}(0, \sigma^2)} |_{\sigma(V)} + \underbrace{pr(X) |_{\sigma(V)}}_{cst}$$

$$X \mid \sigma(V) \sim \mathcal{N}\left(p_V(X), \sigma^2\right)$$

Cor $X \sim \mathcal{N}(0, K)$ Gaussian vector on $\mathbb{R}^d$ $d \geq 2$

K: non-degenerate

ϕ: $m \times d$ matrix rank m $0 < m < d$

$$\mathbb{E}[X \mid \phi X = y] = K\phi^T(\phi K \phi^T)^{-1} y$$

$$X \mid \phi X = y \sim \mathcal{N}(\mathbb{E}[X \mid \phi X = y], C)$$

$$C = K - K\phi^T(\phi K \phi^T)^{-1}\phi K$$

$$\mathbb{E}[X \mid \phi X = y] = \arg \left| \begin{array}{ll} \min & z^T K^{-1} z \\ \text{over} & z \in \mathbb{R}^d \\ \text{s.t.} & \phi z = y \end{array} \right.$$

(optimal recovery)

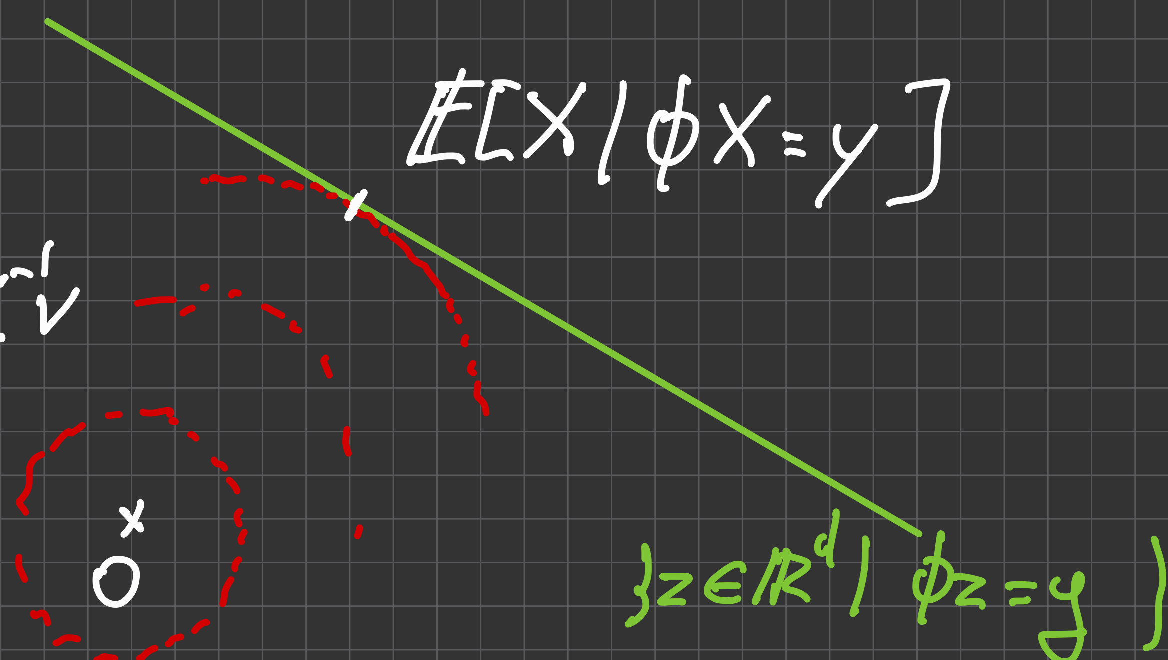
For $v \in \mathbb{R}^d$

$$v^T C v = \min_{w \in \mathbb{R}^m} (v - \phi^T w)^T K (v - \phi^T w)$$

Geometric interpretation

$\mathbb{R}^d$

$\langle u, v \rangle_K = u^T K^{-1} v$



$$\mathrm{pdf}(X) \propto \exp\left(- \frac{z^T K^{-1} z}{2}\right)$$

$$C = \mathrm{Cov}\left(X - \mathbb{E}[X \mid \phi X]\right)$$

$$v^T C v = v^T K^{\phi^\perp} v = \min_{w \in \mathbb{R}^m} (v - \phi^T w)^T K (v - \phi^T w)$$

$v \in \mathbb{R}^d$

Schur complement of matrix K

operator K

w.r. to $\{\phi^T w, w \in \mathbb{R}^m\} \subset \mathbb{R}^d$

(Shorted operator)

<u>proof</u> $X | \phi X$

$v^T X | \phi X \qquad \forall v \in \mathbb{R}^d$

<u>Prop</u> $v^T X | \phi X \sim \mathcal{N}(\mathbb{E}[v^T X | \phi X], \sigma^2)$

$$\mathbb{E}[v^T X | \phi X] = w^T \phi X$$

$$w = \arg\min_{w} \mathbb{E}[|v^T X - w^T \phi X|^2]$$

$$F(w) = \mathbb{E}[|v^T X - w^T \phi X|^2]$$

$$= \mathbb{E}[v^T X X^T v + w^T \phi X X^T \phi^T w - 2 w^T \phi X X^T v]$$

$$= v^T K v + w^T \phi K \phi^T w - 2 w^T \phi K v$$

$$\mathbb{E}[XX^T] = K$$

$$\nabla_w F(w) = 0 \iff \phi K \phi^T w = \phi K v$$

$$w = (\phi K \phi^T)^{-1} \phi K v$$

$$\Rightarrow \mathbb{E}[v^T X | \phi X] = w^T \phi X$$

$$= v^T K \phi^T (\phi K \phi^T)^{-1} \phi X$$

$$\Rightarrow$$

$$\mathbb{E}[X | \phi X] = K \phi^T (\phi K \phi^T)^{-1} \phi X$$

representer formula

$$\sigma^2 = \mathbb{E}\left[\,\left|v^T X - \mathbb{E}[v^T X \mid \phi X]\right|^2\right]$$

$$\Downarrow$$

$$C = \mathrm{Cov}\big(X - \mathbb{E}[X \mid \phi X]\big)$$

$$= \mathbb{E}\left[(X - \mathbb{E}[X \mid \phi X])(X - \mathbb{E}[X \mid \phi X])^T\right]$$

$$= \mathbb{E}[XX^T] + \mathbb{E}\left[\mathbb{E}[X \mid \phi X]\,\mathbb{E}[X \mid \phi X]^T\right]$$

$$- \mathbb{E}\left[X\,\mathbb{E}[X \mid \phi X]^T\right] - \mathbb{E}\left[\mathbb{E}[X \mid \phi X]\,X^T\right]$$

$$K = \mathbb{E}[XX^T]$$

$$\mathbb{E}[X \mid \phi X] = K\phi^T(\phi K \phi^T)^{-1}\phi X$$

$$= K + K\phi^T(\phi K\phi^T)^{-1}\cancel{\phi K\phi^T(\phi K\phi^T)^{-1}}\phi K$$

$$- K\phi^T(\phi K\phi^T)^{-1}\phi K - K\phi^T(\phi K\phi^T)^{-1}\phi K$$

$$= K - K\phi^T(\phi K\phi^T)^{-1}\phi K$$

$v^T X | \phi X$ Gaussian $\forall v$

$$\Downarrow$$

$X | \phi X$ Gaussian

$$\sim \mathcal{N}(\mathbb{E}[X|\phi X], C]$$

Now let us prove

$$\boxed{\begin{array}{l} (1)\quad \mathbb{E}[X|\phi X = y] = \arg \left| \begin{array}{ll} \min & z^T K^{-1} z \\ \text{over} & z \in \mathbb{R}^d \\ \text{s.t.} & \phi z = y \end{array} \right. \\ \qquad \text{(optimal recovery)} \\ \qquad \text{For } v \in \mathbb{R}^d \\ (2)\quad v^T C v = \min\limits_{w \in \mathbb{R}^m} (v - \phi^T w)^T K (v - \phi^T w) \end{array}}$$

We have

$$\mathbb{E}[X | \phi X = y] = K \phi^T (\phi K \phi^T)^{-1} y$$

Write $v := K\phi^T(\phi K \phi^T)^{-1} y$

$$\phi v = \phi K \phi^T (\phi K \phi^T)^{-1} y = y$$

$\Rightarrow$ v satisfies const. of (1)

Let $w \in \mathbb{R}^d \,/\, \phi w = y$

$$w = v + w - v$$

$$w^T K^{-1} w = v^T K^{-1} v + (w-v)^T K^{-1} (w-v) + 2\, v^T K^{-1} (w-v)$$

$$v^T K^{-1} (w-v) = y^T (\phi K \phi^T)^{-1} \phi \cancel{K K^{-1}} (w-v)$$

$= 0$ because $\phi(w-v) = 0$

$\Rightarrow$
$$w^T K^{-1} w = v^T K^{-1} v + (w-v)^T K^{-1} (w-v)$$

$\Rightarrow$ $v = \arg \left|\begin{array}{l} \min\ z^T K^{-1} z \\ \text{s.t. } \phi z = y \end{array}\right.$

Now let us prove (2)

$$\mathbb{E}[v^T X \mid \phi X] = w^T \phi X$$

$$w = \underset{\bar{w}}{\arg\min}\ \mathbb{E}[|v^T X - \bar{w}^T \phi X|^2]$$

$$\sigma^2 = v^T C v = \mathbb{E}[|v^T X - w^T \phi X|^2]$$

$$\sigma^2 = \min_{\bar{w}} \mathbb{E}[|v^T X - \bar{w}^T \phi X|^2]$$

$$< \min_{\bar{w}} \mathbb{E}[(v - \phi^T \bar{w})^T (X X^T)(v - \phi^T \bar{w})]$$

$$= \min_{\bar{w}} (v - \phi^T \bar{w})^T K (v - \phi^T \bar{w})$$

$$v^T C v = \min_{w} (v - \phi^T w)^T K (v - \phi^T w)$$

Conditional covariance and precision matrix

Y, Z finite dim rand. vect.

$\mathrm{Cov}(Y)$: cov matrix of Y

$(\mathrm{Cov}(Y))_{ij} = \mathrm{Cov}(Y_i, Y_j)$

$\mathrm{Cov}(Y|Z)$: cond cov of Y given Z

$(\mathrm{Cov}(Y|Z))_{i,j} = \mathrm{Cov}(Y_i, Y_j | Z)$

$\mathrm{Cov}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^T]$

$\mathrm{Cov}(Y, Z) = \mathbb{E}[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])^T]$

If $\mathrm{Cov}(Y)$ invertible

$\mathrm{Cov}(Y)^{-1}$: precision matrix of Y

X : r.v. on $\mathbb{R}^n$

$n \geqslant 3$ $X = (X_1, \dots, X_n) \sim \mathcal{N}(m, \Theta)$

$X_1, X_2 / X_3, \dots, X_n$

$\operatorname{Cor}(X_1, X_2 / X_3, \dots, X_n)$

as a function $\Theta^{-1}_{1,2}$

X r.v on $\mathbb{R}^n$

M $m \times n$ matrix

$\operatorname{Cov}(MX) = M \operatorname{Cov}(X) M^T$

$n \geqslant 3$ $X = (X_1, \dots, X_n) \sim \mathcal{N}(m, \Theta)$

$Y = (X_1, X_2)$ $Z = (X_3, \dots, X_n)$

$m = (m_y, m_z)$

$$\Theta = \begin{pmatrix} \Theta_{yy} & \Theta_{y,z} \\ \Theta_{zy} & \Theta_{zz} \end{pmatrix}$$

$\Theta_{yy} = Cov(Y)$ $\qquad$ $\Theta_{y,z} = Cov(Y, Z)$

$Y \sim \mathcal{N}(m_y, \Theta_{yy})$ $\qquad$ $Z \sim \mathcal{N}(m_z, \Theta_{zz})$

Prop $\mathbb{E}[Y \mid Z] = m_y + \Theta_{yz} \Theta_{zz}^{-1} (z - m_z)$

$$Cov(Y \mid Z) = \Theta_{yy} - \Theta_{yz} \Theta_{zz}^{-1} \Theta_{zy}$$

proof

$Z = \phi X$ $\qquad$ $\phi : (n-2) \times n$

previous prop $\rightarrow$ know

$X \mid \phi X$

$$\mathbb{E}[Y|Z] = m_Y + \mathbb{E}[Y - m_Y \mid Z - m_Z]$$

$$Y = Y - m_Y + m_Y \qquad \overset{\shortparallel}{\beta\,(\underbrace{Z - m_Z}_{(n-2)\text{ vec.}})}$$

β : $2 \times (n-2)$ matrix

$$\mathbb{E}[Y|Z] = m_Y + \beta\,(Z - m_Z)$$

$$\mathrm{Cov}(Y - m_Y - \beta(Z - m_Z),\ Z - m_Z) = 0$$

$$\shortparallel$$

$$\mathbb{E}[(Y - m_Y - \beta(Z - m_Z))(Z - m_Z)^T]$$

C

$$\mathbb{E}[Y|Z] = m_Y + \mathbb{E}[Y - m_Y - \beta(Z - m_Z) \mid Z - m_Z]$$
$$+ \mathbb{E}[\beta(Z - m_Z) \mid Z - m_Z]$$

$$Y = Y - m_Y - \beta(Z - m_Z) + \beta(Z - m_Z) + m_Y$$

$$\mathbb{E}[Y \mid Z] = \mathbb{E}[Y - m_Y - \beta(Z - m_Z) \mid Z - m_Z] + \beta(Z - m_Z) + m_Y$$

$$Y - m_Z - \beta(Z - m_Z) \perp Z - m_Z$$

$$\Downarrow$$

$$Y - m_Z - \beta(Z - m_Z) \text{ indep from } Z - m_Z$$

$$\Downarrow$$

$$\mathbb{E}[(Y - m_Y - \beta(Z - m_Z) \mid Z - m_Z]$$

$$\|$$

$$\mathbb{E}[Y - m_Y - \beta(Z - m_Z)] = 0$$

$$\mathbb{E}[Y|Z] = m_Y + \beta(Z - m_Z)$$

$$\mathrm{Cov}(Y - m_Y - \beta(Z - m_Z), Z - m_Z) = 0$$

$$\parallel$$

$$\mathbb{E}\left[(Y - m_Y - \beta(Z - m_Z))(Z - m_Z)^T\right] = 0$$

$$\parallel$$

$$\Theta_{YZ} - \beta\,\Theta_{ZZ} = 0$$

$$\beta = \Theta_{YZ}\,\Theta_{ZZ}^{-1}$$

$$\mathbb{E}[Y|Z] = m_Y + \Theta_{YZ}\,\Theta_{ZZ}^{-1}(Z - m_Z)$$

$$\mathrm{Cov}(Y|Z) = \mathrm{Cov}(Y - m_Y - \beta(Z - m_Z))$$

why?

$$Y|_Z = \underbrace{Y - m_Y - \beta(Z - m_Z)}_{\text{indep from } Z} + \underbrace{m_Y + \beta(Z - m_Z)}_{\text{cst given } Z}$$

$\underline{Q}$ Why $Y - m_Y - \beta(Z - m_Z)$

and $(Z - m_Z)$

are jointly Gaussian

Because $\forall v \in \mathbb{R}^2, \forall w \in \mathbb{R}^{n-2}$

$$v^T(Y - m_Y - \beta(Z - m_Z)) + w^T(Z - m_Z) = \lambda^T(X - m_X)$$

$\lambda \in \mathbb{R}^n$

$$\mathrm{Cov}(Y|Z) = \mathrm{Cov}(Y - m_Y - \beta(Z - m_Z))$$

$$= \mathbb{E}[(Y - m_Y - \beta(Z - m_Z))(Y - m_Y - \beta(Z - m_Z))^T]$$

$$= \underbrace{\Theta_{YY} - \Theta_{YZ}\Theta_{ZZ}^{-1}\Theta_{ZY}}$$

$$\beta = \Theta_{YZ}\Theta_{ZZ}^{-1}$$

$$\Theta = \begin{pmatrix} \Theta_{YY} & \Theta_{YZ} \\ \Theta_{ZY} & \Theta_{ZZ} \end{pmatrix}$$

Schur complement of block ZZ of matrix Θ

$$\mathrm{Cov}(Y|Z) = \Theta_{YY} - \Theta_{YZ}\Theta_{ZZ}^{-1}\Theta_{ZY}$$

$$\begin{aligned}
\mathrm{Var}[\rho^T Y \mid Z] &= \rho^T \mathrm{Cov}(Y|Z)\,\rho \\
&= \mathrm{Var}[\rho^T Y - \rho^T \beta (Z - m_Z)] \\
&= \inf_{v \in \mathbb{R}^{n-2}} \mathrm{Var}[\rho^T Y - v^T (Z - m_Z)] \\
&= \inf_{v \in \mathbb{R}^{n-2}} (\rho, -v)\,\Theta \begin{pmatrix} \rho \\ -v \end{pmatrix} \\
&= \underbrace{\inf_{v \in \mathbb{R}^{n-2}} (\rho, v)\,\Theta \begin{pmatrix} \rho \\ v \end{pmatrix}}_{\text{variational formulation of the Schur complement}}
\end{aligned}$$

Shorted Gperator (if $\dim(Z) = \infty$)

$$\mathrm{Var}[\rho^T Y \mid Z] = \mathrm{Var}[\rho^T Y - \rho^T \mathbb{E}[Y|Z]]$$

$$\rho^T Y = \underbrace{\rho^T Y - \rho^T \mathbb{E}[Y|Z]}_{\text{indep from } Z} + \underbrace{\rho^T \mathbb{E}[Y|Z]}_{\text{cst given } Z}$$

$$\mathrm{Var}[\rho^T Y|Z] = \mathrm{Var}[\rho^T Y - \rho^T \mathbb{E}[Y|Z])$$

$$= \mathrm{Var}[\rho^T Y - \rho^T \beta(Z - m_Z)]$$

$$= \inf_{v \in \mathbb{R}^{n-2}} \mathrm{Var}[\rho^T Y - v^T (Z - m_Z)]$$

$$\mathbb{E}[Y|Z] = m_Y + \beta(Z - m_Z)$$

$$\mathrm{Var}[\rho^T Y - \rho^T \beta(Z - m_Z)]$$

$$= \mathrm{Var}[\rho^T Y - \rho^T \mathbb{E}[Y - m_Y|Z]]$$

$\downarrow$

$$= \mathrm{Var}\left[\rho^T (Y - m_Y) - \rho^T \mathbb{E}[(Y - m_Y) | Z]\right]$$

$$= \mathrm{Var}\left[\rho^T (Y - m_Y) - \mathbb{E}\left[\rho^T (Y - m_Y) | Z - m_Z\right]\right]$$

$\mathbb{E}\left[\rho^T (Y - m_Y) | Z - m_Z\right]$

is $\perp$ proj. of $\rho^T (Y - m_Y)$ onto

$\mathrm{span}_v \{ v^T (Z - m_Z) \}$

$$= \min_v \mathrm{Var}\left[\rho^T (Y - m_Y) - v^T (Z - m_Z)\right]$$

<u>Prop</u>

$$\mathrm{Cov}(Y_1, Y_2 | Z) = \frac{-\Theta_{1,2}^{-1}}{\Theta_{1,1}^{-1} \Theta_{2,2}^{-1} - (\Theta_{1,2}^{-1})^2}$$

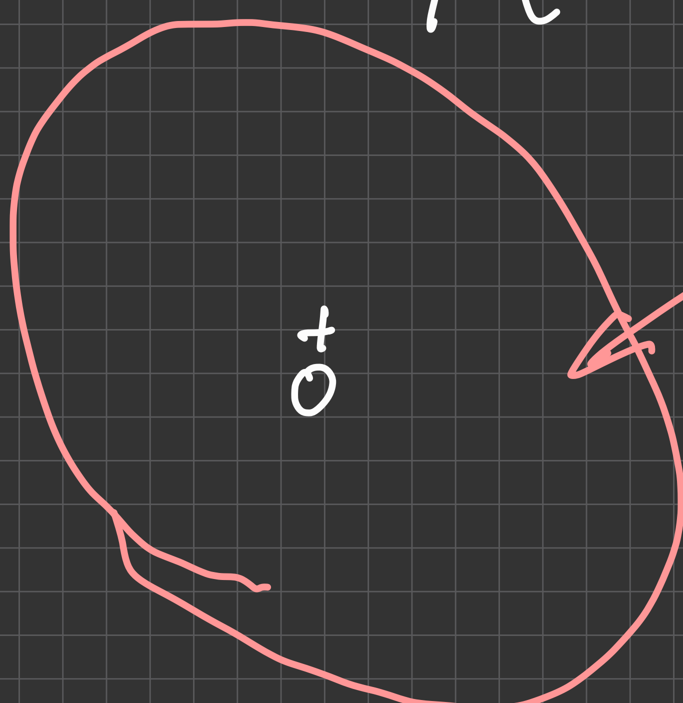
$$\mathrm{Var}[Y_1 | Z] = \frac{\Theta_{2,2}^{-1}}{\Theta_{1,1}^{-1} \Theta_{2,2}^{-1} - (\Theta_{1,2}^{-1})^2}$$

$$\mathrm{Cor}[Y_1, Y_2 \mid Z] := \frac{\mathrm{Cov}(Y_1, Y_2 \mid Z)}{\sqrt{\mathrm{var}(Y_1 \mid Z)}\,\sqrt{\mathrm{var}(Y_2 \mid Z)}}$$

$$= -\frac{\Theta^{-1}_{1,2}}{(\Theta^{-1}_{1,1})^{1/2}\,(\Theta^{-1}_{2,2})^{1/2}}$$

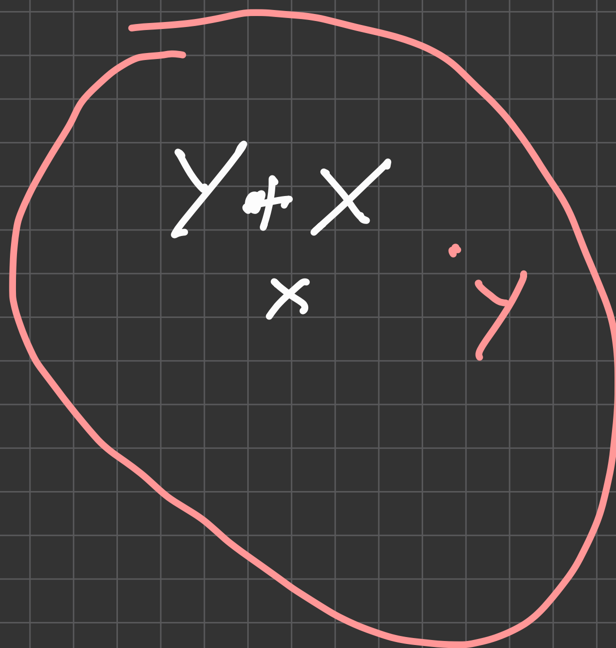
$$X \sim \mathcal{N}(m, \Theta)$$

$$p(x) \propto \exp\left(-\frac{x^T \Theta^{-1} x}{2}\right)$$



$x^T \Theta^{-1} x \leq \alpha$

$Y + X$



$$\left\{ y \;\middle|\; (y-(Y+X))^T \Theta^{-1} (y-(Y+X)) \le \alpha \right\}$$

Proof

$$\Theta^{-1} = \begin{pmatrix} (\Theta_{YY} - \Theta_{YZ}\Theta_{ZZ}^{-1}\Theta_{ZY})^{-1} & 0 \\ -\Theta_{ZZ}^{-1}\Theta_{ZY}(\Theta_{YY} - \Theta_{YZ}\Theta_{ZZ}^{-1}\Theta_{ZY})^{-1} & 0 \end{pmatrix}$$

$$(\Theta^{-1})_{Z,Z} = (\Theta_{ZZ} - \Theta_{ZY}\Theta_{YY}^{-1}\Theta_{YZ})^{-1}$$

$$(\Theta^{-1})_{YZ} = -\Theta_{YY}^{-1}\Theta_{YZ}(\Theta_{ZZ} - \Theta_{ZY}\Theta_{YY}^{-1}\Theta_{YZ})^{-1}$$

$$\mathrm{Cor}(Y|Z) = \Theta_{YY} - \Theta_{YZ}\Theta_{ZZ}^{-1}\Theta_{ZY}$$

$$\left(\mathrm{Cor}(Y \mid Z)\right)^{-1} = \begin{pmatrix} \Theta_{1,1}^{-1} & \Theta_{1,2}^{-1} \\ \Theta_{2,1}^{-1} & \Theta_{2,2}^{-1} \end{pmatrix}$$

$$\Theta = \left(\begin{array}{c|c} & \\ \hline & \end{array}\right)$$

$$\Theta^{-1} = \left(\begin{array}{c|c} & \\ \hline & \end{array}\right)$$

Gaussian Processes

T: arbitrary set

<u>Def</u> A family of rand. var $\{X_t, t \in T\}$ is a (centered) Gaussian Process (GP) iif $\forall n \in \mathbb{N}$, $\forall t_1, \dots, t_n \in T$ $\forall \lambda_1, \dots, \lambda_n \in \mathbb{R}$

$$\lambda_1 X_{t_1} + \dots + \lambda_n X_{t_n}$$

is a (centered) Gaussian rand. var

<u>Ex</u> $T = \{1, 2, \dots, n\}$

$X = (X_1, \dots, X_n)$ is a GP iif X is a Gaussian vector

<u>Def</u> A stochastic process X is a collection $\{X_t, t \in T\}$ of rand. var on a prob space

$(\Omega, \mathcal{U}, \mathbb{P})$

Def For a stochastic process X we define $\sigma(X)$ as the σ-algebra generated by events

$$(X_{t_1}, \dots, X_{t_k}) \in A \qquad A \in \mathcal{B}(\mathbb{R}^k)$$

Thm The probability distribution of X is uniquely determined by its finite dimensional distributions

$$\mu_{t_1, \dots, t_k}(A) = \mathbb{P}[(X_{t_1}, \dots, X_{t_k}) \in A]$$
$$A \in \mathcal{B}(\mathbb{R}^k)$$

Ex $(X_t)_{t \in T}$

$(X_{t_1}, \dots, X_{t_k})$ is a Gaussian vector

because $\sum_i \lambda_i X_{t_i}$ is Gaussian $\forall \lambda \in \mathbb{R}^k$

$$(X_{t_1}, \dots, X_{t_k}) \sim \mathcal{N}\left(\left(\dots \mathbb{E}[X_{t_i}] \dots \right), \left(\mathrm{Cov}(X_{t_i}, X_{t_j}) \right) \right)$$

$$\mathbb{E}[X_t], \quad \mathrm{Cov}(X_t, X_s) \quad \forall t, s$$

<u>Def</u> Let $(X_t)_{t \in T}$ be a GP

The covariance function of X is the function

$$\Gamma : T \times T \longrightarrow \mathbb{R}$$

$$\Gamma(s,t) := \mathrm{Cov}(X_s, X_t)$$

$\left(= \mathbb{E}[X_s X_t] \text{ if } X \text{ is centered} \right)$

Thm The prob. law of X is uniquely determined $m(t) = \mathbb{E}[X_t]$

$\Gamma(s,t) = \operatorname{Cov}(X_s, X_t)$

$X \sim GP(m, \Gamma) \Leftrightarrow$ X is a GP with mean m, cov. fct. Γ

$\updownarrow$

$\sim \mathcal{N}(m, \Gamma)$

Ex $\Gamma: T \times T \to \mathbb{R}$

$s, t \quad \operatorname{Cov}(X_s, X_t)$

$\Gamma(s,t) = \Gamma(t,s)$

$\forall k, t_1, \dots, t_k \in T$

$\Theta_{i,j} = \Gamma(t_i, t_j)$ must be positive

$\Leftrightarrow \forall c \in \mathbb{R}^k, \quad \sum_{i,j} c_i c_j \Gamma(t_i, t_j) \geq 0$

Def We say that a function Γ on $T \times T$ is symmetric and of positive type iif

$$\Gamma(s,t) = \Gamma(t,s) \quad \forall s, t$$

$\forall k, t_1, \dots, t_k, c \in \mathbb{R}^k$

$$\sum_{i,j} c_i c_j \Gamma(t_i, t_j) \geq 0$$

Thm Let $m : T \longrightarrow \mathbb{R}$

$\Gamma : T \times T \longrightarrow \mathbb{R}$

symmetric positive

then $\exists$ a GP $X \sim GP(m, \Gamma)$

proof Kolmogorov existence theorem

Ex $T = \mathbb{R}_+$

$\Gamma(s,t) = \min(s,t)$ $\Rightarrow$ $X \sim GP(0,\Gamma)$

$\downarrow$ Brownian Motion

$T = [0,1]^2$ $\quad \Gamma(s,t) = \min(s_1,t_1)\min(s_2,t_2)$

$\to X \sim GP(0,\Gamma)$

$\to$ Brownian Sheet

$$T = \mathbb{L}^2(E,\mathcal{E},\mu)$$

$$= \left\{ f : E \to \mathbb{R} \;\middle|\; \int_E f^2 \, d\mu < \infty \right\}$$

$$\Gamma(f,g) = \langle f, g \rangle_{\mathbb{L}^2}$$

$$= \int_E f(x)\, g(x)\, \mu(dx)$$

Gaussian measure

Gaussian field

<u>Ex</u> $\xi_1 \sim GP(0, K_1)$ $\quad$ ξ_1, ξ_2 indep

$\xi_2 \sim GP(0, K_2)$

$\xi_1 + \xi_2$

ξ_1 and ξ_2 indep $\Longleftrightarrow$ $\forall k, t_1, \dots, t_k$
$k', s_1, \dots, s_{k'}$

$(\xi_1(t_1), \dots, \xi_1(t_k))$

indep from $(\xi_2(s_1), \dots, \xi_2(s_{k'}))$

$$\xi_1 + \xi_2 \sim GP(0, K_1 + K_2)$$

<u>Ex</u> $\xi_1 \mid \xi_1 + \xi_2$?

Kolmogorov existence theorem

The prob. dist of stochastic process X on T is uniquely determined by its finite dim distributions

$$* \quad \mu_{t_1,\dots,t_k}(A) = \mathbb{P}[(X_{t_1},\dots,X_{t_k}) \in A] \quad \forall A \in \mathcal{B}(\mathbb{R}^k)$$

$$\forall k \in \mathbb{N}, \forall t_1,\dots,t_k \in T$$

$\overset{(*)}{\Longrightarrow}$ two consistency properties

(1) $\forall k \in \mathbb{N}$, $t_1, \dots, t_k \in T$, $A_1 \times \dots \times A_k \in \mathcal{B}(\mathbb{R}^k)$ and π a permutation of $\{1,\dots,k\}$

$$\mu_{t_1,\dots,t_k}(A_1 \times \dots \times A_k) = \mu_{t_{\pi_1}\dots t_{\pi_k}}(A_{\pi_1} \times \dots \times A_{\pi_k})$$

(2) $\forall k \in \mathbb{N}$, $t_1,\dots,t_k \in T$, $A_1 \times \dots \times A_{k-1} \in \mathcal{B}(\mathbb{R}^{k-1})$

$$\mu_{t_1,\dots,t_{k-1}}(A_1 \times \dots \times A_{k-1}) = \mu_{t_1,\dots,t_k}(A_1 \times \dots \times A_{k-1} \times \mathbb{R})$$

$A_1 \times A_2 = \{ (x_1, x_2) \mid x_1 \in A_1,\ x_2 \in A_2 \}$

$$A_1 \times A_2 \times \ldots \times A_{k-1} \times \mathbb{R} = \left\{ (x_1, \ldots, x_k) \in \mathbb{R}^k \;\middle|\; \begin{array}{l} x_1 \in A_1 \\ \vdots \\ x_{k-1} \in A_{k-1} \\ x_k \in \mathbb{R} \end{array} \right\}$$

Thm If $\mu_{t_1, \ldots, t_k}$ is a system of distributions satisfying (1) & (2) then $\exists$ some prob. space $(\Omega, \mathcal{U}, \mathbb{P})$, a stochastic process $(X_t)_{t \in T}$ having $\mu_{t_1, \ldots, t_k}$ as its finite dimensional distributions

---

$$\mu : \underset{k \geq 1}{(t_1, \ldots, t_k) \in T^k} \longrightarrow \mu_{t_1, \ldots, t_k} \in \mathcal{S}(\mathbb{R}^k)$$

Brownian Motion as a Gaussian Process

Def For $T = \mathbb{R}_+$ and $\Gamma(s,t) = \min(s,t)$ the centered $GP(0,\Gamma)$ noted $(B_t, t \in \mathbb{R}_+)$ is called a Brownian Motion

Ex Assume that X is a BM

- $X_0 = 0 \qquad (\sim \mathcal{N}(0,0))$
- $X_t \sim \mathcal{N}(0,t)$

$X_t - X_s \sim \mathcal{N}(0, \; t-s \;)$

$$t-s = \mathbb{E}[(X_t - X_s)^2] = \min(t,t) + \min(s,s) - 2\min(t,s) \qquad t > s$$

$$\mathbb{E}[X_t^2] + \mathbb{E}[X_s^2] - 2\,\mathbb{E}[X_t X_s]$$

$X \sim GP(0,\Gamma) \quad \Gamma(t,s) = \min(s,t)$

$$\Gamma(t,t) + \Gamma(s,s) - 2\Gamma(t,s)$$

$$\mathrm{Cov}(X_t - X_s, X_r) = \mathrm{Cov}(X_t, X_r) - \mathrm{Cov}(X_s, X_r)$$

$t > s > r$ $\quad = \min(t,r) - \min(s,r)$

$= 0 \quad \Rightarrow X_t - X_s$ indep X_r

$t > s > r_1, \dots, r_k \quad \Rightarrow \quad X_t - X_s$ indep from $X_{r_1}, \dots, X_{r_k}$

$\Rightarrow X_t - X_s$ and X_r are indep.

<u>Prop</u> $(X_t, t \in \mathbb{R}^+)$ is a B.M

$$\Updownarrow$$

(i) $X_0 = 0$ a.s.

(ii) $\forall 0 \leq s \leq t$

$X_t - X_s \sim \mathcal{N}(0, t-s)$

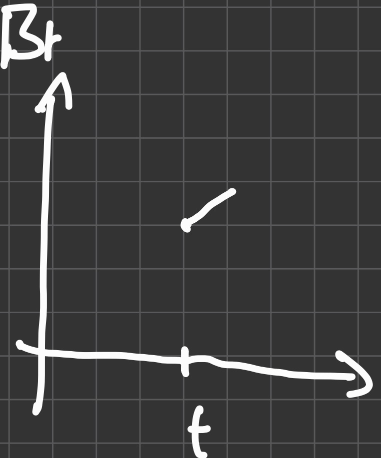
indep from $\sigma(X_r, r \leq s)$

<u>proof</u> (i)
(ii) $\Rightarrow X_t$ is a GP because

$\forall k, t_1, \dots, t_k, \lambda_1, \dots, \lambda_k$

$$\underbrace{\sum_i \lambda_i X_{t_i}}_{\text{Gaussian}} = \underbrace{c_1 X_{t_1} + \sum_{i \geq 2} c_i \underbrace{(X_{t_{i+1}} - X_{t_i})}_{\text{indep Gaussian}}}_{\text{Gaussian}}$$

Ex



$$\frac{X_t - X_s}{t-s} \sim \frac{\mathcal{N}(0, t-s)}{t-s}$$

$$t \downarrow s$$

Cor $(X_t, t \in \mathbb{R}_+)$ is a BM.

$$\Updownarrow$$

$X_0 = 0$ a.s.

and $\forall\ t_0 = 0 < t_1 < \dots < t_n$

the r.v. $X_{t_i} - X_{t_{i-1}}$ are indep

and $X_{t_i} - X_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$

$$f(y_1, \dots, y_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(t_1(t_2 - t_1) \cdots (t_n - t_{n-1}))^{1/2}}$$

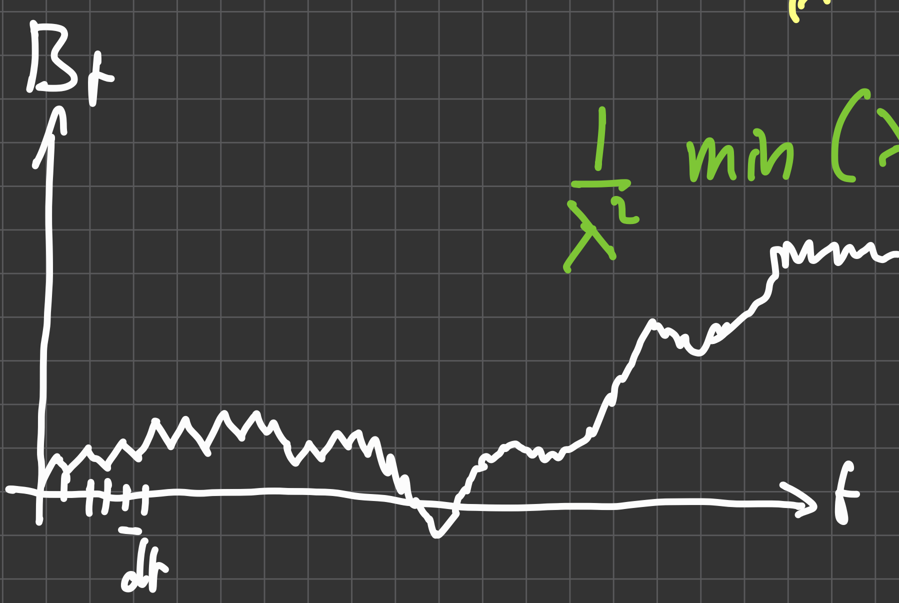
$$\times \exp\left(- \sum_{i=1}^{n} \frac{(y_i - y_{i-1})^2}{2(t_i - t_{i-1})} \right)$$

<u>Ex</u> Let B be a BM.

(i) $-B$ is a BM

(ii) $\forall \lambda > 0$, $B^\lambda_t = \frac{1}{\lambda} B_{\lambda^2 t}$ is a B.M.

$$\mathrm{Cov}(B^\lambda_t, B^\lambda_s) = \frac{1}{\lambda^2} \mathrm{Cov}(B_{\lambda^2 t}, B_{\lambda^2 s}) = \frac{1}{\lambda^2} \min(\lambda^2 t, \lambda^2 s) = \cancel{\frac{\lambda^2}{\lambda^2}} \min(t, s)$$



$C^{1/2}(\mathbb{R}_+)$

$$B_{t+dt} - B_t = dB_t \sim \mathcal{N}(0, dt)$$

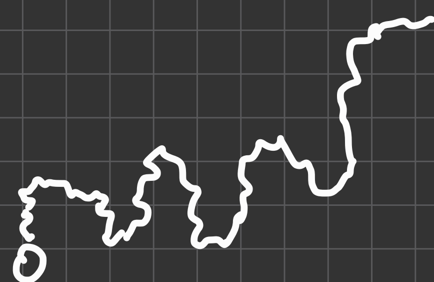
$$\mathbb{E}[B_t^2] = t$$

$$(\mathbb{E}[B_t^2])^{1/2} \sim \sqrt{t}$$

Fick - law

diffusive process

Brown:



$$B_t = (B_t^1, B_t^2)$$

↳ 2 dim. Brownian Motion

Wiener process

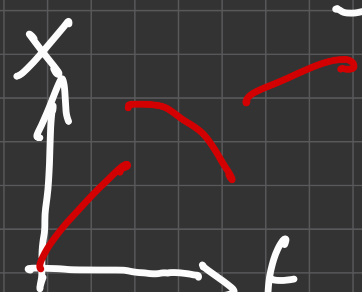
Ex $f: \mathbb{R}_+ \longrightarrow \mathbb{R}$

$$\Gamma(s,t) = f(s)\, f(t)$$

$$X \sim GP(0, \Gamma) \qquad Z \sim \mathcal{N}(0,1)$$

$$X_t = f(t)\, Z$$

$$\mathbb{E}[X_t X_s] = f(t) f(s)\, \mathbb{E}[Z^2]$$
$$= f(t) f(s) = \Gamma(s,t)$$



$X(t)$: Stochastic Process

$\sigma(X(r) \mid 0 \le r \le s)$

history of X up to time s

X is a Markov Process $\Leftrightarrow$ $\quad t > s > 0$

$\forall B \subset \mathbb{R}^n$, open

$$\mathbb{P}[X_t \in B \mid \sigma(X_r \mid 0 \le r \le s)] = \mathbb{P}[X_t \in B \mid X_s]$$

A B.M. is a Markov Process

$t > s$ $\qquad B_t = \underbrace{B_t - B_s}_{\text{indep from } \sigma(B_r, r \leq s)} + B_s$

$$\mathbb{P}[B_t \in A \mid B_s, \sigma(B_r, r \leq s)]$$
$$= \mathbb{P}[B_t \in A \mid B_s]$$

Markov Prop

Wiener process

$W_t = (W_t^1, \dots, W_t^n)$

W_t^i: iid 1d BM.

W is a Markov Process

d-dim B.M.

Gaussian Process Regression

$$(\xi(x))_{x \in \mathcal{X}} \sim GP(0, K)$$

$$\mathbb{E}[\xi(x)] = 0 \quad \forall x \in \mathcal{X}$$

$$\mathbb{E}[\xi(x)\,\xi(x')] = K(x,x') \quad \forall x, x' \in \mathcal{X}$$

$\mathcal{X}$ arbitrary

$$\begin{aligned} \xi : \mathcal{X} &\longrightarrow \text{Gaussian space} \\ x &\longrightarrow \xi(x) \sim \mathcal{N}(0, K(x,x)) \end{aligned}$$

Let $X := (X_1, \dots, X_N) \in \mathcal{X}^N$

Write $\xi(X) := (\xi(X_1), \dots, \xi(X_N))$

<u>Ex</u> $\xi(X) \sim \mathcal{N}(0, K(X,X))$

$K(X,X)$: $N \times N$ matrix with entries $K(X_i, X_j)$

Def Call K non-degenerate iif

$K(X,X)$ is invertible $\forall N,\ X_1, \dots, X_N /$ $X_i \neq X_j$ for $i \neq j$

Ex $\mathcal{X} = \mathbb{R}^d \iff K(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$
non-degenerate

Assume $X_i \neq X_j$ for $i \neq j$
K non degenerate

Thm Let $Y \in \mathbb{R}^N$. Then $\forall x \in \mathcal{X}$

(1) $\mathbb{E}[\xi(x) \mid \xi(X) = Y] = m(x)$
$= K(x,X)\, K(X,X)^{-1} Y$

$K(x,X)$: N-vector with entries
$K(x, X_i)$

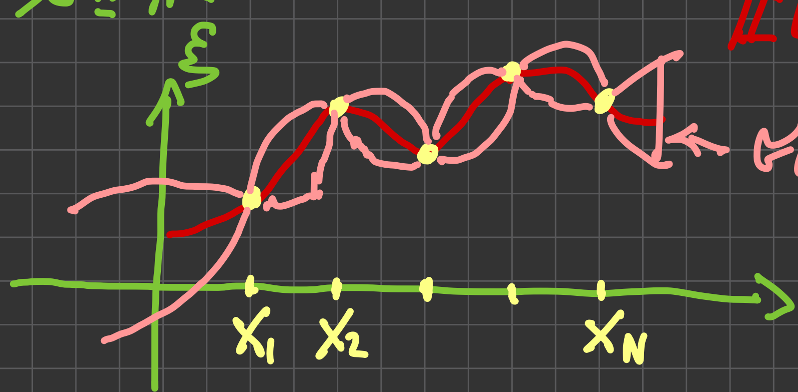
(2) $\mathrm{Var}[\xi(x) \mid \xi(X) = Y] = \sigma^2(x)$

$$= K(x,x) - K(x,X)K(X,X)^{-1}K(X,x)$$

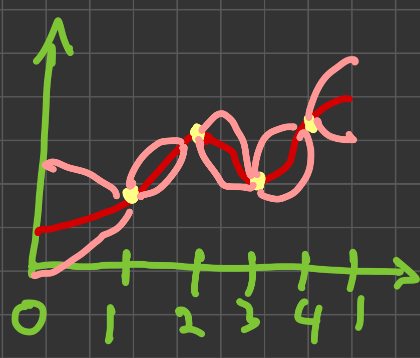
$\downarrow$ does not depend on Y

Rk $\mathcal{X} = \mathbb{R}$

$\mathbb{E}[\xi(x) \mid \xi(X) = Y]$



$\mathcal{X} = [0,1]$



$\sigma(x)$: Acquisition function

(3) $\xi(x) \mid \xi(X) = Y \sim \mathcal{N}(m(x), \sigma^2(x))$

$$(\xi(x))_{x\in\mathcal{X}} \mid \xi(X) = Y \sim GP(m(x), \sigma)$$

$$G(x,x') = K(x,x') - K(x,X)K(X,X)^{-1} K(X,x')$$

<u>proof</u>

$$(\xi(x), \xi(X_1), \ldots, \xi(X_N))$$

Gaussian vector on $\mathbb{R}^{N+1}$

Get (1) + (2) from previous thm on cond. Gaussian vectors

$$Z = (\xi(x), \xi(X_1), \ldots, \xi(X_N))$$

Z is a centered Gaussian vector

$$Z_1 \mid Z_2, \ldots, Z_{N+1}$$

<u>proof of (3)</u>

$$x_1, \ldots, x_m \in \mathcal{X}^m$$

$$\big(\xi(x_1), \dots, \xi(x_m), \xi(X_1), \dots, \xi(X_N)\big)$$

Gaussian vector on $\mathbb{R}^{m+N}$

$$\big(\xi(x_1), \dots, \xi(x_m)\big) \big| \xi(X)$$

Gaussian vector on $\mathbb{R}^m$

$\Rightarrow$ linear combination of its entries are Gaussian

$$\Rightarrow \big(\xi(x)\big)_{x \in \mathcal{X}} \big| \xi(X) \quad \text{is a GP}$$

we know its mean

$$\mathrm{Cov}\big(\xi(x_1), \xi(x_2) \big| \xi(X)\big)$$

$$= K(x_1, x_2) - K(x_1, X)\, K(X, X)^{-1} K(X, x_2)$$

Rk
$$\mathbb{E}[\xi(x) \mid \xi(X) = Y] = m(x)$$
$$= K(x, X)\, K(X, X)^{-1} Y$$
$$= \sum_{i=1}^{N} Y_i\, \psi_i(x)$$

$$\psi_i(x) = K(x, X)\, K(X, X)^{-1}_{\cdot, i}$$
$$= \sum_j K(x, X_j)\, (K(X, X))^{-1}_{j, i}$$

ψ_i : Optimal Recovery splines

$$\psi_i(x) = \mathbb{E}\left[\xi(x) \mid \xi(X_j) = \delta_{ij} \ \forall j\right]$$

elementary gambles / bets gamblets

Ex $f(t) = \sin(t)\, e^{t^2} - \sin(t+1)$

Approximate

$$I = \int_0^1 f(t)\, dt$$

$t_k = \frac{k}{N} \quad k \in \{0, 1, \dots, N\}$

given $f(t_k)$ try to approximate I,

Idea f deterministic $\Leftrightarrow B_t$: 1d BM

$B \sim GP(0, K)$

$K(t, t') = \min(t, t')$

$$I \simeq \mathbb{E}\left[\int_0^1 B_t\, dt \,\middle|\, B_{t_i} = f(t_i) \;\forall i\right]$$

$$= \int_0^1 \underbrace{\mathbb{E}\left[B_t \mid B_{t_i} = f(t_i) \;\forall i\right]} dt$$

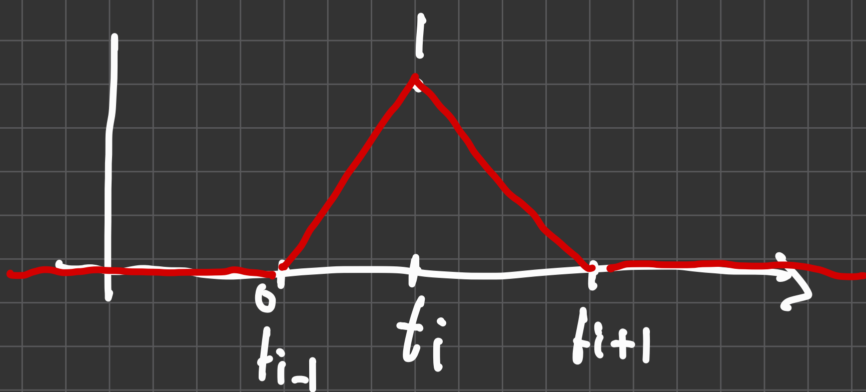
$$\sum_i f(t_i)\,\psi_i(t)$$

min (t, t_j)

$$\psi_i(t) = \sum_j K(t, t_j)\,\Theta^{-1}_{j,i}$$

$$\Theta_{i,j} = \min(t_i, t_j) = K(t_i, t_j)$$

$$\psi_i(t) = \mathbb{E}\left[B_t \,\middle|\, \begin{matrix} B_{t_i} = 1 \\ B_{t_j} = 0 \quad \forall j \neq i \end{matrix}\right]$$



$$I \approx \int_0^1 \underbrace{\sum_{i=0}^{N} f(t_i)\,\psi_i(t)}_{\text{piecewise linear interpolat of } f}$$

rediscover trapezoidal quadrature rule

$$\int_{[0,1]^2} f(t)\,dt$$

$$\downarrow$$

$$f \longleftrightarrow X_t \sim GP(0,\Gamma)$$

$$\Gamma(s,t) = \min(s_1,t_1)\min(s_2,t_2)$$

<u>1d</u> $\displaystyle\int_0^1 f(t)\,dt$

$$f \longleftrightarrow B_t$$

$$f \longleftrightarrow X_t = \int_0^t B_s\,ds \sim GP(0,\Gamma)$$

$$\begin{aligned}
\Gamma(s,t) &= \mathbb{E}[X_t, X_s] \\
&= \mathbb{E}\left[\int_0^s B_r\,dr \int_0^t B_{r'}\,dr'\right] \\
&= \mathbb{E}\left[\int_{\substack{0\le r\le s \\ 0\le r'\le t}} B_r B_{r'}\,dr\,dr'\right] \\
&= \int_{\substack{0\le r\le s \\ 0\le r'\le t}} \min(r,r')\,dr\,dr'
\end{aligned}$$

$$I = \int_0^1 f(t)\,dt = \int_0^1 \underbrace{\mathbb{E}[X_t \mid X_{t_i} = f(t_i)\ \forall i]}_{\text{cubic spline interpolant of } f}\,dt$$

$$f \longleftrightarrow \int_0^t \int_0^s B_s\,ds \iff \text{splines of order } 5$$

<u>Ex</u> — B_t: 1d BM

$GP(0, K)$

$$\mathbb{E}[B_t B_{t'}] = K(t, t') = \min(t, t')$$

$$= \mathbb{E}\Big[\int_{\substack{0 \leq r \leq s \\ 0 \leq r' \leq t}} B_r B_{r'}\,dr\,dr'\Big]$$

$$B_r B_{r'} \leq |B_r B_{r'}|$$

$$\mathbb{E}\Big[\int_{\substack{0 \leq r \leq s \\ 0 \leq r' \leq t}} |B_r B_{r'}|\,dr\,dr'\Big] < \infty$$

$$\int_{\substack{0 \le r \le s \\ 0 \le r' \le t}} \mathbb{E}[|B_r B_{r'}|] \, dr \, dr' \overset{?}{<} \infty$$

$$\overset{\text{``}}{\le} \underbrace{\mathbb{E}[B_r^2]^{1/2}}_{\sqrt{r}} \underbrace{\mathbb{E}[B_{r'}^2]^{1/2}}_{\sqrt{r'}}$$

$$\le \int_{\substack{0 \le r \le s \\ 0 \le r' \le t}} \sqrt{r}\sqrt{r'} \, dr \, dr' < \infty$$

Ex $I = \int_{[0,1]^{50}} f(x) \, dx$

$f : \mathbb{R}^{50} \longrightarrow \mathbb{R}$

Approximate I

Sol $X_1, \dots, X_N$ iid $\sim \text{Uniform}([0,1]^{50})$

$f \leftrightarrow \xi_N \; \mathcal{N}(0, K)$

$$I \simeq \mathbb{E}\left[\int_{[0,1]^{50}} \xi(x)\,dx \,\middle|\, \xi(X) = f(X)\right]$$

$\rightarrow$ quadrature rule in dim 50

$$I \simeq \int_{[0,1]^{50}} \sum_{i=1}^{N} f(X_i)\, \psi_i(x)$$

$$\psi_i(x) = K(x, X)\, K(X,X)^{-1}_{:,i}$$

$$\simeq \sum_{i=1}^{N} f(X_i)\, c_i$$

$$c_i = \int_{[0,1]^{50}} \psi_i(x)\, dx$$

$\downarrow$ quadrature coefficients can be pre-computed

Bayesian Quadrature Rule

Ex $\xi_1 \sim GP(0, K_1)$

$\xi_2 \sim GP(0, K_2)$

ξ_1, ξ_2 indep

$$\xi_1 + \xi_2 \sim GP(0, K)$$

$$K = K_1 + K_2$$

What is

$\xi_1 | \xi_1 + \xi_2$?

Ex $X = (X_1, \dots, X_N) \in \mathcal{X}^N$

$$\xi_1(X) \,\Big|\, (\xi_1 + \xi_2)(X) = Y$$

Sol $\left(\xi_1(X), \xi_1(X) + \xi_2(X) \right)$

$$\sim \mathcal{N}(0, C)$$

$$C = \begin{pmatrix} K_1(X,X) & K_1(X,X) \\ K_1(X,X) & K(X,X) \end{pmatrix}$$

$$\xi_1(X) \,\big|\, \xi_1(X) + \xi_2(X) = Y$$

$$\sim \mathcal{N}(m, C')$$

$$m = K_1(X,X)\, K(X,X)^{-1} Y$$

$$C' = K_1(X,X) - K_1(X,X)\, K(X,X)^{-1} K_1(X,X)$$

<u>Ex</u> $\xi_1 \,|\, \xi_1 + \xi_2 \sim$

$$\xi_1 \,|\, \xi_1 + \xi_2 \sim GP(m, C)$$

$m(x) = K_1(x,\cdot)K(\cdot,\cdot)^{-1}(\xi_1+\xi_2)$

$$\left.\begin{aligned} C = K_1 - \underbrace{K_1K^{-1}K_1}_{\text{operator}} \end{aligned}\right) \rightarrow \text{What is this}$$

Need to define covariance operators

Later $\Updownarrow$

Gaussian fields

Gaussian Measures

Wiener integral

B_t: 1d BM

$f \in C([0,1])$

What is $\int_0^1 f(s)\, dB_s$?

$t_i = \frac{i}{n}$

$$\int_0^1 f(s)\, ds \simeq \sum_i f(t_i)(t_{i+1} - t_i)$$

take limit $n \to \infty$

$$\int_0^1 f(s)\, dB_s \simeq \underbrace{\sum_{i=0}^{n-1} f(t_i)(B_{t_{i+1}} - B_{t_i})}_{I_n} \xrightarrow{n\to\infty} ?$$

$$I_n \sim \mathcal{N}\left(0, \sum_{i=0}^{n-1} f^2(t_i)(t_{i+1}-t_i)\right)$$

$$\mathbb{E}\left[\left(\sum_{i=0}^{n-1} f(t_i)(B_{t_{i+1}} - B_i)\right)^2\right]$$

$$= \sum_{i,j} f(t_i) f(t_j) \underbrace{\mathbb{E}\left[(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_j)\right]}_{\substack{=0 \quad i \neq j \\ = t_{i+1} - t_i \quad i = j}}$$

$$\mathbb{E}\left[(B_{t_{i+1}} - B_{t_i})^2\right] = t_{i+1} - t_i$$

$$I_n \sim \mathcal{N}\left(0, \sum_{i=0}^{n-1} f^2(t_i)(t_{i+1}-t_i)\right)$$

$$\Big\downarrow n \to \infty \qquad\qquad \Big\downarrow n \to \infty$$

$$I \sim \mathcal{N}\left(0, \int_0^1 f^2(t)\,dt\right)$$

$$\int_0^1 f(s)\,dB_s \sim \mathcal{N}\left(0, \int_0^1 f^2(t)\,dt\right)$$

$$\mathbb{L}^2(\mathcal{X},\mu) = \{ f: \mathcal{X} \to \mathbb{R} \mid \int f^2(x)\mu(dx) < \infty \}$$

Gaussian space

$$G: \mathbb{L}^2(\mathcal{X},\mu) \longrightarrow H$$

$$f \longrightarrow G(f) \sim \mathcal{N}(0, \|f\|^2_{\mathbb{L}^2(\mathcal{X},\mu)})$$

$g \in C([0,1])$

What is

$$\left(\int_0^1 f(s)\,dB_s , \int_0^1 g(s)\,dB_s \right) \quad ?$$

$\sim \mathcal{N}(0, C)$

$$\approx \left(\sum_{i=0}^{n-1} f(t_i)(B_{t_{i+1}} - B_{t_i}), \sum_{i=0}^{n-1} g(t_i)(B_{t_{i+1}} - B_{t_i}) \right)$$

$\sim \mathcal{N}(0, C_n)$

$$C_n = \begin{pmatrix} \sum_{i=0}^{n-1} f^2(t_i)(t_{i+1}-t_i) & \text{''} \\ \sum_{i=0}^{n-1} f(t_i)\,g(t_i)(t_{i+1}-t_i) & \sum_{i=0}^{n-1} g^2(t_i)(t_{i+1}-t_i) \end{pmatrix}$$

$$\Big\downarrow{\scriptstyle n\to\infty}$$

$$C = \begin{pmatrix} \int_0^1 f^2(t)\,dt & \text{''} \\ \int_0^1 f(t)\,g(t)\,dt & \int_0^1 g^2(t)\,dt \end{pmatrix}$$

Ex Applications of the Wiener Integral

Ex B_t 1d BM

$$\int_0^1 f(s)\,dB_s \sim \mathcal{N}\left(0, \int_0^1 f^2(s)\,ds\right)$$

$$f(s) = \begin{cases} s & s \le t \\ 0 & s > t \end{cases}$$

$$\int_0^t s\,dB_s \sim \mathcal{N}\Big(0, \underbrace{\int_0^t s^2\,ds}_{= t^3/3}\Big)$$

Ex $$\Big(\underbrace{B_t}_{= \int_0^t dB_s}, \int_0^t \cos(s)\,dB_s\Big) \sim \mathcal{N}(0, C)$$

$$C = \begin{pmatrix} t & \int_0^t \cos(s)\,ds \\ \int_0^t \cos(s)\,ds & \int_0^t \cos^2(s)\,ds \end{pmatrix}$$

Ex Estimating the response function of of a neuron

$x(t) \to$ neuron $\to y(t)$

$$y(t) = \int_0^t K(t-s)\, dx(s) + Z_t$$

$s \to K(s)$: linear response kernel

$t \to Z_t$: stochastic process whose dist. is unknown and indep. of $t \to x(t)$

Sol $$y_t(B,Z) = \int_0^t K(t-s)\, dB_s + Z_t$$

$$f \in \mathbb{L}^2([0,t], dx)$$

$$\mathbb{E}\left[y_t(B,Z) \int_0^t f(s)\, dB_s \right]$$

$$= \int_0^t f(s) K(t-s)\, ds$$

$$\mathbb{E}\left[y_t(B,Z) \int_0^t f(s)\, dB_s \right]$$

$$= \mathbb{E}\left[\int_0^t K(t-s)\, dB_s \int_0^t f(s)\, dB_s \right]$$

$$+ \underbrace{\mathbb{E}\left[Z_t \int_0^t f(s)\, dB_s \right]}_{= 0}$$

$$= \int_0^t K(t-s) f(s)\, ds$$

$$\| $$

$$\mathbb{E}\left[\int_0^t K(t-s)\, dB_s \int_0^t f(s)\, dB_s \right]$$

Generalization

$(E, \mathcal{E})$: measurable space

μ: σ-finite measure $(E, \mathcal{E})$

$\Updownarrow$

$\exists\ (A_i)_{1 \leq i \leq \infty}$ s.t. $E = \bigcup_i A_i$ and
$\forall i,\ \mu(A_i) < \infty$

<u>Ex</u> $E = \mathbb{R}^d$
$\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$ (Borel subsets of $\mathbb{R}^d$)
$\mu(dx) = dx$ Lebesgue measure

<u>Def</u>
A Gaussian measure of intensity μ is an isometry (linear map preserving the inner product) from $\mathbb{L}^2(E, \mathcal{E}, \mu)$ to a Gaussian space

$$G : \mathbb{L}^2(E, \mathcal{E}, \mu) \longrightarrow H$$
$$f \longrightarrow G(f)$$

$\forall \lambda \in \mathbb{R},\ f, g \in \mathbb{L}^2(E, \mathcal{E}, \mu)$

$$G(\lambda f + g) = \lambda G(f) + G(g)$$

$$\mathbb{E}[G(f)G(g)] = \langle f, g \rangle_{\mathbb{L}^2(E, \mathcal{E}, \mu)}$$

$$\mathbb{L}^2(E, \mathcal{E}, \mu) := \left\{ f : E \to \mathbb{R} \;\middle|\; \int_E f^2(x)\mu(dx) < \infty \right\}$$

<u>Ex</u> $f \in \mathbb{L}^2(E, \mathcal{E}, \mu)$

$$G(f) \sim \mathcal{N}(0, \underbrace{\mathbb{E}[G(f)^2]}_{= \|f\|^2_{\mathbb{L}^2(E, \mathcal{E}, \mu)}})$$

$A \in \mathcal{E}$, $\mu(A) < \infty$ $\quad f = \mathbb{1}_A$

$$G(A) := G(\mathbb{1}_A) \sim \mathcal{N}(0, \mu(A))$$

<u>Ex</u> $E = [0, 1]$

$\mathcal{E} = \mathcal{B}([0, 1])$

$\mu(dx) = dx$

$$\mu([a,b]) = b-a \qquad \|f\|^2_{L^2(E,\mathcal{E},\mu)}$$

$$f : [0,1] \to \mathbb{R} \mid \int_0^1 f^2(x)\,dx < \infty$$

$$G_\mu(f) \sim \mathcal{N}\left(0, \int_0^1 f^2(x)\,dx\right)$$

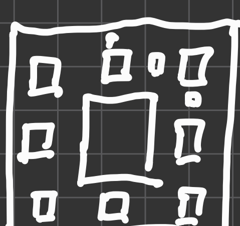
$$\nu(dx) = \cos^2(\pi x)\,dx$$

$$\nu([a,b]) = \int_a^b \cos^2(\pi x)\,dx$$

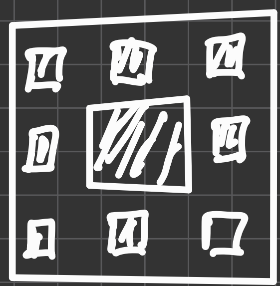
$$f : [0,1] \to \mathbb{R} \mid \int_0^1 f^2(x)\cos^2(\pi x)\,dx < \infty$$

$$G_\nu(f) = \mathcal{N}\Big(0, \int_0^1 \underbrace{f^2(x)\cos^2(\pi x)\,dx}_{= \|f\|^2_{L^2(E,\mathcal{E},\nu)}}\Big)$$

<u>Ex</u> $E = [0,1]^2$



μ : measure on Sierpinski carpet



$f : [0,1]^2 \longrightarrow \mathbb{R}$

$$G_\mu(f) = \mathcal{N}\left(0, \int_S f^2(x)\, \mu(dx)\right)$$

<u>Ex</u> $(E, \mathcal{E}, \mu)$

$A_1, \dots, A_n \in \mathcal{E}$ $\quad \mu(A_i) < \infty$

$A_i \cap A_j = \emptyset$ for $i \neq j$

$$(G(A_1), \dots, G(A_n)) \sim \mathcal{N}(0, C)$$

$G(A_1) = G(\mathbb{1}_{A_1})$

$C_{i,j} = 0 \quad i \neq j$

$C_{i,i} = \mu(A_i)$

$G(A_i) \in H$ (Gaussian space)

$\mathbb{E}[G(A_i)G(A_j)] = \int \mathbb{1}_{A_i} \mathbb{1}_{A_j} \mu(dx) = 0$

Existence and construction of Gaussian Measures

<u>Thm</u> If $(E, \mathcal{E})$ is a measurable space and μ is a σ-finite measure on $(E, \mathcal{E})$ such that $\mathbb{L}^2(E, \mathcal{E}, \mu)$ is separable (i.e. it contains a dense countable subset, e.g. $E = \mathbb{R}_+$, $\mu(dx) = dx$), then there exists a Gaussian measure of intensity μ on $(E, \mathcal{E})$

<u>proof</u> $\mathbb{L}^2(E, \mathcal{E}, \mu)$ is separable

$\Rightarrow \exists\, \varphi_n$: orthonormal basis of $\mathbb{L}^2(E, \mathcal{E}, \mu)$

$\varphi_n \in \mathbb{L}^2(E, \mathcal{E}, \mu)$

$$\langle \varphi_n, \varphi_m \rangle_{\mathbb{L}^2} = \delta_{n,m} = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

$\forall f \in \mathbb{L}^2(E, \mathcal{E}, \mu)$

$$f = \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle_{\mathbb{L}^2} \varphi_n$$

ξ_n : iid $\mathcal{N}(0,1)$ r.v

Define

$$G(f) := \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle_{\mathbb{L}^2} \xi_n$$

$G(f)$ is Gaussian

$$\mathbb{E}[G(f)] = 0$$

$\forall f, g \in \mathbb{L}^2(E, \mathcal{E}, \mu)$

$$g = \sum_{n=1}^{\infty} \langle g, \varphi_n \rangle_{\mathbb{L}^2} \varphi_n$$

$\mathbb{E}[G(f)G(g)] =$

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \langle f, \varphi_n \rangle \xi_n \sum_{m=1}^{\infty} \langle g, \varphi_m \rangle \xi_m\right]$$

$$= \sum_{n,m}^{\infty} \langle f, \varphi_n \rangle \langle g, \varphi_m \rangle \underbrace{\mathbb{E}[\xi_n \xi_m]}_{\delta_{n,m}}$$

$$= \sum_{n} \langle f, \varphi_n \rangle \langle g, \varphi_n \rangle = \langle f, g \rangle_{\mathbb{L}^2}$$

$$\|f\|_{\mathbb{L}^2}^2 = \sum_n \langle f, \varphi_n \rangle^2$$

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \langle f, \varphi_n \rangle \xi_n \sum_{m=1}^{\infty} \langle g, \varphi_m \rangle \xi_m\right]$$

$$\|$$

$$\mathbb{E}\left[\sum_{n,m=1}^{\infty} \langle f, \varphi_n \rangle \langle g, \varphi_m \rangle \xi_n \xi_m\right]$$

$$\mathbb{E}\left[\sum_{n,m=1}^{\infty} \langle f, \varphi_n \rangle \langle g, \varphi_m \rangle |\xi_n| |\xi_m|\right] < \infty$$

Rk $f = \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle_{\mathbb{L}^2} \varphi_n$ $\quad \xi_n \sim N(0,1)$ iid

$$G(f) = \sum_{n=1}^{\infty} \langle f, \varphi_n \rangle_{\mathbb{L}^2} \xi_n$$

Define (formally)

$$\xi = \sum_{n=1}^{\infty} \xi_n \varphi_n$$

ξ is not a random element of $\mathbb{L}^2(E, \mathcal{E}, \mu)$

$\|\xi\|_{\mathbb{L}^2}^2 = \sum_{n=1}^{\infty} \xi_n^2$ is a.s. infinite

However we formally have

$$\langle \xi, f \rangle_{\mathbb{L}^2(E,\mathcal{E},\mu)} = \sum_{n=1}^{\infty} \xi_n \langle \varphi_n, f \rangle_{\mathbb{L}^2} = G(f)$$

$\forall m$, $\xi^m := \sum_{n=1}^{m} \xi_n \varphi_n$ is a random element of $\mathbb{L}^2(E, \mathcal{E}, \mu)$

$$G(f) = \lim_{m \to \infty} \langle \xi^m, f \rangle_{\mathbb{L}^2(E,\mathcal{E},\mu)}$$

G: is not a prob. measure on $\mathbb{L}^2(E, \mathcal{E}, \mu)$ but a cylinder measure (limit of prob. measures)

In the sense that the cylinder measure associated with ξ can be defined as a weak limit of Gaussian measures on
$$\mathbb{L}^2(E, \mathcal{E}, \mu)$$

Gaussian vectors $\quad X$ r.el. of $\mathbb{R}^d$

$(\mathbb{L}^2(E, \mathcal{E}, \mu))$

$$G(f) \sim \mathcal{N}(0, \|f\|_{L^2}^2)$$

$T = \mathbb{L}^2(E, \mathcal{E}, \mu)$

$\Gamma(f, g) = \langle f, g \rangle_{\mathbb{L}^2}$

$G(f)$: GP with mean zero and cov. function Γ

Ex: $E = [0,1]$ $\quad \mu(dx) = dx$

$$\mathbb{L}^2(E, \mathcal{E}, \mu) = \mathbb{L}^2([0,1], \mathcal{B}([0,1]), dx)$$

$$\varphi_n : \{1, \sqrt{2}\sin(2\pi n t), \sqrt{2}\cos(2\pi n t)\}$$

$$\xi^m = \underset{\sim \mathcal{N}(0,1)}{\xi_0} + \sum_{n=1}^{m} \sqrt{2}\sin(2\pi n t)\, \xi_{n,s} + \sum_{n=1}^{m} \sqrt{2}\cos(2\pi n t)\, \xi_{n,c}$$

$\xi_{n,s} \sim \mathcal{N}(0,1)$, $\xi_{n,c} \sim \mathcal{N}(0,1)$

ξ^m: rand. el. of $\mathbb{L}^2([0,1], dx)$

ξ^∞: is not a random element of $\mathbb{L}^2([0,1], dx)$

$$\|\xi^\infty\|_{L^2} = \xi_0^2 + \sum_{n=1}^{\infty} \xi_{n,s}^2 + \xi_{n,c}^2 = \infty$$

$$\|\xi^\infty\|_{H^{-1}} \simeq 1 + \sum_{n=1}^{\infty} \frac{\xi_{n,s}^2}{n^2} + \frac{\xi_{n,c}^2}{n^2} < \infty$$

H^{-1} ↪ Sobolev space

ξ^∞ rand. element of $H^{-1}([0,1], dx)$

$$H^1([0,1], dx) =$$
$$= \{ f : [0,1] \to \mathbb{R} \,/\, \int_0^1 |\nabla f|^2(x)\, dx < \infty \}$$

$$H^{-1}([0,1], dx)$$
$$= \{ f : [0,1] \to \mathbb{R} \,/\, \sup_g \frac{\left(\int_0^1 f g\right)^2}{\int_0^1 |\nabla g|^2 dx} < \infty \}$$

$$\xi^m = \sum_{n=1}^{m} \varphi_n \xi_n \in \mathbb{L}^2(E, \mathcal{E}, \mu)$$

$$V = \{ f : \sum_n \frac{\langle f, \varphi_n \rangle^2}{n^2} < \infty \}$$

$$\mathbb{L}^2(E, \mathcal{E}, \mu) \subset V$$

$$\xi^m \longrightarrow \xi^\infty \in V$$

$$\mathbb{P}[\xi^\infty \in \mathbb{L}^2(E, \mathcal{E}, \mu)] = 0$$
$$\mathbb{P}[\xi^\infty \in V] = 1$$

$\underline{\text{Ex}}$: $E = [0,1]$ $\quad \mu(dx) = dx$

$$\mathbb{L}^2(E, \mathcal{E}, \mu) = \mathbb{L}^2([0,1], \mathcal{B}([0,1]), dx)$$

$$\varphi_n : \{ \underbrace{1}_{\varphi_0}, \underbrace{\sqrt{2}\sin(2\pi n t)}_{\varphi_{n,s}}, \underbrace{\sqrt{2}\cos(2\pi n t)}_{\varphi_{n,c}} \}$$

$$f = \langle f, \varphi_0 \rangle \varphi_0 + \sum_n \langle f, \varphi_{n,s} \rangle \varphi_{n,s}$$
$$\langle f, \varphi_{n,c} \rangle \varphi_{n,c}$$

$$G(f) = \langle f, \varphi_0 \rangle \underbrace{\xi_0}_{\mathcal{N}(0,1)} + \sum_n \langle f, \varphi_{n,s} \rangle \overbrace{\xi_{n,s}}^{\mathcal{N}(0,1)}$$
$$+ \langle f, \varphi_{n,c} \rangle \underbrace{\xi_{n,c}}_{\mathcal{N}(0,1)}$$

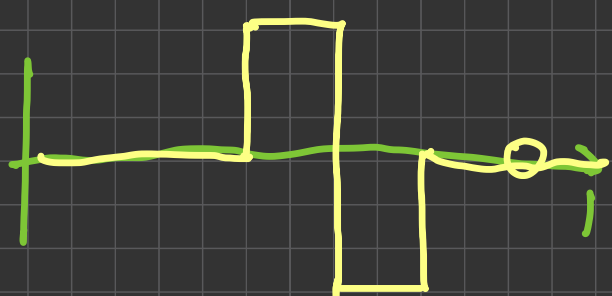
$$G(f) \sim \mathcal{N}\left(0, \int_0^1 f^2(x)\, dx\right)$$

Haar functions

$$\{\varphi_0, \varphi_{j,n} \mid j = 1, \dots, 2^{n-1}, n = 1, 2, \dots)$$

$\varphi_0(t) \equiv 1$ and for $k = 2j-1$

$$\varphi_{j,n} = \begin{cases} 2^{(n-1)/2} & \frac{k-1}{2^n} \leq t < \frac{k}{2^n} \\ -2^{(n-1)/2} & \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \\ 0 & \text{else.} \end{cases}$$



$$f = \langle f, \varphi_0 \rangle \varphi_0 + \sum_{j,n} \langle f, \varphi_{j,n} \rangle \varphi_{j,n}$$

$$G(f) = \langle f, \varphi_0 \rangle \underset{\mathcal{N}(0,1)}{\underset{\sim}{\xi_0}} + \sum_{j,n} \langle f, \varphi_{j,n} \rangle \underset{\text{iid}}{\underset{\mathcal{N}(0,1)}{\underset{\sim}{\xi_{j,n}}}}$$

$$G(f) \sim \mathcal{N}\left(0, \int_0^1 f^2(x)\,dx\right)$$

Brownian Motion from a Gaussian Measure

$(E, \mathcal{E}, \mu) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), dx)$

$(\varphi_n)_{n \in \mathbb{N}}$: orthonormal basis of
$\mathbb{L}^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), dx)$
(Fourier, Haar, Wavelets)

$$G : \mathbb{L}^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), dx) \longrightarrow H$$
$$f = \sum_n \langle f, \varphi_n \rangle \varphi_n \longrightarrow G(f) = \sum_n \langle f, \varphi_n \rangle \xi_n$$

ξ_n iid $\mathcal{N}(0,1)$

<u>Thm</u> $B_t = G(\mathbb{1}_{[0,t]})$ is a 1d B.M.

<u>Proof</u> $(B_t, t \in \mathbb{R}_+)$: Centered GP

because $\forall \lambda_1, \dots, \lambda_n \in \mathbb{R}$

$$\sum_i \lambda_i B_{t_i} = G\left[\sum_i \lambda_i \mathbb{1}_{[0,t_i]}\right] \text{ is Gaussian}$$

$$\Gamma(s,t) = \mathbb{E}[B_s B_t]$$

$$= \int_{\mathbb{R}_+} \mathbb{1}_{[0,s]} \mathbb{1}_{[0,t]} \, dx$$

$$= \min(s,t)$$

$\Rightarrow B_t$ is a ld BM.

Wiener integral with respect to a Brownian Motion

<u>Def</u> If B is a B.M.

and G the associated Gaussian Measure

$(B_t = G(\mathbb{1}_{[0,t]}))$

we write

$$\int_0^\infty f(s)\, dB_s := G(f)$$

for $f \mid \int_0^\infty f^2(x)\, dx < \infty$

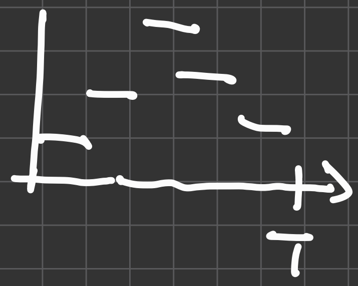
$$\int_0^t f(s)\, dB_s := G(f \mathbb{1}_{[0,t]})$$

The mapping $f \longrightarrow \int_0^\infty f(s)\, dB_s$ is called the Wiener integral w.r. to the BM. B.

<u>Rk</u> If $u < v$

$$\int_u^v dB_s = G(]u, v])$$
$$= B_v - B_u$$

<u>Def</u> A function $f \in L^2(0,T)$ $\left(\int_0^T f^2(x)\, dx < \infty\right)$ is called a step function if $\exists$ a partition $0 = t_0 < t_1 < \dots < t_m = T$ s.t.
$f(t) = f_k$ for $t_k \leq t < t_{k+1}$ $(k = 0, \dots, m-1)$



Prop If $f \in L^2(0,T)$ is a stepfunction then

$$\int_0^T f(s)\, dB_s = \sum_{k=0}^{m-1} f_k \left(B_{t_{k+1}} - B_{t_k}\right)$$

Lemma If $f \in L^2(0,T)$ then $\exists$ a sequence of stepfunctions $f_n \in L^2(0,T)$ |

$$\int_0^T |f - f_n|^2(s)\, ds \xrightarrow[n\to\infty]{} 0$$

Moreover by isometry

$$\int_0^T f_n(s)\, dB_s \xrightarrow[n\to\infty]{\mathbb{L}^2(\Omega)} \int_0^T f(s)\, dB_s$$

Lemma $\forall a, b \in \mathbb{R}$, $f, g \in L^2(0,T)$

$$\int_0^T (af + bg)\, dB_s = a\int_0^T f(s)\, dB_s + b\int_0^T g(s)\, dB_s$$

$$\mathbb{E}\left[\int_0^T f(s)\, dB_s\right] = 0$$

$$\mathbb{E}\left[\int_0^T f(s)\, dB_s \int_0^T g(s)\, dB_s\right] = \int_0^T f(s) g(s)\, ds$$

$$\int_0^T f(s)\, dB_s \sim \mathcal{N}\left(0, \int_0^T f^2(s)\, ds\right)$$

Wiener integral

(Itô integral: (f stochastic process) beyond ACM 118)

Stationary processes.

Def A stochastic process $(X_t)_{t\in\mathbb{R}}$ is stationary if

$$\mathrm{dist}\left((X_t)_{t\in\mathbb{R}}\right) = \mathrm{dist}\left((X_{t+s})_{t\in\mathbb{R}}\right) \quad \forall s \in \mathbb{R}$$

X_t: iid $\mathcal{N}(0,1)$

$$\Gamma(t,s) = \begin{cases} 1 & \text{if } t=s \\ 0 & \cdot \quad t\neq s \end{cases}$$

$(X_t)_{t\in\mathbb{R}}$. $GP(0,\Gamma)$

$$\Gamma(s,t) = \Gamma(t-s)$$

X stationary

$$\Gamma(s,t) = \exp\left(-\frac{(t-s)^2}{2\sigma^2}\right)$$

B_t^1, B_t^2: indep 1d BM

$$B_t = \begin{cases} B_t^1 & t \geq 0 \\ B_{-t}^2 & t < 0 \end{cases}$$

φ: continuous function on $\mathbb{R}$ /

$\varphi(t) \equiv 0$ for $t \notin [-1, 1]$

$$X_t = \int_{t-1}^{t+1} \varphi(s-t) \, dB_s$$

. X_t is a GP

$\mathbb{E}[X_t] = 0$

$$\Gamma(s,t) = \mathbb{E}[X_t X_s]$$

$$= \int_{\mathbb{R}} \varphi(u-t) \varphi(u-s) \, du$$

$$= \int_{\mathbb{R}} \varphi(x) \varphi(x+t-s) \, dx$$

$= \Gamma(t-s)$

cov. fct invariant under time shifts

Gaussian vectors in Euclidean Spaces

E: Euclidean space of dimension d
(E is isomorphic to $\mathbb{R}^d$)

$\langle u, v\rangle$ or $\langle u, v\rangle_E$: scalar product of E

Rk E endowed with scalar product $\langle u, v\rangle$ is also known as a finite-dimensional Hilbert space

Ex. $E = \mathbb{R}^d$ $\quad \langle u, v\rangle = u^T v \qquad u, v \in \mathbb{R}^d$

. $E = \mathbb{R}^{n\times n}$ $\quad \langle A, B\rangle = \sum_{i,j=1}^{n} A_{i,j} B_{i,j}$

. Let $\varphi_1, \ldots, \varphi_d$ be d linearly indep. elements of

$$\mathbb{L}^2((0,1), dx) = \{ f : [0,1] \to \mathbb{R} \mid \int_0^1 f^2(x)\, dx < \infty \}$$

$$E = \text{Span}\{\varphi_1, \ldots, \varphi_d\}$$

$$\langle \psi, \varphi \rangle = \int_0^1 \psi(x)\, \varphi(x)\, dx \qquad \forall \varphi, \psi \in E$$

<u>Def</u> $X : \Omega \to E$ is a Gaussian vector iff

$$\langle u, X \rangle \text{ is a Gaussian r.v. } \forall u \in E$$

<u>Ex</u> $E = \text{Span}\{\varphi_1, \ldots, \varphi_d\} = \{ \varphi = \sum_{i=1}^{d} c_i \varphi_i \mid c \in \mathbb{R}^d \}$

$$\langle \psi, \varphi \rangle = \int_0^1 \varphi(x)\, \psi(x)\, dx$$

$$X = \sum_{i=1}^{d} X_i \varphi_i \qquad \text{rand. element of } E$$

$X_1, \dots, X_d$: iid $\mathcal{N}(0,1)$

$\Rightarrow X$ is a Gaussian vector on E

$\psi \in E$

$$\langle \psi, X \rangle = \underbrace{\sum_{i=1}^{d} X_i \int_0^1 \psi(x) \varphi_i(x)}_{\text{is Gaussian}}$$

$\Downarrow$

X is a Gaussian vector on E

Only need $(X_1, \dots, X_d)$ to be a Gaussian vector of $\mathbb{R}^d$

Rk X Gaussian vector on E

$\downarrow$

$e_1, \dots, e_d$ basis of E

$(\langle e_1, X \rangle, \dots, \langle e_d, X \rangle)$

$\langle u, X \rangle \quad \forall u \in E$

$$\langle u, X \rangle \sim \mathcal{N}(\underset{\substack{\Uparrow \\ E}}{\langle u, m_x \rangle}, \; ? \;)$$

Def A mapping $B: E \times E \to \mathbb{R}$

is called a symmetric bilinear form on E

iif $\forall \lambda, u, v, w \in \mathbb{R} \times E \times E \times E$

$B(\lambda u + w, v) = \lambda B(u, v) + B(w, v)$

& $B(u, v) = B(v, u)$

Ex $E = \mathbb{R}^d \qquad B(u, v) = u^T v$

$B(u, v) = u^T A v$

A: $d \times d$ symmetric matrix

Ex $E = \text{Span}\{\varphi_1, \dots, \varphi_d\}$

$$\varphi_i : [0,1] \to \mathbb{R} \qquad \int_0^1 \varphi_i^2(x)\,dx < \infty$$

$$B(\varphi, \psi) = \int_0^1 \varphi(x)\,\psi(x)\,dx$$

$$(*) \quad B(\varphi, \psi) = \int_0^1\!\!\int_0^1 \varphi(x)\, G(x,y)\, \psi(y)\,dx\,dy$$

G: symmetric function of x & y

$(G(x,y) = G(y,x))$

Ex In previous ex all bilinear sym forms are of the form $*$

$$\psi = \sum_{i=1}^{d} u_i \varphi_i \qquad \varphi = \sum_{i=1}^{d} v_i \varphi_i$$

$$B(\varphi, \psi) = \sum_{i,j=1}^{d} u_i v_j\, B(\varphi_i, \varphi_j)$$

$$A_{i,j} = B(\varphi_i, \varphi_j)$$

A sym. matrix

$\phi_1, \dots, \phi_d$: basis of E that is bi-orthogonal to $\varphi_1, \dots, \varphi_d$

$$\langle \phi_i, \varphi_j \rangle = \delta_{i,j} \quad \forall i,j$$

$$B(\varphi_i, \varphi_j) = \int_0^1 \int_0^1 \varphi_i(x) \phi_i(x) B(\varphi_i, \varphi_j) \phi_j(y) \varphi_j(y) \, dx \, dy$$

$\Rightarrow \forall \psi, \varphi \in E$

$$B(\varphi, \psi) = \int_0^1 \int_0^1 \varphi(x) \, G(x,y) \, \psi(y) \, dx \, dy$$

$$G(x,y) = \sum_{i,j=1}^{d} \phi_i(x) \, B(\varphi_i, \varphi_j) \, \phi_j(y)$$

<u>Def</u> $Q : E \longrightarrow \mathbb{R}$ is a positive quadratic form on E, iif $\exists$ a symmetric bilinear form B on E / $Q[u] = B[u,u]$

& $Q[u] \geqq 0 \quad \forall u \in E$

<u>Rk</u> Can you recover B from Q ? YES

$$B[u,v] = \frac{1}{2}\left[Q[u+v] - Q[u] - Q[v]\right]$$

<u>Ex</u> $E = \mathbb{R}^d$, $Q[u] = u^T u$

$$Q[u] = u^T A u$$

A: symmetric positive $d \times d$ matrix

<u>Ex</u> $E = \mathrm{Span}\{\varphi_1, \dots, \varphi_d\}$ $\quad \varphi_i : [0,1] \to \mathbb{R}$

$$\int_0^1 \varphi_i^2(x)\, dx < \infty$$

$$Q[\varphi] = \int_0^1 \varphi^2_{(x)}\, dx$$

$$Q[\varphi] = \int_0^1 \int_0^1 \varphi(x)\, G(x,y)\, \varphi(y)\, dx\, dy$$

G: symmetric / $\int_0^1 \int_0^1 \varphi(x)\, G(x,y)\, \varphi(y)\, dx\, dy \geq 0$

$\downarrow$

positive kernel

If $\nabla \varphi_i \in L^2([0,1], dx)$

$$Q[\varphi] = \int_0^1 |\nabla \varphi|^2(x)\, dx$$

<u>Prop</u> If X is a Gaussian vector on E, then $\exists\, m_X \in E$ and a positive quadratic form Q_X on E / $\forall u \in E$

$$\mathbb{E}[\langle u, X \rangle] = \langle u, m_X \rangle$$

$$\mathrm{Var}[\langle u, X \rangle] = Q_X[u]$$

$$\mathbb{E}[\exp(i\langle u, X\rangle)] = \exp\left(i\langle u, m_X\rangle - \frac{1}{2} Q_X[u]\right)$$

$$\langle u, X\rangle \sim \mathcal{N}(\langle u, m_X\rangle, Q_X[u])$$

We say

$$X \sim \mathcal{N}(m_X, Q_X)$$

<u>proof</u> $(e_1, \dots, e_d)$ orthonormal basis of E

$$X = \sum_{i=1}^{d} X_i e_i$$

$X_i = \langle e_i, X\rangle$ are Gaussian

$$m_X = \sum_{i=1}^{d} \mathbb{E}[X_i] e_i$$

$$u = \sum_{i=1}^{d} u_i e_i$$

$$Q_X[u] = \mathrm{Var}[\langle u, X\rangle]$$

$$= \sum_{i,j} u_i u_j \, \mathrm{Cov}(X_i, X_j)$$

$$\langle u, X \rangle \sim \mathcal{N}\left(\langle u, m_X \rangle, \ Q_X[u]\right)$$

$$m_X^e = \begin{pmatrix} \mathbb{E}[\langle e_1, X \rangle] \\ \vdots \\ \mathbb{E}[\langle e_d, X \rangle] \end{pmatrix}$$

C^e: $d \times d$ matrix

$$C^e_{i,j} = \mathrm{Cov}(\langle e_i, X \rangle, \langle e_j, X \rangle)$$

X is defined indep from basis

m^e_X, C^e_X: depend on basis

$(f_1, \dots, f_d)$: another orthonormal basis of E

$$f_i = \sum_{j=1}^{d} A_{i,j} e_j$$

A: $d \times d$ orthonormal matrix $(A^T A = I_d)$

$$X^f = \begin{pmatrix} \langle f_1, X \rangle \\ \langle f_d, X \rangle \end{pmatrix}$$

C^f: $d \times d$ matrix

$$C^f_{i,j} = \mathrm{Cov}\left[\langle f_i, X \rangle, \langle f_j, X \rangle\right]$$

$$m^f_X = A\, m^e_X$$

$$C^f = A\, C^e A^T$$

<u>Prop</u> $(e_1, \dots, e_d)$ orthonormal basis of E

$$X = \sum_{i=1}^{d} X_i e_i$$

$$X \sim \mathcal{N}(m_X, Q_X)$$

Then the r.v. $X_1, \dots, X_d$ are indep.

iif $(\mathrm{Cov}(X_i, X_j))_{i,j}$ is diagonal

$\Leftrightarrow$ Q_X is diagonal in basis $e_1, \dots, e_d$

$$Q_X[u] = B[u,u]$$

$$B[e_i, e_j] = \lambda_i \delta_{i,j}$$

<u>Def</u> γ is a symmetric positive endomorphism on E iif γ is a linear function mapping E to E s.t. $\forall u, v \in E$

$$\langle u, \gamma[v] \rangle = \langle \gamma[u], v \rangle$$

$$\langle u, \gamma[u] \rangle \geq 0$$

<u>Rk</u> Q_X is a quadratic form iif $\exists$ a symmetric positive endomorphism associated to Q_X s.t. $\forall u \in E$

$$Q_X[u] = \langle u, \gamma_X[u] \rangle$$

The matrix of γ_X in the basis $(e_1, \dots, e_d)$ is $(\mathrm{Cov}(X_i, X_j))_{i,j}$ $\quad X_i = \langle X, e_i \rangle$

$$\langle v, \gamma_X[u] \rangle = \frac{1}{4}\left[Q_X[u+v] - Q_X[u-v]\right]$$

<u>Thm</u> It holds true that

1. If γ is a symmetric positive endomorphism on E, then there exists a centered Gaussian vector X s.t. $\gamma_X = \gamma$

2. Let X be a centered Gaussian vector $(\varepsilon_1, \dots, \varepsilon_d)$ a basis of E diagonalizing γ_X

$$\gamma_X \varepsilon_j = \lambda_j \varepsilon_j \qquad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0 = \lambda_{r+1} = \dots = \lambda_d = 0$$

Then

$$X = \sum_{j=1}^{r} Y_j \varepsilon_j$$

Y_j indep $\mathcal{N}(0, \lambda_j)$ r.v.

topological support of X is $\operatorname{span}\{\varepsilon_1, \dots, \varepsilon_r\}$

Thm Let $X \sim \mathcal{N}(m_X, Q_X)$

$v_1, \dots, v_m \in E$

$$X \,|\, \langle v_1, X\rangle, \dots, \langle v_m, X\rangle \sim \mathcal{N}(m_X^{\perp}, Q_X^{\perp})$$

$$m_X^{\perp} = \mathbb{E}[X \,|\, \langle v_1, X\rangle, \dots, \langle v_m, X\rangle]$$

$$= m_X + \sum_{i,j=1}^{m} \langle v_i, X - m_X\rangle \, \theta_{i,j}^{-1} \, \gamma_X v_j$$

$$\theta_{i,j} = \langle v_i, \gamma_X v_j \rangle$$

$$Q_X^{\perp}[u] = \min_{v \in \mathrm{Span}\{v_1, \dots, v_m\}} Q_X[u - v]$$

$$Q_X^{\perp}[u] = \langle u, \gamma_X u\rangle - \sum_{i,j} \langle u, \gamma_X v_i\rangle \, \theta_{i,j}^{-1} \, \langle u, \gamma_X v_j\rangle$$

$$= \langle u, \gamma_X^{\perp} u\rangle$$

$$\gamma_x^{\perp} = \gamma_x - \sum_{i,j} \theta_{i,j}^{-1} \; \gamma_x v_i \otimes \gamma_x v_j$$

$$\gamma_x v_i \otimes \gamma_x v_j : E \longrightarrow E$$

$$v \to \gamma_x v_i \langle v, \gamma_x v_j \rangle$$

<u>proof</u> $v \in E$

Find $\langle v, X \rangle \,/\, \langle v_1, X \rangle, \dots, \langle v_m, X \rangle$

$\langle v, X \rangle, \langle v_1, X \rangle, \dots, \langle v_m, X \rangle$

Gaussian vector of $\mathbb{R}^{m+1}$

Gaussian measures on a Hilbert space

H: separable Hilbert space

$\langle \cdot, \cdot \rangle_H$: inner product on H

We say a linear operator

$$T : H \longrightarrow H$$

is symmetric and positive w.r.t. $\langle \cdot, \cdot \rangle_H$

iif
$$\begin{cases} \langle \varphi, T\phi \rangle_H = \langle \phi, T\varphi \rangle_H & \forall \varphi, \phi \in H \\ \langle \phi, T\phi \rangle_H \geq 0 & \forall \phi \in H \end{cases}$$

Trace of T

$$tr[T] = \sum_{i=1}^{\infty} \langle Te_i, e_i \rangle_H \quad \left(\text{indep from } (e_i)_{i \in \mathbb{N}^*} \right)$$

$(e_i)_{i \in \mathbb{N}^*}$: any orthonormal basis of H

Def T is trace class $\Longleftrightarrow$ $tr[T] < \infty$

Gaussian measure

Def $\mathcal{P}(\mathcal{X})$: set of Borel prob measures on a topological space $\mathcal{X}$, and, when $\mathcal{X}$ is a Banach space, we let $\mathcal{P}_2(\mathcal{X})$ denote those with finite second moments

$$(\mathbb{E}[\|X\|^2] < \infty)$$

Def A Borel prob. measure $\mu \in \mathcal{P}(\mathcal{H})$ on the Hilbert space $\mathcal{H}$ is said to be a Gaussian measure if the random vector X of $\mathcal{H}$ defined by μ is such that

$\langle \phi, X \rangle_{\mathcal{H}}$ is a Gaussian rand. var

$$\forall \phi \in \mathcal{H}$$

Thm If $\mu \in \mathcal{P}(H)$ is a Gaussian measure

then $\mu \in \mathcal{P}_2(H)$, i.e.

$$\mathbb{E}_{X\sim\mu}[\|X\|^2] < \infty$$

Furthermore, $\exists\, m \in H$ and a linear symmetric positive trace class operator

$$T: H \to H$$

s.t. the random vector X of H defined by μ is s.t.

$$\langle \phi, X \rangle_H \sim \mathcal{N}(\langle \phi, m \rangle_H, \langle \phi, T\phi \rangle_H) \quad \forall \phi \in H$$

In particular

$$\mathbb{E}_{X\sim\mu}[\langle \phi_1, X-m \rangle_H \langle \phi_2, X-m \rangle_H] = \langle \phi_1, T\phi_2 \rangle_H \quad \forall \phi_1, \phi_2 \in H$$

We say that $\mu \in \mathcal{P}(H)$ is a Gaussian measure with mean m & covariance operator T

$$X \sim \mathcal{N}(m, T)$$

<u>Thm</u> Given $m \in H$ and a linear symmetric positive trace class operator $T: H \to H$ $\exists$ a Gaussian measure $\mu \in \mathcal{P}(H)$ with mean m and covariance operator T

<u>proof</u> Let $e_1, e_2, \dots$ be an orthonormal basis of H diagonalizing T

$$Te_i = \lambda_i e_i$$

T positive, trace class $\Rightarrow$ $\lambda_i \geq 0$

$$\sum_i \lambda_i < \infty$$

$Z_1, Z_2, \dots$ iid $\mathcal{N}(0,1)$ r.v.

$$X := m + \sum_i \sqrt{\lambda_i}\, Z_i\, e_i$$

$$X \sim \mathcal{N}(m, T)$$

T trace class $\Rightarrow$ X is a random element of H

Gaussian fields on Hilbert Space

H: separable Hilbert space

$\langle \cdot, \cdot \rangle_H$: inner product

$\| \cdot \|_H$: norm

<u>Def</u> (*) A centered Gaussian field ξ on H is an isometry mapping H to a centered Gaussian space

$$\xi : H \longrightarrow \text{Gaussian Space (centered)}$$

$$\phi \longrightarrow \xi(\phi) \sim \mathcal{N}(0, \|\phi\|_H^2)$$

$$\mathrm{Cov}(\xi(\phi), \xi(\varphi)) = \langle \phi, \varphi \rangle_H$$

<u>Thm</u> Let $\phi_1, \phi_2, \dots$ be an orthonormal basis of H

$Z_1, Z_2, \dots$ iid $\mathcal{N}(0,1)$

Then
$$\begin{aligned}
\xi : H &\longrightarrow \text{Gaussian space} \\
\phi &\longrightarrow \xi(\phi) \\
\overset{\shortparallel}{\sum_{i=1}^{\infty} c_i \phi_i} &\longrightarrow \overset{\shortparallel}{\sum_{i=1}^{\infty} c_i Z_i}
\end{aligned}$$

is a Gaussian field in the sense of Def $*$

<u>proof</u> $\mathbb{E}[\xi(\phi)^2] = \sum_i c_i^2 = \|\phi\|_H^2$

<u>Rk</u> Formally $\xi = \sum_i Z_i \phi_i$

$\phi = \sum_i c_i \phi_i$

$\xi(\phi) = \langle \xi, \phi \rangle_H$

If $\dim(H) = \infty$ then ξ is not a rand. element of H

$$\|\xi\|_H^2 = \sum_i Z_i^2 = \infty \quad \text{a.s.}$$

However

$$V = \left\{ \sum_i c_i \phi_i \,\middle|\, c_i \in \mathbb{R} \,\&\, \sum_i \frac{c_i^2}{i^2} < \infty \right\}$$

$$\|\phi\|_V^2 = \sum_i \frac{c_i^2}{i^2} \quad \text{for } \phi = \sum_i c_i \phi_i$$

then ξ is a.s. an element of V

$(V \supset H)$

$$\|\xi\|_V^2 = \sum_i \frac{Z_i^2}{i^2} < \infty \quad \text{a.s.}$$

ξ : Gaussian measure on V

<u>Ex</u> $H = L^2(\Omega)$

$\Omega \subset \mathbb{R}^d$

$$\|f\|_H^2 = \int_\Omega f^2(x)\,dx$$

ξ is not a random element of $L^2(\Omega)$

but of $H^{-s}(\Omega)$ for $s > \frac{d}{2}$

ϕ_i: eigenfunctions of $(-\Delta)$ on $H_0^1(\Omega)$

$$-\Delta \phi_i = \lambda_i \phi_i$$

Weyl's estimate: $\lambda_i \propto i^{2/d}$

$$H^{-s}(\Omega) := \left\{ \sum_k c_k \phi_k \,\middle|\, \sum_k \frac{c_k^2}{\lambda_k^s} < \infty \right\}$$

$$\Rightarrow \|\xi\|_{H^{-s}(\Omega)} < \infty \quad \text{a.s.}$$

ξ is a Gaussian measure on $H^{-s}(\Omega)$

Gaussian measures and fields on Banach spaces

Gaussian fields:

$\mathcal{B}$: Banach space (separable)

$\mathcal{B}^*$: dual space (space of linear functionals on $\mathcal{B}$)

$[\cdot,\cdot]$: dual pairing between $\mathcal{B}^*$ and $\mathcal{B}$

$\phi \in \mathcal{B}^*$, $u \in \mathcal{B}$, $[\phi, u]$

Let $T: \mathcal{B}^* \longrightarrow \mathcal{B}$

be a symmetric, positive, not necessarily injective, linear operator

. $[\varphi, T\phi] = [\phi, T\varphi] \quad \forall \varphi, \phi \in \mathcal{B}^*$

. $[\phi, T\phi] \geq 0 \quad \forall \phi \in \mathcal{B}^*$

Def ξ is a Gaussian field on B with mean $v \in B$ and covariance operator T

$$\xi \sim \mathcal{N}(v, T)$$

iif $\xi : B^* \longrightarrow$ Gaussian space s.t.

$$\xi(\phi) := [\phi, \xi] \sim \mathcal{N}([\phi, v], [\phi, T\phi]) \quad \forall \phi \in B^*$$

Ex $B = \mathbb{R}^N$

$\|x\|^2 = x^T A x$

A: $N \times N$ positive definite symmetric

$[x, y] = x^T y$

$B^* = \mathbb{R}^N$

$$\|x\|_*^2 = \sup_{y \in B} \left(\frac{[x, y]}{\|y\|} \right)^2 = x^T A^{-1} x$$

Ex D bounded domain of $\mathbb{R}^d$

$$B = \{ f : D \to \mathbb{R} \mid \int_D f^2(x)\,dx < \infty \}$$

$$\|f\|^2 = \int_D f^2(x)\,dx \qquad [\phi, f] = \int_D \phi f$$

$$B^* = B \qquad \|\cdot\|_* = \|\cdot\|$$

Ex $B = H_0^1(D)$

$$\|f\|^2 = \int_D |\nabla f|^2(x)\,dx$$

$$[\phi, f] = \int_D \phi f$$

B: closure $C_0^\infty(D)$ w.r.t. $\|\cdot\|$

$B^* = H^{-1}(D)$

$$\|\phi\|_*^2 = \int_D \phi(x)\, G(x,y)\, \phi(y)\,dx\,dy$$

G: Green's function of $(-\Delta)$

s.l. of $\begin{cases} -\Delta u = \phi & \text{on } D \\ u = 0 & \partial D \end{cases}$

is $u(x) = \int_D G(x,y)\,\phi(y)\,dy$

<u>Ex</u> $\mathcal{B} = H_0^1(0,1)$

$= \{ f : (0,1) \to \mathbb{R} \mid \int_0^1 |\nabla f|^2(x)\,dx < \infty,$
$f(0) = f(1) = 0 \}$

$$[\phi, \underset{\mathcal{B}}{\underset{\cap}{u}}] = \int_0^1 \phi(x)\,u(x)\,dx$$

$\mathcal{B}^* = H^{-1}(0,1)$

$$[\varphi, u] = \int_0^1 \nabla\varphi\,\nabla u(x)\,dx$$

$\mathcal{B}^* = H_0^1(\Omega)$

$\phi \in H^{-1}(\Omega)$ $\qquad u \longrightarrow \int_0^1 \phi(x)\, u(x)\, dx$

Linear functional on B

$\varphi \in H_0^1(\Omega)$ $\qquad u \longrightarrow \int_0^1 \nabla\varphi(x) \nabla u(x)\, dx$

Linear functional on B

$H^{-1}(\Omega)$: Sobolev space

" $\left\{ \phi \,\middle|\, \left| \int_0^1 \phi(x)\, u(x)\, dx \right| \leq C \, \|u\|_{H_0^1} \right\}$

$$H^{-s}(\Omega) = \left\{ \phi \,\middle|\, \left| \int_0^1 \phi(x)\, u(x)\, dx \right| \leq C \, \|u\|_{H_0^s(\Omega)} \right\}$$

$\xi \sim \mathcal{N}(v, T)$

Let $\phi_1, \dots, \phi_m \in \mathcal{B}^*$

Write $[\phi, \xi] = ([\phi_1, \xi], \dots, [\phi_m, \xi])$

Ex $[\phi, \xi] \sim \mathcal{N}([\phi, v], [\phi, T\phi])$

$[\phi, T\phi]$: $m \times m$ matrix

$[\phi, T\phi]_{i,j} = [\phi_i, T\phi_j]$

Assume $[\phi, T\phi]$ invertible

Thm $\mathbb{E}[\xi \mid [\phi, \xi]]$

$$= v + \sum_{i,j=1}^{m} T\phi_i \, \Theta_{i,j}^{-1} \, [\phi_j, \xi - v]$$

$\Theta_{i,j} = [\phi_i, T\phi_j]$

$$\zeta \mid [\phi, \zeta] = \mathcal{N}\left(\mathbb{E}[\zeta \mid [\phi, \zeta]], \phi^{\perp}(T)\right)$$

<u>Def</u> The short $\phi^{\perp}(T) : \mathcal{B}^* \longrightarrow \mathcal{B}$
of the symmetric positive operator $T : \mathcal{B}^* \to \mathcal{B}$
to the annihilator $\phi^{\perp} = \{ v \in \mathcal{B} \mid [\phi_i, v] = 0 \ \forall i \}$
is the symmetric positive operator defined
by

$$[\varphi, \phi^{\perp}(T)\varphi] = \inf_{\phi \in \mathrm{Span}\{\phi_1, \dots, \phi_m\}} [(\varphi - \phi), T(\varphi - \phi)]$$

<u>proof</u> $\psi \in \mathcal{B}^*$

$$[\psi, \zeta] \,\big|\, [\phi^1, \zeta], \dots, [\phi^m, \zeta]$$

$\psi \in \mathrm{Span}(\phi^1, \dots, \phi^m)$

$$[\psi,\xi] \sim \mathcal{N}\left(\mathbb{E}[\psi,\xi] \mid [\phi^1,\xi],\dots,[\phi^m,\xi]\right), \ [\psi, \phi^\perp(T)\psi]\ )$$

$$\parallel$$

$$\inf_{\phi \in \operatorname{span}\{\phi_1,\dots,\phi_m\}} [\psi-\phi, T(\psi-\phi)] = 0$$

$$([\psi,\xi], [\phi^1,\xi], \dots, [\phi^m,\xi])$$

$\to$ Gaussian vector of $\mathbb{R}^{m+1}$

Canonical Gaussian Field

$(\mathcal{B}, \|\cdot\|)$: Separable Banach Space

$\|u\|^2 = [Q^{-1}u, u]$

$Q: \mathcal{B}^* \longrightarrow \mathcal{B}$

continuous, symmetric, positive and invertible

$$[Q^{-1}v, v] = \|v\|^2$$

$$[\phi, Q\phi] = \|\phi\|_*^2 = \sup_{v \in B} \frac{([\phi, v])^2}{\|v\|^2}$$

$$[Q^{-1}u, v] = \langle u, v \rangle$$

$$[\phi, Q\varphi] = \langle \phi, \varphi \rangle_*$$

$$Q : B^* \longrightarrow B$$

$$[\phi, Q\varphi] = \frac{1}{4}\Big[[\phi+\varphi, Q(\phi+\varphi)] - [(\phi-\varphi), Q(\phi-\varphi)]\Big]$$

Canonical $\overline{\text{Gaussian}}$ Field

$$\xi \sim \mathcal{N}(0, Q)$$

ξ: Linear isometry

$(B^*, \|\ \|_*) \longrightarrow$ Gaussian space

$[\phi, \xi] \sim \mathcal{N}(0, \|\phi\|_*^2)$

$$\mathbb{E}\big[[\varphi,\xi][\phi,\xi]\big] = \langle \varphi, \phi \rangle_*$$
$$\shortparallel$$
$$[\varphi, Q\phi]$$

<u>Ex</u> $B = \mathbb{R}^N$

$\|x\|^2 = x^T A x$

A: $N \times N$ positive definite matrix

$[x, y] = x^T y$

$B^* = \mathbb{R}^N$

$$\|x\|_*^2 = \sup_{y \in B} \left(\frac{[x, y]}{\|y\|} \right)^2 = x^T A^{-1} x$$

$$Q = A^{-1}$$

$$\xi \sim \mathcal{N}(0, Q)$$

$$\phi \in B^* = \mathbb{R}^N$$

$$[\phi, \xi] = \phi^T \xi \sim \mathcal{N}(0, \phi^T A^{-1} \phi)$$

ξ: Gaussian vector
mean 0 and cov matrix A^{-1}

f: pdf (ξ)

$$f(x) \propto \exp\left(-\frac{x^T A x}{2}\right)$$

$$\propto \exp\left(-\frac{\|x\|^2}{2}\right)$$

$\phi_1, \phi_2, \dots, \phi_m \in \mathcal{B}^*$ linear indep.

$[\phi, v] = ([\phi_1, v], \dots, [\phi_m, v])$

$v \in \mathcal{B}$

Prop

$$\xi \mid [\phi, \xi] \sim \mathcal{N}\left(\mathbb{E}[\xi \mid [\phi, \xi]), \phi^{\perp}(Q)\right)$$

$$\mathbb{E}[\xi \mid [\phi, \xi]) = \sum_{i,j} [\phi_i, \xi]\, \Theta_{i,j}^{-1}\, Q \phi_j$$

$$\Theta_{i,j} = [\phi_i, Q\phi_j]$$

$\phi^{\perp} = \{ v \in \mathcal{B} \mid [\phi, v] = 0 \}$

Short $\phi^{\perp}(Q): \mathcal{B}^* \longrightarrow \mathcal{B}$

$$[\varphi, \phi^{\perp}(Q)\varphi] = \inf_{\phi \in \operatorname{span}\{\phi_1, \dots, \phi_m\}} [(\varphi - \phi), Q(\varphi - \phi)]$$

$$= \inf_{\phi \in \operatorname{span}\{\phi_1, \dots, \phi_m\}} \|\varphi - \phi\|_*^2$$

Prop $\mathbb{E}[\xi \mid [\phi, \xi] = Y]$ $\quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}$

is the minimizer of

$$\begin{cases} \min \|v\| \\ \text{s.t. } v \in \mathcal{B} \ \& \ [\phi, v] = Y \end{cases}$$

$$\phi^{\perp}(Q) = Q - \sum_{i,j} \Theta_{i,j}^{-1} \, Q\phi_i \otimes Q\phi_j$$

for $\varphi \in \mathcal{B}^*$

$$(Q\phi_i \otimes Q\phi_j)(\varphi) = [\varphi, Q\varphi_j] \, Q\varphi_i$$

$$Q\phi_i \otimes Q\phi_j : \ \mathcal{B}^* \longrightarrow \mathcal{B}$$

$$\varphi \longrightarrow [\varphi, Q\phi_j] \, Q\phi_i$$

Ex $\mathcal{B}$ = space of functions on $\mathbb{R}$

$\mathcal{B}^*$ = Space of generalized functions (distributions) on $\mathbb{R}$

For $\phi \in \mathcal{B}^*$ $\quad \int_{\mathbb{R}} \phi(x) u(x)\, dx$

Assume that $\mathcal{B} \subset C^1(\mathbb{R})$

Then $[\phi, u] = \int_{\mathbb{R}} \phi(x) u(x)\, dx$

$\forall x \in \mathbb{R}, \; \delta_x \in \mathcal{B}^*$

$\delta_x \circ \partial_x \in \mathcal{B}^*$

$[\delta_x, u] = u(x)$

$[\delta_x \circ \partial_x, u] = \partial_x u(x)$

For $\phi \in B^*$

$$[\phi, Q\phi] = \int_{\mathbb{R}^2} \phi(x) K(x,y) \phi(y)\, dx\, dy$$

$$K : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

$$\delta_x \in B^*$$

$$K(x,y) = [\delta_x, Q\delta_y]$$

$$\underbrace{\xi \sim \mathcal{N}(0,Q)}_{\text{cov operator}} \iff \underbrace{\xi \sim \mathcal{N}(0,K)}_{\text{cov function}}$$

$X_1, \dots, X_N \in \mathbb{R}$

What is

$$u(x) = \mathbb{E}\left[\xi(x) \,\middle|\, \begin{array}{l} \xi(X_1) = Y_1, \dots, \xi(X_N) = Y_N \\ \partial_x \xi(X_1) = Z_1, \dots, \partial_x \xi(X_N) = Z_N \end{array}\right]$$

$$\Updownarrow$$

$$u(x) = \mathbb{E}\left[[\delta_x, \xi] \,\middle|\, \begin{array}{l} [\phi_1, \xi] = Y_1 \dots [\phi_N, \xi] = Y_N \\ [\phi_{N+1}, \xi] = Z_1, \dots, [\phi_{2N}, \xi] = Z_N \end{array}\right]$$

$$\phi_1 = \delta_{X_1}, \dots, \phi_N = \delta_{X_N}$$

$$\phi_{N+1} = \delta_{X_1} \circ \partial_x, \dots, \phi_{2N} = \delta_{X_N} \circ \partial_x$$

$$u(x) = K(x, \phi)\, K(\phi, \phi)^{-1} \begin{pmatrix} Y \\ Z \end{pmatrix}$$

$K(x, \phi)$: $2N$ vector with entries

$$K(x, \phi_i) = \int K(x, y)\, \phi_i(y)\, dy$$

$K(\phi,\phi)$: $2N \times 2N$ matrix

with entries

$$K(\phi_i, \phi_j) := \int \phi_i(x) K(x,y) \phi_j(y) \, dx \, dy$$

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \\ Z_1 \\ \vdots \\ Z_N \end{pmatrix}$$

Optimal Recovery

$$\xi \sim \mathcal{N}(0, Q)$$

$$\phi_1, \dots, \phi_m \in \mathcal{B}^*$$

Player I

$u \in \mathcal{B}$

Player II

sees $[\phi, u] = ([\phi_1, u], \dots, [\phi_m, u])$

choose $\psi : \mathbb{R}^m \to \mathcal{B}$

max

min

$$\mathcal{E}(u, \psi) = \frac{\|u - \psi([\phi, u])\|}{\|u\|}$$

$$\|u\|^2 = [Q^{-1} u, u]$$

$$\sup_{u \in \mathcal{B}} \inf_{\psi} \mathcal{E}(u, \psi) = 0$$

$$m < \dim(\mathcal{B})$$

$$\inf_{\psi} \sup_{u \in \mathcal{B}} \mathcal{E}(u, \psi) \neq 0$$

<u>Thm</u> $\mathbb{E}[\xi \mid [\phi, \xi] = [\phi, u]]$

is the minimizer of

$$\min_{v \text{ s.t. } [\phi, v] = [\phi, u]} \max_{u \in B} \frac{\|u - v\|^2}{\|u\|^2}$$

$$\Psi\big[[\phi, u]\big] = \sum_{i,j} [\phi_i, u]\, \Theta^{-1}_{i,j}\, Q \phi_j$$

$$\Theta_{i,j} = [\phi_i, Q\phi_j]$$

Ex

		p	$1-p$
		II	
I	q	3	−2
	$1-q$	−2	1

Gain of Player II

Loss of Player I

Average Gain of Player II

$$3qp + 1(1-p)(1-q) - 2(1-q)p - 2(1-p)q$$

$$= 1 + p(8q-3) - 3q$$

$q = \frac{3}{8}$ Av. Gain Player II $= 1 + p \times 0 - \frac{9}{8}$

$= -1/8$

The lifted game

$\mathcal{P}_2(\mathbb{B})$: set of prob. measures on $\mathbb{B}$ with finite second moments

I

$\mu \in \mathcal{P}_2(\mathbb{B})$

II

$\Psi : \mathbb{R}^n \to \mathbb{R}$

$$\mathcal{E}(\mu, \Psi) = \frac{\mathbb{E}_{v\sim\mu}\left[\|v - \Psi[[\phi, v]]\|^2\right]}{\mathbb{E}_{v\sim\mu}\left[\|v\|^2\right]}$$

Thm If $\dim(\operatorname{span}\{\phi_1, \dots, \phi_m\}) < \dim(\mathcal{B})$

$$\Downarrow$$

$$\inf_{\psi} \sup_{\mu \in \mathcal{P}_2(\mathcal{B})} \mathcal{E}(\mu, \psi) = \sup_{\mu} \inf_{\psi} \mathcal{E}(\mu, \psi) = 1$$

Optimal strategy for Player I

$$\xi - \mathbb{E}[\xi \mid [\phi, \xi]] \sim \mathcal{N}(0, \Phi^{\perp}(Q))$$

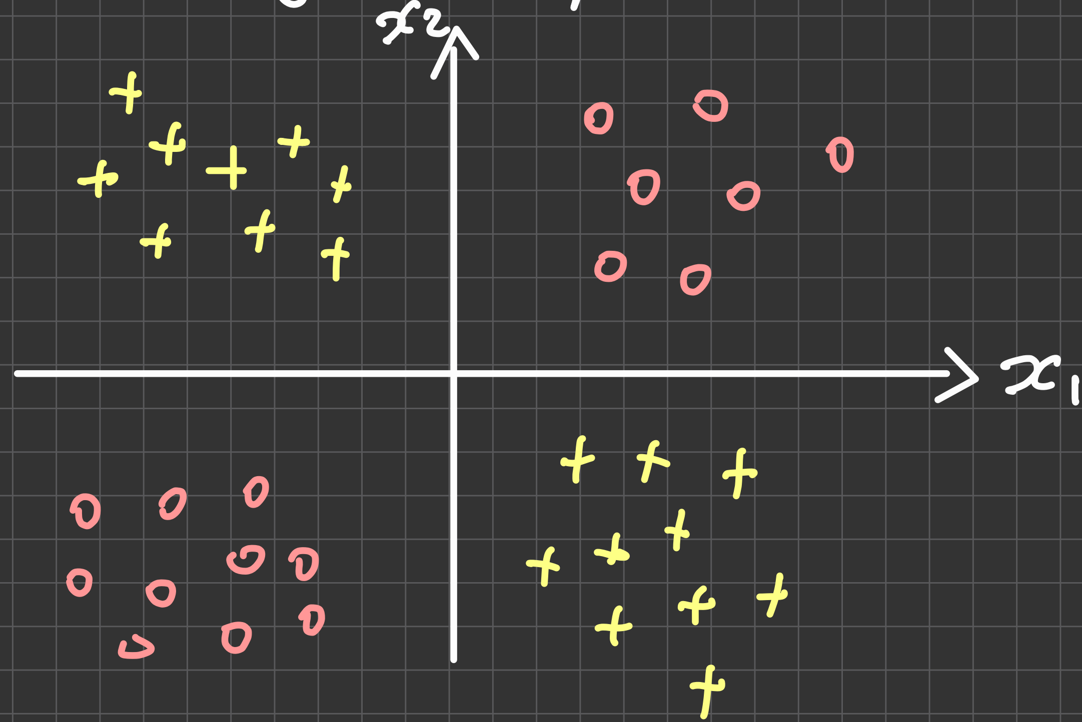
$$\xi \sim \mathcal{N}(0, Q)$$

Optimal strategy of Player II is

$$\psi[[\phi, v]] = \mathbb{E}[\xi \mid [\phi, \xi] = [\phi, v]]$$

Kernel Methods

Motivating example : XOR



N data points $(X_i, Y_i)_{1 \leq i \leq N}$

$X_i \in \mathbb{R}^2$, $Y_i \in \{0, +\}$

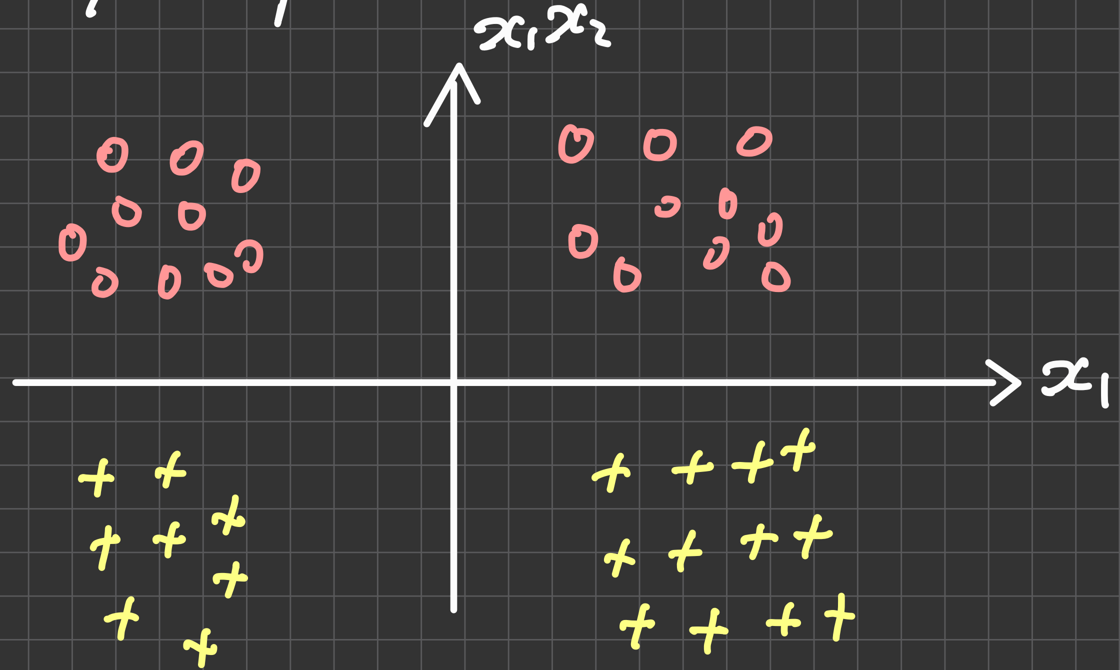
are not linearly separable

idea Consider lifted data points

$(\phi(X_i), Y_i)_{1 \leq i \leq N}$

for $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\quad \phi(x) = \begin{pmatrix} x_1 \\ \\ x_1 x_2 \end{pmatrix}$

These lifted N cluster points are linearly separable



Construction of kernels

<u>Def</u> (Inner product)

Let H be a vector space over $\mathbb{R}$

A function $\langle \cdot , \cdot \rangle_H : H \times H \longrightarrow \mathbb{R}$

is said to be an inner product on H if

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_H = \alpha_1 \langle f_1, g \rangle_H + \alpha_2 \langle f_2, g \rangle_H$

2. $\langle f, g \rangle_H = \langle g, f \rangle_H$

3. $\langle f, f \rangle_H \geq 0$

   $\langle f, f \rangle_H = 0 \Rightarrow f = 0$

norm $\|f\|_H = \sqrt{\langle f, f \rangle_H}$

Hilbert space: $(H, \langle \cdot, \cdot \rangle_H)$ s.t.

H is complete (contains limit of Cauchy sequences)

Def $\mathcal{X}$: non-empty set

A function $k: \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$

is called a kernel if $\exists$ a $\mathbb{R}$-Hilbert space

and a map $\phi: \mathcal{X} \longrightarrow H$ s.t.

$\forall x, x' \in \mathcal{X}$

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_H$$

Rk ϕ called Feature map

H: Feature space

Rk No conditions on $\mathcal{X}$

ex $\mathcal{X} = \{ \text{text documents} \}$

$H = \mathbb{R}^d$

$\phi(x) = (\phi_1(x), \dots, \phi_d(x))$

$\phi_i(x)$: # of times word i appears in doc. x

$\{1, \dots, d\} = \{\text{cat}, \text{dog}, \text{kennel}, \text{furry}, \dots\}$

Ex Given k is ϕ unique?

$k(x, x') = x \cdot x'$

$\mathcal{X} = \mathbb{R}$ $\quad \phi(x) = x \quad H = \mathbb{R}$

$\phi = \begin{pmatrix} x \\ 0 \end{pmatrix} \quad H = \mathbb{R}^2$

$\phi = \begin{pmatrix} x/\sqrt{2} \\ x/\sqrt{2} \end{pmatrix} \quad H = \mathbb{R}^2$

$$k(x, x') = \phi^T(x)\, \phi(x')$$
$$= \Phi^T(x) A^T A\, \Phi(x') \quad \text{if } A^T A = I$$

$$\mathcal{H} = \overline{\operatorname{span}\{x \to k(x, y)\}}$$

<u>Lemma</u> (Sums of kernels are kernels)

Given $\alpha > 0$ and k, k_1 and k_2 all kernels on $\mathcal{X}$

αk and $k_1 + k_2$ are kernels on $\mathcal{X}$

<u>proof</u>

$$k \Longleftrightarrow \phi, H$$

$$\alpha k \Longleftrightarrow \sqrt{\alpha}\,\phi, H$$

$$\left.\begin{array}{l} k_1 \Longleftrightarrow \phi_1, H_1 \\ k_2 \Longleftrightarrow \phi_2, H_2 \end{array}\right| \Rightarrow$$

$$k_1 + k_2 \Longleftrightarrow \phi = (\phi_1, \phi_2) \quad H = H_1 \times H_2$$

$$(k_1 + k_2)(x, x') = \langle \phi(x), \phi(x') \rangle_H$$

$$= \langle \phi_1(x), \phi_1(x') \rangle_{H_1} + \langle \phi_2(x), \phi_2(x') \rangle_{H_2}$$

Rk Difference of kernels may not be kernels

Lemma $\mathcal{X}, \tilde{\mathcal{X}}$: sets

$A : \mathcal{X} \longrightarrow \tilde{\mathcal{X}}$

Let k be a kernel on $\tilde{\mathcal{X}}$

Then $K(x, x') = k(A(x), A(x'))$

is a kernel on $\mathcal{X}$

Warping kernels

proof

$k \leftrightarrow \phi, H$

$K \leftrightarrow \phi \circ A, H$

$$K(x,x') = \left\langle \phi \circ A(x),\ \phi \circ A(x') \right\rangle_H$$

<u>Lemma</u> Product of kernels are kernels

Let k_1 be a kernel on $\mathcal{X}_1$

k_2 " $\mathcal{X}_2$

Then

(i) $K\big((x_1, x_2), (x_1', x_2')\big) = k_1(x_1, x_1')\, k_2(x_2, x_2')$

is a kernel on $\mathcal{X}_1 \times \mathcal{X}_2$

(ii) If $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$

$\Gamma(x,x') = k_1(x,x')\, k_2(x,x')$

is a kernel on $\mathcal{X}$

proof Only need to prove (ii)

since (ii) $\Rightarrow$ (i)

(ii) $\Rightarrow$ (i) If (ii) is true then

$(x_1, x_2, x_1', x_2') \to k_1(x_1, x_1')$ kernel on $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$

$(x_1, x_2, x_1', x_2') \to k_2(x_2, x_2')$ kernel on $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$

$(x_1, x_2, x_1', x_2') \to k_1(x_1, x_1')\, k_2(x_2, x_2')$ kernel on $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$

proof of (ii)

(ii) If $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$

$\Gamma(x, x') = k_1(x, x')\, k_2(x, x')$

is a kernel on $\mathcal{X}$

Need the following

Thm $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$x, x' \to k(x, x')$$

is a kernel on $\mathcal{X}$ iif

$\forall x_1, \dots, x_m \in \mathcal{X}$

$$\Theta = (\Theta_{i,j}) \qquad \Theta_{i,j} = k(x_i, x_j)$$

is symmetric & positive

Let $x_1, \dots, x_m \in \mathcal{X}$

$$\Theta^1_{i,j} = k_1(x_i, x_j) \qquad \Theta^2_{i,j} = k_2(x_i, x_j)$$

$$\Theta^1 = L L^T \quad \Big((\quad) = L \text{ Cholesky factor} \Big)$$

$\searrow$ Cholesky factorization of Θ^1

$c \in \mathbb{R}^m$

$$\sum_{i,j} c_i \Gamma(x_i, x_j) c_j = \sum_{i,j} c_i \Theta^1_{i,j} \Theta^2_{i,j} c_j$$

$$= \sum_{i,j,k} c_i c_j L_{ik} L_{jk} \Theta^2_{i,j}$$

$$= \sum_k (Z^k)^T \Theta^2 Z^k \geq 0$$

$$Z_k = \begin{pmatrix} c_1 L_{1k} \\ \vdots \\ c_m L_{mk} \end{pmatrix}$$

<u>Lemma</u> (Polynomial kernels)

Let $x, x' \in \mathbb{R}^d$, $d \geq 1$

$m \geq 1$ $\qquad c \geq 0$ $(c \in \mathbb{R})$

$m \in \mathbb{N}$

Then $k(x, x') = (\langle x, x' \rangle + c \quad)^m$

is a valid kernel

<u>proof</u> $\phi(x) = \begin{pmatrix} x \\ \sqrt{c} \end{pmatrix} \in \mathbb{R}^{d+1} = H$

$\phi(x) \cdot \phi(x') = x^T x' + c$

is a kernel

$\Downarrow$ stability under product

$(\langle x, x' \rangle + c)^m$ is a kernel

<u>Def</u> The space l_p of p-summable sequences is defined as all sequences $(a_i)_{i \geq 1}$ for which

$$\sum_{i=1}^{\infty} |a_i|^p < \infty$$

l_2: Hilbert space $\langle a, b \rangle = \sum_i a_i b_i$

<u>Lemma</u> Given a non-empty set $\mathcal{X}$ and a sequence of functions $(\phi_i(x))_{i \geq 1}$

in ℓ_2 where $\phi_i : \mathcal{X} \to \mathbb{R}$ is the ith coordinate of the feature map $\phi(x)$

Then

$$K(x, x') = \sum_{i=1}^{\infty} \phi_i(x)\, \phi_i(x')$$

is a well defined kernel on $\mathcal{X}$

<u>proof</u>

$$|K(x, x')| = \left|\sum_i \phi_i(x)\, \phi_i(x')\right| \underset{C.S.}{\leq} \|\phi(x)\|_{\ell_2} \|\phi(x')\|_{\ell_2}$$

K is well defined $\forall x, x' \in \mathcal{X}$

<u>Prop</u> (Taylor series kernel)

Assume we can define the Taylor series

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad |z| < r, \; z \in \mathbb{R}$$

for $r \in (0, \infty]$, with $a_n \geq 0 \quad \forall n$

Define $\mathcal{X} = \{ x \in \mathbb{R}^d \mid \|x\| < \sqrt{r} \}$

Then

$$K(x, x') = \sum_{n=0}^{\infty} a_n \langle x, x' \rangle^n \quad (= f(\langle x, x' \rangle))$$

is a valid kernel on $\mathcal{X}$

proof

$$K(x, x') = \sum_{n=0}^{\infty} a_n \langle x, x' \rangle^n$$

converges

CS inequ $|\langle x, x' \rangle| \leq \|x\| \|x'\| < r$

Ex $\quad k(x, x') = \underbrace{\exp(\langle x, x' \rangle)}_{\text{exponential kernel}}$ is a kernel on $\mathbb{R}^d$

$$f(z) = e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

Ex $G(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$

is a kernel on $\mathbb{R}^d$

$\downarrow$

Gaussian kernel

Normalized kernels

Prop Let K be a kernel on $\mathcal{X}$

Let

$$K'(x,x') = \begin{cases} 0 & \text{if } K(x,x)=0 \\ & \text{or } K(x',x')=0 \\ \dfrac{K(x,x')}{\sqrt{K(x,x)}\sqrt{K(x',x')}} & \text{otherw.} \end{cases}$$

Then K' is a kernel on $\mathcal{X}$

proof Let $x_1, \dots, x_m \in \mathcal{X}$, $c \in \mathbb{R}^m$

$K \leftrightarrow \phi, H$

$$\sum_{i,j} c_i c_j K'(x_i, x_j)$$

$$= \sum_{i,j} c_i c_j \frac{k(x_i, x_j)}{\sqrt{k(x_i, x_i) k(x_j, x_j)}}$$

$$= \sum_{i,j} c_i c_j \frac{\langle \phi(x_i), \phi(x_j) \rangle_H}{\|\phi(x_i)\|_H \|\phi(x_j)\|_H}$$

$$= \left\| \sum_i c_i \frac{\phi(x_i)}{\|\phi(x_i)\|_H} \right\|_H^2 \geq 0$$

$$K' \Longleftrightarrow \phi'(x) = \frac{\phi(x)}{\|\phi(x)\|_H}, \quad H$$

<u>Ex</u> $G(x, x') = \exp\left(- \frac{\|x - x'\|^2}{2\sigma^2} \right)$

is a kernel on $\mathbb{R}^d$

$\downarrow$

Gaussian kernel

<u>Sol</u>

$$K(x,x') = \exp\left(\frac{\langle x, x' \rangle}{\sigma^2}\right)$$

is a kernel on $\mathbb{R}^d$

$$G(x,x') = \frac{K(x,x')}{\sqrt{K(x,x)\,K(x',x')}}$$

$$= \frac{e^{x \cdot x'/\sigma^2}}{e^{\|x\|^2/\sigma^2}\, e^{\|x'\|^2/\sigma^2}}$$

$$= \exp\left(-\frac{1}{2\sigma^2}\|x - x'\|^2\right)$$

is a kernel

__Def__ A symmetric function
$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
is of positive type if $\forall\, m \geq 1$, $c \in \mathbb{R}^m$
$x_1, \dots, x_m \in \mathcal{X}$

$$\sum_{i,j} c_i c_j \, k(x_i, x_j) \geq 0$$

(k is also called PDS, positive definite symmetric)

The function k is strictly positive definite (or non-degenerate) if

$$\sum_{i,j} c_i c_j \, k(x_i, x_j) = 0 \;\; \& \;\; x_i \neq x_j \text{ for } i \neq j \implies c = 0$$

<u>Lemma</u>

Let H be any Hilbert space, $\mathcal{X}$ a non-empty set and $\phi : \mathcal{X} \to H$

Then $k(x, x') = \langle \phi(x), \phi(x') \rangle_H$ is a symmetric function of positive type (PDS)

<u>proof</u> $\sum_{i,j} c_i c_j \, k(x_i, x_j) = \sum_{i,j} c_i c_j \langle \phi(x_i), \phi(x_j) \rangle_H$

$$= \left\| \sum_i c_i \, \phi(x_i) \right\|_H^2 \geq 0$$

<u>Lemma</u>

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be PDS

Then $\forall x, x'$

$$\left(k(x, x') \right)^2 \leq k(x, x) \, k(x', x')$$

<u>proof</u> Let $\Theta = \begin{pmatrix} k(x, x) & k(x, x') \\ k(x', x) & k(x', x') \end{pmatrix}$

$$\Theta \geq 0 \Rightarrow \det(\Theta) \geq 0$$

$$\det(\Theta) = k(x, x) \, k(x', x') - \left(k(x, x') \right)^2 \geq 0$$

Thm Reproducing Kernel Hilbert Spaces (RKHS)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a PDS function.

Then there exists a Hilbert space $\mathcal{H}$ and a mapping $\phi : \mathcal{X} \to \mathcal{H}$ s.t.

(1) $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}} \quad \forall x, x' \in \mathcal{X}$

Furthermore $\mathcal{H}$ can be taken to be a space of functions $\mathcal{X} \to \mathbb{R}$ with the following property known as the reproducing property

(2) $h(x) = \langle h, k(x,\cdot) \rangle_H$

H is called the reproducing kernel Hilbert Space (RKHS) associated to k

<u>Rk</u> (1) & (2) $\Rightarrow$ $(H, \langle \cdot, \cdot \rangle_H)$ is unique

<u>proof</u>

proof of (1). Based on GPs

k PDS $\Rightarrow$ k is a covariance function

$\Rightarrow \exists$ a GP $\xi \sim \mathcal{N}(0,k)$

i.e. $k(x,x') = \mathbb{E}[\xi(x)\,\xi(x')]$

$\Rightarrow$ H = Gaussian Space containing all $(\xi(x))_{x\in\mathcal{X}}$

$$\phi(x) = \xi(x)$$

$$\langle \cdot , \cdot \rangle_H = L_2 \text{ inner product in } H$$

<u>Q</u> Is H the RKHS associated with k?

$\rightarrow$ no!

This H satisfies (1) but not (2)

$$H \neq \{ \mathcal{X} \rightarrow \mathbb{R} \}$$
$$= \{ \mathcal{X} \rightarrow \text{Gaussian space} \}$$

<u>proof</u> (1) & (2)

$\forall x \in \mathcal{X}$, define $\phi(x) : \mathcal{X} \rightarrow \mathbb{R}$

$\forall x' \in \mathcal{X}$, $\phi(x)(x') = k(x, x')$

$$\phi(x)(\cdot) = k(x, \cdot)$$

$$H_0 = \left\{ \sum_{i \in I} c_i \phi(x_i) \mid c_i \in \mathbb{R},\ x_i \in \mathcal{X},\ \mathrm{Car}(I) < \infty \right\}$$

$$\langle \cdot, \cdot \rangle : H_0 \times H_0 \longrightarrow \mathbb{R}$$
$$(f, g) \longrightarrow \langle f, g \rangle$$

$$f = \sum_{i \in I} a_i \phi(x_i)$$
$$g = \sum_{j \in J} b_j \phi(x_j)$$

$$\langle f, g \rangle = \sum_{\substack{i \in I \\ j \in J}} a_i b_i k(x_i, x_j') = \sum_{j \in J} b_j f(x_j') = \sum_{i \in I} a_i g(x_i)$$

$\langle \cdot, \cdot \rangle$: is symmetric, bilinear

$\langle f, g \rangle$: does not depend on particular representation of f & g

$$k \text{ PDS} \Rightarrow \langle f, f \rangle = \sum_{i,j \in I} a_i a_j k(x_i, x_j) \geq 0$$

$$\forall f = \sum_{i \in I} a_i \phi(x_i) \in H_0$$

$$\sum_{i,j=1}^{m} c_i c_j \langle f_i, f_j \rangle = \left\langle \sum_{i=1}^{m} c_i f_i, \sum_{i=1}^{m} c_i f_i \right\rangle \geq 0$$

$$\forall f_1, \dots, f_m \in H_0$$
$$c_1, \dots, c_m \in \mathbb{R}$$

$\Downarrow$

$\langle \cdot, \cdot \rangle$ is PDS in H_0

Prev. lemma $\Rightarrow$

$$\langle f, \phi(x) \rangle^2 \leq \langle f, f \rangle \langle \phi(x), \phi(x) \rangle$$
$$\forall f \in H_0,\ x \in \mathcal{X}$$

By def of $\langle \cdot, \cdot \rangle$

$$\forall f = \sum_{i \in I} a_i \phi(x_i) \in \mathcal{H}_0$$

$$(*) \quad \boxed{f(x) = \sum_{i \in I} a_i k(x_i, x) = \langle f, \phi(x) \rangle}$$

$$|f(x)|^2 \leq \langle f, f \rangle k(x, x)$$

$\Rightarrow \langle \cdot, \cdot \rangle$: defines an inner product on $\mathcal{H}_0$ which thereby becomes a pre-Hilbert space

$\Rightarrow \langle \cdot, \cdot \rangle$: defines an inner product on $\mathcal{H}_0$ which thereby becomes a pre-Hilbert space

$\mathcal{H}_0$ can be closed (completed) to form a Hilbert space in which $\mathcal{H}_0$ is dense

By Cauchy - Schwartz inequ.

$\forall x \in \mathcal{X}$

$$f \longrightarrow \langle f, \phi(x) \rangle$$

is Lipschitz

and therefore continuous

H_0 is dense in H $\Rightarrow$ reproducing prop *

holds in H

<u>Def</u> A Hilbert space H of functions is a RKHS if the evaluation functionals

$f \longrightarrow f(x)$ are bounded, i.e.

$\forall x \in \mathcal{X}, \exists C > 0$

$$|f(x)| \leq C \; \|f\|_H \quad , \forall f \in H$$

<u>Thm</u> (Riesz representation)

If $\mathcal{A}: H \to \mathbb{R}$ is a bounded linear operator in a Hilbert space H, there exists some $g_{\mathcal{A}} \in H$ s.t.

$$\mathcal{A}[f] = \langle f, g_{\mathcal{A}} \rangle_H \quad \forall f \in H$$

<u>Prop</u> (Reproducing prop)

If H is a RKHS then $\forall x \in \mathcal{X}$ there exists a function $k_x \in H$ s.t.

$$f(x) = \langle k_x, f \rangle_H \quad \forall f \in H$$

<u>Rk</u> $k(x,y) = \langle k_x, k_y \rangle_H$

$$= \langle k_y, k_x \rangle = k(y,x)$$

kernel associated with RKHS H

$$\sum_{i,j} c_i c_j \, k(x_i, x_j) = \Big\| \sum_i c_i k_{x_i} \Big\|_H^2 \geq 0$$

<u>Thm</u> For every (positive definite symmetric function) kernel on $\mathcal{X} \times \mathcal{X}$ there exists a unique RKHS with k as its reproducing kernel.
Conversely the reproducing kernel of a RKHS is unique (and PDS)

$$k(x,y) = \langle k_x, k_y \rangle_H$$

Ex RKHS associated with GP

Let $\xi \sim \mathcal{N}(0, A^{-1})$ $A^{-1}(i,j)$

Gaussian vector on $\mathbb{R}^N$

A: $N \times N$ PDS matrix

Here : $\mathcal{X} = \{1, 2, \cdots, N\}$

RKHS: $\mathbb{R}^N$ endowed with norm

$\|u\|^2 = u^T A u$ $\langle u, v \rangle = u^T A v$

$H: \{1, \dots, N\} \longrightarrow \mathbb{R} \iff \mathbb{R}^N$

$i \longrightarrow v_i$ $v \in \mathbb{R}^N$

Feature Map ($\mathbb{E}[\xi_i \xi_j] = k(i,j)$)

$\phi(i): \mathcal{X} \longrightarrow \mathbb{R}$ $\cdot \to k(i, \cdot)$

$j \longrightarrow \phi(i)(j) = K(i,j) = A^{-1}_{i,j}$

$$\langle \phi(i), \phi(j) \rangle = A^{-1}_{i,\cdot}\, A\, A^{-1}_{\cdot,j} = A^{-1}_{i,j}$$
$$= k(i,j)$$

Reproducing property

$$u : \{1, \dots, N\} \longrightarrow \mathbb{R}$$

$$\langle u, \phi(i) \rangle = u(i)$$
$$= u^T A A^{-1}_{\cdot,i}$$

<u>Ex</u> Let $\xi \sim \mathcal{N}(0, \mathcal{L}^{-1})$ be a GP on $\Omega \subset \mathbb{R}^d$

$\mathcal{L} : H^s_0(\Omega) \longrightarrow H^{-s}(\Omega)$ endowed with norm $\|u\|^2 = \int_\Omega u \mathcal{L} u$ for $u \in H^s_0(\Omega)$

$s > \frac{d}{2}$, $\mathcal{L}$: symm. pos. definite (elliplic)

$$\int u \mathcal{L} v = \int v \mathcal{L} u$$

$$\int u \mathcal{L} u \geq 0$$

$$\int u \mathcal{L} u = 0 \Rightarrow u = 0$$

$$u \in H_0^s(\Omega) \;:\quad u : \Omega \longrightarrow \mathbb{R}$$
$$\qquad\qquad\qquad\qquad x \longmapsto u(x)$$

$$\int u \mathcal{L} v = \int_\Omega u(x)\, \mathcal{L} v(x)\, dx$$

kernel: Green's function of $\mathcal{L}$

$$G(x, y) = \int_\Omega \delta_x \mathcal{L}^{-1} \delta_y$$

$$\phi(x) = G(x, \cdot)$$

$$\langle \phi(x), \phi(y) \rangle = \int_\Omega \phi(x)(x')\, \mathcal{L}\, \phi(y)(x')\, dx'$$

$$\langle u, v \rangle = \int_\Omega u \mathcal{L} v$$

$$\phi(x): \Omega \to \mathbb{R}$$
$$x' \to G(x,x')$$
$$\langle \phi(x), \phi(y) \rangle$$
$$\shortparallel$$
$$G(x,y) = \int_\Omega G(x,x')\,\mathcal{L}\,G(x',y)\,dx'$$
$$= \int_\Omega G(x,x')\,\underbrace{\delta(x'-y)}_{\rightsquigarrow \delta_y}\,dx'$$
$$\int f(x)\,\delta(x)\,dx = f(0)$$

$$u(x) = \int u\,\mathcal{L}\,G(\cdot,x) = \langle u, \phi(x) \rangle$$

H: the reproducing kernel Hilbert space associated to k

Any Hilbert space H st.

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

with $k(x, x') = \langle \phi(x), \phi(x') \rangle$

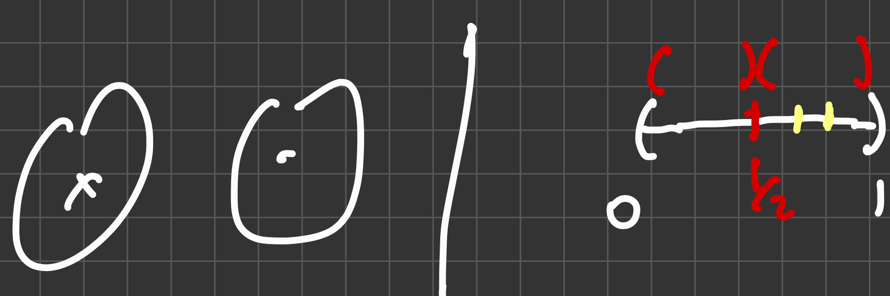
$\forall x, x'$

is called a feature space associated

and ϕ is called a feature map.

Mercer's thm

Hausdorff space: topological space in which each pair of distinct points can be separated by disjoint open sets



Borel measure: Any measure defined on σ-algebra of Borel sets

Finite measure: measure that always takes finite values

Thm Let $\mathcal{X}$ be a compact Hausdorff space and μ a finite Borel measure with support $\mathcal{X}$.

Suppose k is a continuous SPD kernel on $\mathcal{X}$, and define the integral operator $T_k : L_2(\mathcal{X}, \mu) \longrightarrow L_2(\mathcal{X}, \mu)$ by

$$(T_k f)(\cdot) = \int_{\mathcal{X}} k(\cdot, x) f(x) \, \mu(dx)$$

which is positive definite, i.e. $\forall f \in L_2(\mathcal{X}, \mu)$

$$\int_{\mathcal{X}^2} k(u,v) f(u) f(v) \, \mu(du) \, \mu(dv) \geq 0$$

(Mercer's condition)

Then there is an orthonormal basis $\{\psi_i\}$ of $L_2(\mathcal{X}, \mu)$
consisting of eigenvectors of T_k s.t. the corresponding sequence of eigenvalues $\{\lambda_i\}$ are non-negative

The eigenfunctions corresponding to non-zero eigenvalues can be taken as continuous functions on $\mathcal{X}$ and $k(u, v)$ has the representation

$T_k \psi_i = \lambda_i \psi_i$

$$k(u, v) = \sum_{i=1}^{\infty} \lambda_i \psi_i(u) \psi_i(v)$$

where the convergence is absolute and uniform

Rk

$$k(u, v) = \left\langle \left(\sqrt{\lambda_i}\, \psi_i(u)\right)_i, \left(\sqrt{\lambda_i}\, \psi_i(v)\right)_i \right\rangle_{\ell_2}$$

$$\phi(x) = \left(\sqrt{\lambda_1}\, \psi_1(x), \sqrt{\lambda_2}\, \psi_2(x), \dots \right)$$

$\mathcal{H} = \ell_2 = \{ (a_1, \dots, \dots, \dots) \}$

$$\langle a, b \rangle = \sum_{i=1}^{\infty} a_i b_i$$

Bochner's theorem

<u>Thm</u> $k(x, x') = \psi(x - x')$ is a kernel on $\mathbb{R}^d$ iif $\exists$ a finite non-negative Borel measure Λ on $\mathbb{R}^d$ /

$$\psi(x - x') = \int_{\mathbb{R}^d} e^{i\omega^T(x - x')} \Lambda(d\omega)$$

<u>Rk</u> One may normalize k. s.t. $\psi(0) = 1$

$\Rightarrow \Lambda$ is a prob. measure and k corresponds to its characteristic function

$$\Rightarrow \psi(x - x') = \mathbb{E}_{\omega \sim \Lambda}\left[e^{i\omega^T(x - x')} \right]$$

ω_j iii $\sim \Lambda$

$$\Psi(x-x') \approx \frac{1}{N} \sum_{j=1}^{N} e^{i\omega_j^T(x-x')}$$

$$\simeq \underline{\left(\frac{e^{i\omega_1^T x}}{\sqrt{N}}, \frac{e^{i\omega_2^T x}}{\sqrt{N}}, \dots, \frac{e^{i\omega_N^T x}}{\sqrt{N}}\right)} \cdot \left(\frac{e^{i\omega_1^T x'}}{\sqrt{N}}, \dots, \frac{e^{i\omega_N^T x'}}{N}\right)$$

<u>Rk</u> $k(x,x') = \Psi(x-x')$

$$(X_1, Y_1) \dots (X_M, Y_M)$$

$$k(X,X)^{-1} \to M$$

$$\phi(x) = (\phi_1(x), \dots, \phi_N(x))$$

$$f(x) = \sum_{i=1}^{N} c_i \phi_i(x)$$

$M >> N$

$$\min_{f} \; \sum_{i=1}^{M} \| f(X_i) - Y_i \|^2 + \lambda \|f\|_K^2$$

inverting $M \times M$ matrix

$$K(x, x') = \sum_{i=1}^{N} \phi_i(x) \phi_i(x')$$

$$\|f\|_K^2 = \sum_{i=1}^{N} c_i^2$$

$$\min_{c} \sum_{i=1}^{M} \Big| \sum_{j=1}^{N} c_j \phi_j(X_i) - Y_i \Big|^2 + \lambda \sum_{j=1}^{N} c_j^2$$

inverting a $N \times N$ matrix

$$\Psi(x - x') = \int_{\mathbb{R}^d} e^{i\omega(x - x')} \, d\Lambda(\omega)$$

Ex $k(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$

$\Updownarrow$

$$\Lambda(\omega) = \left(\frac{2\pi}{\sigma^2}\right)^{-d/2} \exp\left(-\sigma^2 \frac{\|\omega\|_2^2}{2}\right)$$

$$k(x,x') = \exp\left(-\frac{\|x-x'\|_1}{\sigma}\right)$$

$\Updownarrow$

$$\Lambda(\omega) = \prod_{i=1}^{d} \frac{\sigma}{\pi(1+\omega_i^2)}$$

Radial kernels

Def The kernel $k(x,x')$ is said to be radial if

$$k(x,x') = \Psi(\|x-x'\|)$$

for some (positive definite) $\Psi : [0,\infty) \to \mathbb{R}$

Thm (Schoenberg's thm)

A continuous function $\psi : [0,\infty) \to \mathbb{R}$ is positive definite and radial (i.e. it defines a kernel $k(x,x') = \psi(\|x-x'\|)$ on $\mathbb{R}^d$ for all d) iff it is of the form

$$\psi(r) = \int_0^\infty \exp(-r^2 t^2)\, \mu(dt)$$

where μ is a finite non-negative Borel measure on $[0,\infty)$

$\Rightarrow$ radial kernels are mixtures of Gaussians

Rk **Thm**, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

s.t. $\forall x_1, \dots, x_N$, $\forall N$

$\Theta_{i,j} = k(x_i, x_j)$

Θ is symm. positive

$$\Rightarrow \quad k(x,x') = \langle \phi(x), \phi(x') \rangle_H$$

Ridge regression

data $(X_1, Y_1), \ldots, (X_n, Y_n)$

$X_i \in \mathbb{R}^d \qquad Y_i \in \mathbb{R}$

Ridge regression: Linear regression with L^2 penalty

$X_{n+1} \to$ predict Y_{n+1}

idea: $a^T X_{n+1} \qquad a \in \mathbb{R}^d$

$$\min_{a \in \mathbb{R}^d} \sum_{i=1}^{n} |Y_i - a^T X_i|^2 + \lambda a^T a$$

Solution

$$X = \begin{array}{cc} & \begin{array}{cccc} X_1 & X_2 & X_j & X_n \end{array} \\ i \rightarrow & \underbrace{\begin{pmatrix} & \vdots & & \vdots \\ X_{i,1} & X_{i,2} & X_{i,j} & X_{i,n} \\ & & & \end{pmatrix}}_{n} d \end{array} \in \mathbb{R}^{d \times n}$$

$$X_{i,1} := (X_1)_i \qquad (X_{i,j}) = (X_j)_i$$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

Loss: $\|Y - X^T a\|_2^2 + \lambda \|a\|_2^2$

$$= Y^T Y - 2 Y^T X^T a + a^T X X^T a + \lambda a^T a$$

$$= Y^T Y - 2 Y^T X^T a + a^T (X X^T + \lambda I) a$$

minimized for

$$a = a^* = (X X^T + \lambda I)^{-1} X Y$$

which is the classical regularized least square solution

Kernel ridge regression

data $(X_1, Y_1), \dots, (X_n, Y_n)$

$X_i \in \mathcal{X}$

$Y_i \in \mathbb{R}$

$k(x, x')$: kernel $\quad \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

ϕ: feature map

H: feature space (RKHS)

$$k(x, x') = \langle \phi(x), \phi(x') \rangle_H$$

Idea: $\quad X_{n+1} \xrightarrow{\text{predict}} Y_{n+1}$

predict Y_{n+1} with $f(X_{n+1})$

$f \in H$

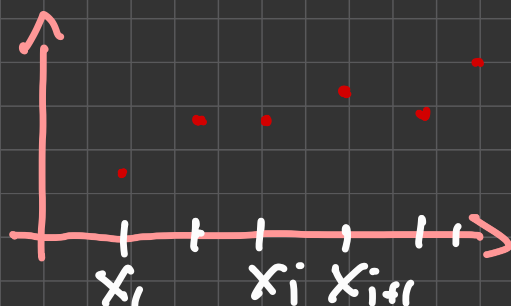
$$(1) \min_{f \in H} \sum_{i=1}^{n} |Y_i - f(X_i)|^2 + \lambda \|f\|_H^2$$

Using $\langle f, \phi(X_i) \rangle = f(X_i)$

$$\Updownarrow$$

(2) $$\min_{f\in H} \sum_{i=1}^{n} |Y_i - \langle f, \phi(X_i)\rangle_H|^2 + \lambda \|f\|_H^2$$

Q: Why $\exists\ \infty$ # of f / $\sum_i |Y_i - f(X_i)|^2 = 0$



Thm The minimizer of (1) $\Leftrightarrow$ (2) is

$$f = \sum_{i=1}^{n} c_i \phi(X_i)$$

Proof $$f = \underbrace{\sum_{i=1}^{n} c_i \phi(X_i)}_{f_\phi} + f_\perp$$

$$f_\phi \in \operatorname{span}\{\phi(X_1), \dots, \phi(X_n)\}$$

$f_\perp$: orthog. complement in H

$\mathcal{L}$: loss function

$$\mathcal{L}(f) = \sum_{i=1}^{n} |Y_i - \langle f_\phi + f_\perp, \phi(X_i)\rangle|^2 + \lambda \|f_\phi + f_\perp\|_H^2$$

$$\langle f_\perp, \phi(X_i)\rangle = 0$$

$$= \sum_{i=1}^{n} \underbrace{|Y_i - \langle f_\phi, \phi(X_i)\rangle|^2}_{\text{does not depend on } f_\perp} + \lambda \underbrace{(\|f_\phi\|_H^2 + \|f_\perp\|_H^2)}_{\text{minimized for } f_\perp = 0}$$

$\Rightarrow \mathcal{L}(f)$ minimized for

$f \in \mathrm{Span}\{\phi(X_1), \dots, \phi(X_n)\}$

$k(X,X)$: Gram matrix $(n \times n)$

$$(k(X,X))_{i,j} = k(X_i, X_j)$$

For $f = \sum_{i=1}^{n} c_i \phi(X_i)$

$$\mathcal{L}(f) = \| Y - k(X,X)c \|_2^2 + \lambda \, c^T k(X,X) c$$

$$\langle \phi(X_i), \phi(X_j) \rangle_H = k(X_i, X_j)$$

Minimizer in c is

$$c^* = (k(X,X) + \lambda I)^{-1} Y$$

for $x \in \mathcal{X}$, $k(X, x)$: n dim vector with entries $k(X_i, x)$

$$f^*(x) = Y^T (k(X,X) + \lambda I)^{-1} k(X, x)$$
$$= k(x, X) \; (k(X,X) + \lambda I)^{-1} Y$$

$$f^*(x) = \sum_{i=1}^{n} c_i^* \underbrace{\phi(X_i)(x)}_{k(X_i, x) = k(x, X_i)}$$

Let ξ be the centered GP with covariance function k

$\xi \sim \mathcal{N}(0, k)$

ξ defined by quadratic norm $\|\cdot\|_H$

$\mathbb{E}[\xi(x)\xi(x')] = k(x, x')$

(H self-dual, using $\langle \cdot, \cdot \rangle_H$ to define duality)

for $\phi \in H$, $\langle \phi, \xi \rangle \sim \mathcal{N}(0, \|\phi\|_H^2)$

$\xi : H \longrightarrow$ Gaussian space
$\phi \longrightarrow \xi(\phi)$

<u>Thm</u> For

$$f_\lambda^* = \underset{f \in H}{\operatorname{argmin}} \sum_{i=1}^n |Y_i - f(X_i)|^2 + \lambda \|f\|_H^2$$

we have

$$f_\lambda^*(x) = \mathbb{E}\left[\xi(x) \,\middle|\, \xi(X_i) = Y_i + Z_i \quad \forall i = 1, \dots, n\right]$$

where Z_i are i.i.d. $\sim \mathcal{N}(0, \lambda)$

If $k(X, X)$ invertible

as $\lambda \downarrow 0$, $f_\lambda^* \to f_0^*$ (pointwize in $\|\cdot\|_H$ norm)

f_0^* is the minimizer of

$$\begin{cases} \min \|f\|_H \\ \text{s.t. } f(X_i) = Y_i \end{cases}$$

and $f_0^*(x) = \mathbb{E}\left[\xi(x) \,\middle|\, \underset{i=1,\dots,n}{\xi(X_i) = Y_i}\right]$

proof Consider Gaussian vector

$$(\underbrace{\xi(X_1) - Z_1}_{V_1}, \dots, \underbrace{\xi(X_n) - Z_n}_{V_n}, \xi(x))$$

$$\mathrm{Cov}(V, V) = \mathrm{Cov}(V) = k(X, X) + \lambda I$$

$$\mathrm{Cov}(V, \xi(x)) = k(X, x)$$

$$\mathbb{E}[\xi(x) | V] = V^T \mathrm{cov}(V)^{-1} \mathrm{cov}(V, \xi(x))$$

$$= \mathrm{Cov}(\xi(x), V)\, \mathrm{Cov}(V)^{-1} V$$

$$\mathrm{Cov}\left[\xi(x) - V^T \mathrm{cov}(V)^{-1} \mathrm{Cov}(V, \xi(x)),\ V\right] = 0$$

Q: Unpack convergence

$$f^*_\lambda(x) = k(x, X)\,(k(X,X) + \lambda I)^{-1} Y$$

$$f^*_0(x) = k(x, X)\, k(X,X)^{-1} Y$$

$$\left|\begin{array}{l} \min \ \|f\|_H \\ \text{s.t. } f(x_i) = y_i \end{array}\right.$$

$$f = f_\phi + f_\perp \qquad f_\phi \in \mathrm{span}\{\phi(x_1), \dots, \phi(x_n)\}$$

$$f_\perp \in \text{orth. comp. in } H$$

$$\left|\begin{array}{l} \min \ \|f\|_H^2 \\ \text{s.t. } f(x_i) = y_i \end{array}\right. \iff \left|\begin{array}{l} \min \ \|f_\phi\|_H^2 + \|f_\perp\|_H^2 \\ \text{s.t. } \langle f, \phi(x_i)\rangle = y_i \end{array}\right.$$

$$\langle f, \phi(x_i)\rangle = \langle f_\phi, \phi(x_i)\rangle$$

minimized for $f_\perp = 0$

$$f = \sum_{i=1}^{n} c_i \phi(x_i)$$

$$\Leftrightarrow \left| \begin{array}{ll} \min & c^T k(X,X)\, c \\ \text{s.t.} & c^T k(X,X) = Y \end{array} \right.$$

$$\Leftrightarrow \quad c^* = k(X,X)^{-1} Y$$

$$f^*_\lambda(x) = k(x,X)\left(k(X,X) + \lambda I\right)^{-1} Y$$

$$f^*_0(x) = k(x,X)\, k(X,X)^{-1} Y$$

$$\Downarrow$$

$$f^*_\lambda(x) \to f^*_0(x) \quad \forall x \in \mathcal{X}$$

$$\Downarrow$$

$$f^*_\lambda \xrightarrow{H} f^*_0 \qquad (\text{in RKHS norm } \|\cdot\|_H)$$

$$\|f^*_\lambda - f^*_0\|_H^2 = \left\| \sum_{i=1}^{n} (c^*_{\lambda,i} - c^*_{0,i})\, \phi(x_i) \right\|_H^2$$

$$c^*_\lambda = \left(k(X,X) + \lambda I\right)^{-1} Y$$

$$c_0^* = k(X, X)^{-1} Y$$

$$\lim_{\lambda \to 0} c_\lambda^* = c_0$$

$$A(\lambda) = k(X, X) + \lambda I$$

$$\frac{d}{d\lambda} A^{-1}(\lambda)\Big|_{\lambda=0} = - A(0)^{-1} \frac{d}{d\lambda} A(0) A^{-1}(0)$$

Representer theorem

<u>Thm</u> Let k be a kernel on $\mathcal{X}$ and let $\mathcal{H}$ be its associated RKHS

For $x_1, \dots, x_n \in \mathcal{X}$ consider the opt. pb

$$(*) \quad \min_{f \in \mathcal{H}} D\left(f(x_1), \dots, f(x_n)\right) + P\left(\|f\|_{\mathcal{H}}^2\right)$$

where P is nondecreasing and

D depends on f only through $f(x_1), \dots, f(x_n)$

If $(*)$ has a minimizer then it has a minimizer of the form

$$f(x) = \sum_{i=1}^{n} c_i k(x, x_i) \quad \text{where } c_i \in \mathbb{R}$$

Furthermore if P is strictly increasing then every minimizer of $(*)$ has this form.

<u>Proof</u> Write

$$J(f) = D(f(x_1), \dots, f(x_n)) + P(\|f\|_H^2)$$

Consider $S := \operatorname{span}\{k(\cdot, x_1), \dots, k(\cdot, x_n)\}$

$$S \subset H$$

S finite dim $\Rightarrow$ S closed

projection thm $\Rightarrow$ $H = S \oplus S^{\perp}$

$\hookrightarrow$ orth comp. in H

$\Rightarrow \forall f \in \mathcal{H}$, f can be written

$$f = f_{\parallel} + f_{\perp} \quad \text{where } f_{\parallel} \in S, \; f_{\perp} \in S^{\perp}$$

$$\langle f_{\perp}, k(\cdot, x_i) \rangle = 0 \quad \forall i$$

reproducing property $\Rightarrow$

$$\begin{aligned} f(x_i) &= \langle f, k(\cdot, x_i) \rangle_H \\ &= \langle f_{\parallel}, k(\cdot, x_i) \rangle_H + \langle f_{\perp}, k(\cdot, x_i) \rangle_H \\ &= f_{\parallel}(x_i) \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} J(f) &= D(f(x_1), f(x_2), \dots, f(x_n)) + P(\|f\|_H^2) \\ &= D(f_{\parallel}(x_1), \dots, f_{\parallel}(x_n)) + P(\|f\|_H^2) \\ &\geq D(f_{\parallel}(x_1), \dots, f_{\parallel}(x_n)) + P(\|f_{\parallel}\|_H^2) \\ &= J(f_{\parallel}) \end{aligned}$$

$$\|f\|_H^2 = \|f_{\parallel}\|_H^2 + \|f_{\perp}\|_H^2$$

$\Rightarrow$ If f is a minimizer of $J(f)$
so is $f_{\parallel}$

second statement $\Leftarrow$ P is strictly $\nearrow$

$\Downarrow$

$J(f) > J(f_{\parallel})$
if $f \notin S$

Alternate proof

$$\min_{f} \; D(f(x_1), \dots, f(x_n)) + P(\|f\|_H^2)$$

$$\Updownarrow$$

$$\min_{\substack{z \in \mathbb{R}^N \\ f \in H \\ \text{s.t. } f(x_i) = z_i}} \; D(z_1, \dots, z_n) + P(\|f\|_H^2)$$

$$\min_{z \in \mathbb{R}^n} \; \min_{\substack{f \in H \\ \text{s.t. } f(x_i) = z_i}} \; D(z_1, \dots, z_n) + P(\|f\|_H^2)$$

Kernel PCA

M centered observations

$x_k : k = 1, \dots, M \ , \ x_k \in \mathbb{R}^N$

$$\sum_{k=1}^{M} x_k = 0$$

PCA diagonalizes the cov. matrix

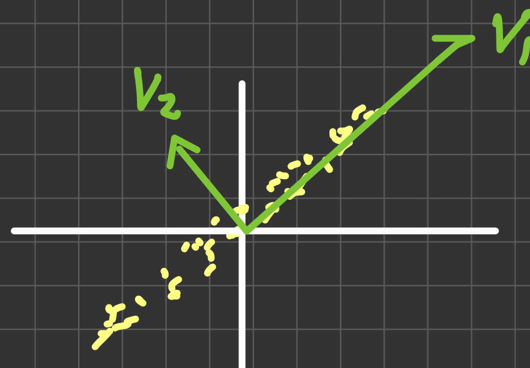
$$C = \frac{1}{M} \sum_{i=1}^{M} x_i x_i^T$$

$\rightarrow$ solve eigenvalue pb

$$* \quad C v = \lambda v$$

for eigenvalues $\lambda \geqslant 0$ and $v \in \mathbb{R}^N \setminus \{0\}$

$N = 2$



$\lambda_1 \gg \lambda_2 > 0$

$(N \gg M)$

$$Cv = \lambda v$$

$$\Updownarrow$$

$$Cv = \frac{1}{M}\sum_{i=1}^{M} (x_i^T v) x_i = \lambda v$$

$\lambda > 0 \Rightarrow v \in \operatorname{span}\{x_1, \dots, x_M\}$

$\lambda = 0 \Rightarrow v \in \operatorname{span}\{x_1, \dots, x_M\} \oplus_{\perp} V_{\perp}$

$\Rightarrow$ All $v \in \operatorname{span}\{x_1, \dots, x_M\}$

$Cv = \lambda v$, $w \in V_{\perp}$ $\Rightarrow$ $C(v+w) = Cv = \lambda v$

$\lambda \neq 0$ you can assume $w = 0$

$V_{\perp} = \{w \in \mathbb{R}^N,\ w^T x_i = 0 \quad \forall i\}$

Generalization to PCA in feature space (Kernel PCA)

H: Hilbert space (feature space)

$\phi: \mathbb{R}^N \to H$ (feature map)

Kernel PCA $\to$ do PCA on $\phi(x_i)_{i=1,\dots,M}$

Assume data is centered in H

$$\sum_{i=1}^{M} \phi(x_i) = 0$$

$$\bar{C}: H \to H$$

Cov matrix in H

$$\bar{C} = \frac{1}{M} \sum_{i=1}^{M} \phi(x_i)\, \phi(x_i)^T$$

$\phi(x_i)\phi(x_i)^T$: linear operator

$$H \to H$$

$$f \to \phi(x_i) \langle \phi(x_i), f \rangle$$

We want eigenpairs of $\bar{C}$

$$(v, \lambda) \qquad \lambda \geq 0$$

$$\bar{C} v = \lambda v$$

by same argument as for PCA (in $\mathbb{R}^N$)

$$v \in \text{Span}\{\phi(x_1), \dots, \phi(x_M)\}$$

$$\bar{C} v = \lambda v$$

$$\Updownarrow$$

(1) $\lambda \phi(x_k)^T v = \phi(x_k)^T \bar{C} v \quad \forall k = 1, \dots, M$

$\rightarrow \exists \alpha_i \quad (i = 1, \dots, M)$ s.t.

(2) $v = \sum_{i=1}^{M} \alpha_i \phi(x_i)$

(1) + (2) $\Rightarrow$

(3) $\lambda \sum_{i=1}^{M} \alpha_i \phi(x_k)^T \phi(x_i) =$

$\frac{1}{M} \sum_{i=1}^{M} \alpha_i \quad \phi(x_k)^T \sum_{j=1}^{M} \phi(x_j) \left(\phi(x_j)^T \phi(x_i) \right)$

$k = 1, \dots, M$

Define $M \times M$ matrix K by

kernel $\Leftrightarrow \phi$

$K_{i,j} = \phi(x_i)^T \phi(x_j) = k(x_i, x_j)$

(3) $\Leftrightarrow M \lambda K \alpha = K^2 \alpha$

$\Leftarrow \quad M \lambda \alpha = K \alpha$

$\rightarrow$ only need to diagonalize K

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_M \text{ : eigenvalues}$$

$$\alpha^1 \quad \dots \quad \alpha^M \text{ eigenvectors}$$

$$\lambda_i \alpha^i = K \alpha^i$$

λ_p: first non zero eigenvalue

Normalize $\alpha^p, \dots, \alpha^M$ by

$$(4) \quad \langle v^k, v^k \rangle_H = 1 \qquad \forall k = p, \dots, M$$

$$(4) \Leftrightarrow 1 = \sum_{i,j=1}^{M} \alpha_i^k \alpha_j^k \, \phi(x_i)^T \phi(x_j)$$

$$= \sum_{i,j=1}^{M} \alpha_i^k \alpha_j^k K_{i,j}$$

$$= \alpha^{k,T} K \alpha^k$$

$$= \lambda_k \, \alpha^{k,T} \alpha^k$$

$$\Leftrightarrow \quad \alpha^{k,T} \alpha^k = 1/\lambda_k$$

For principal component extraction need to compute projections on eigenvectors $v^k \in H$, $k = p, \dots, M$

Let x be a test point with an image $\phi(x) \in H$

$$v^{k,T} \phi(x) = \sum_{i=1}^{M} \alpha_i^k \phi(x_i)^T \phi(x)$$
$$= \sum_{i=1}^{M} \alpha_i^k \, k(x_i, x)$$

KPCA

1. Compute matrix

$$K_{i,j} = k(x_i, x_j) \qquad (M \times M \text{ matrix})$$

2. Compute eigenvectors

$$\lambda_j \alpha^j = K \alpha^j$$

normalize $\Leftrightarrow$ $\alpha^{j,T} \alpha^j \lambda_j = 1$

3. Compute projections of a test point onto eigenvectors

$$\langle v^j, \phi(x) \rangle = \sum_{i=1}^{M} \alpha_i^j \, k(x_i, x)$$

Centering in feature space

drop assumption $\sum_i \phi(x_i) = 0$

Use

$$\tilde{\phi}(x_i) = \phi(x_i) - \frac{1}{M} \sum_{i=1}^{M} \phi(x_i)$$

instead (these are centered in $\mathcal{H}$)

prev. analysis holds with cov. matrix

$$\tilde{K}_{i,j} = \tilde{\phi}(x_i) \tilde{\phi}^T(x_j)$$

eig pb $\rightarrow$ $\tilde{\lambda} \tilde{\alpha} = \tilde{K} \tilde{\alpha}$

$\tilde{\alpha}$ $\rightarrow$ expansion coeff. of an eigenvector

$$\tilde{v} = \sum_{i=1}^{M} \tilde{\alpha}_i \, \tilde{\phi}(x_i)$$

express $\tilde{K}$ in terms of K

$\mathbb{1}$: $M \times M$ matrix, $\mathbb{1}_{i,j} = 1 \quad \forall i,j$

$$\tilde{K}_{i,j} = \tilde{\phi}(x_i)^T \tilde{\phi}(x_j)$$

$$= \left(\phi(x_i) - \frac{1}{M} \sum_{m=1}^{M} \phi(x_m)\right)^T \left(\phi(x_j) - \frac{1}{M} \sum_{n=1}^{M} \phi(x_n)\right)$$

$$= K - \frac{1}{M} \mathbb{1} K - \frac{1}{M} K \mathbb{1} + \frac{1}{M^2} \mathbb{1} K \mathbb{1}$$

Normalize the $\tilde{v}^k$ in H

$$\Updownarrow$$

$$\tilde{\lambda}_k \left(\tilde{\alpha}^{k,T} \tilde{\alpha}^k\right) = 1$$

Feature extraction

$t \in \mathbb{R}^N$: test point

$$\tilde{v}^{k,T} \tilde{\phi}(t) = \sum_{i=1}^{M} \tilde{\alpha}_i^k \tilde{\phi}(x_i)^T \tilde{\phi}(t)$$

$$= \sum_{i=1}^{M} \tilde{\alpha}_i^k \tilde{K}(x_i, t)$$

L test points: $t_1, \dots, t_L$

$$K^{test}_{i,j} = \phi(t_i)^T \phi(x_j) = k(t_i, x_j)$$

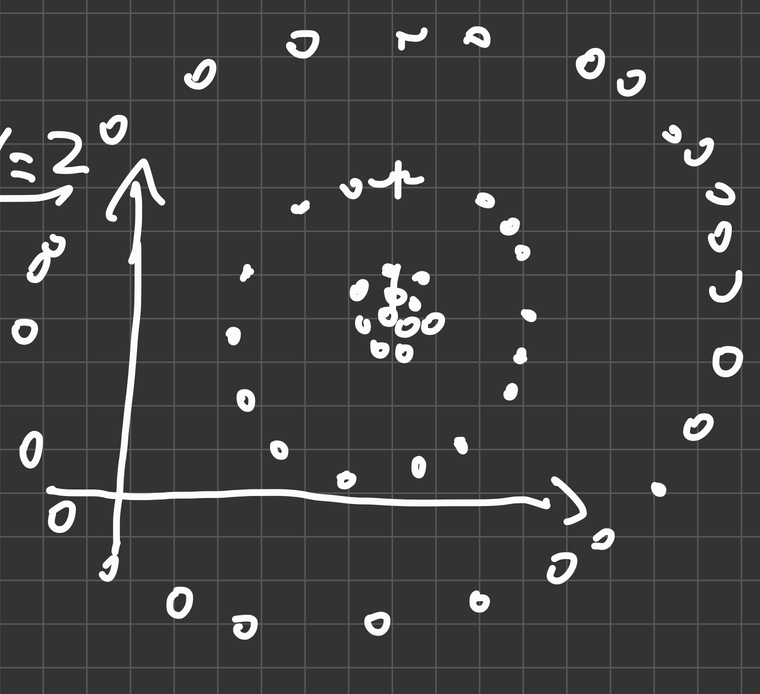
$$\tilde{K}^{test}_{i,j} = \left(\phi(t_i) - \frac{1}{M}\sum_{m=1}^{M} \phi(x_m)\right)^T \left(\phi(x_j) - \frac{1}{M}\sum_{n=1}^{M} \phi(x_n)\right)$$

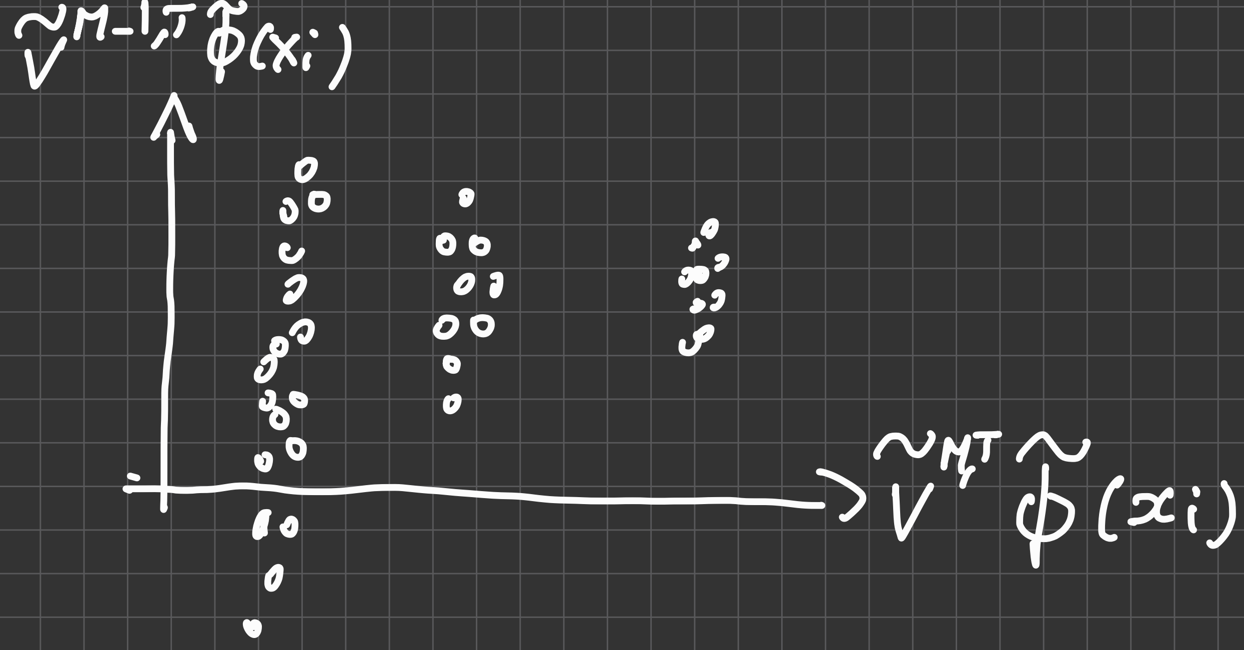
$$\Updownarrow$$

$$\tilde{K}^{test} = K^{test} - \frac{1}{M}\mathbb{1}K - \frac{1}{M}K^{test}\mathbb{1}^T + \frac{\mathbb{1}K\mathbb{1}^T}{M^2}$$

$\mathbb{1}$: $L \times M$ matrix $\mathbb{1}_{i,j} = 1$

Ex $N=2$





<u>Ex</u> Image denoising

x_i: images (MNIST)

x^*: noisy image

$$P_d \tilde{\phi}(x^*) = (\tilde{V}^{M-d+1})^T \tilde{\phi}(x^*)\ \tilde{V}^{M-d+1} + \cdots + (\tilde{V}^{M})^T \tilde{\phi}(x^*)\ \tilde{V}^{M}$$

projection of $\tilde{\phi}(x^*)$ onto
span eigenvectors associated with
d largest eigenvalues

x^*: noisy image

y^*: denoised image

$$y^* = \underset{Y \in \mathcal{X}}{\arg\min} \ \| \tilde{\phi}(y) - P_d \tilde{\phi}(x^*) \|_H$$

Kernel linear discriminant analysis

Fisher Linear discriminant analysis (LDA)

LDA

classification pb

C classes

Each class has N_i m-dim samples

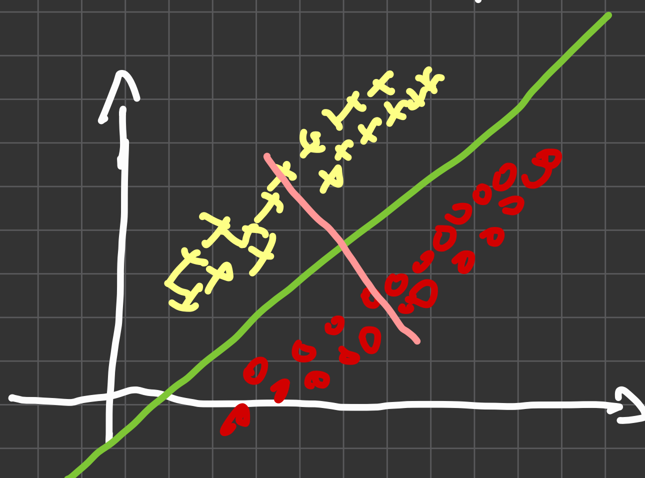
$i = 1, \dots, C$

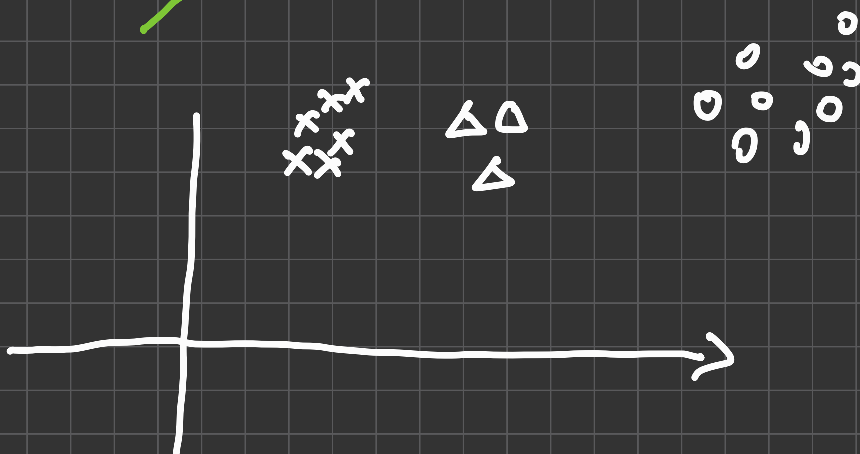
For each class i

samples: $\{ x_1^i, \dots, x_{N_i}^i \}$

$$X = m\underbrace{\begin{array}{cccccccc} x_1^1 & \cdots & x_{N_1}^1 & & x_1^i & \cdots & x_{N_i}^i & x_1^c & x_{N_c}^c \\ \end{array}}_{N_1 + \cdots + N_c = N}$$

→ seek to separate classes by projecting the x_j^i onto hyperplane of dim $C-1$





LDA with 2 classes

$N_1 + N_2 = N$ $\quad m$ dim samples $\{x_1, \ldots, x_N\}$

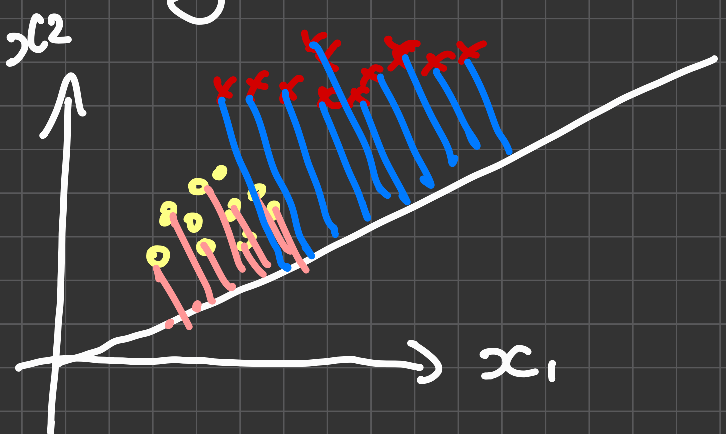
N_1: class 1

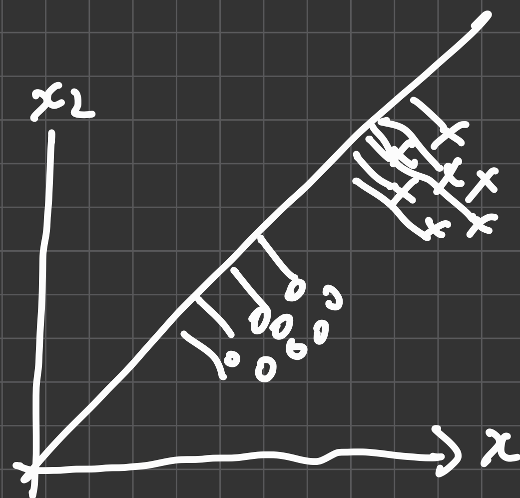
N_2: 2

seek obtain y by projecting x_i onto a line

(C-1 dim space)

$$y = w^T x$$





Need a good measure of separation

ω_i : set of points of class 1

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x$$

$$\tilde{\mu}_i = w^T \mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} w^T x$$

idea $J(w) = |\tilde{\mu}_1 - \tilde{\mu}_2|$

$$= |w^T(\mu_1 - \mu_2)|$$

scatter of class i

$$\tilde{s}_i^2 = \sum_{x \in \omega_i} (w^T x - \tilde{\mu}_i)^2$$

measure of variability within class i after projection

$$\tilde{s}_1^2 + \tilde{s}_2^2$$

Quantity to be maximized

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Need to find

$$w^* = \underset{w}{\operatorname{argmax}}\ J(w)$$

Scatter matrices

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T$$

$$S = S_1 + S_2$$

$$\tilde{s}_i^2 = w^T S_i w$$

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S w$$

Between class scatter matrix

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$w^T S_B w = |\tilde{\mu}_1 - \tilde{\mu}_2|$$

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{w^T S_B w}{w^T S w}$$

$$\frac{dJ}{dw} = 0 \iff (w^T S w) S_B w - (w^T S_B w) S w = 0$$

$$\iff S_B w - J(w) S w = 0$$

$\Leftrightarrow$ $$\underbrace{S_B w = \lambda S w}_{\text{generalized eigenvalue pb}} \qquad \lambda = J(w)$$

Sol

$$w^* = \operatorname{argmax} J(w) = \operatorname*{argmax}_w \frac{w^T S_B w}{w^T S w}$$

$$= S^{-1}(\mu_1 - \mu_2)$$

$\downarrow$ Fisher liner discriminant

LDA with C classes

C classes

seek $(C-1)$ projection vectors w_i

$$W = \underbrace{\begin{bmatrix} w_1 & \cdots & w_{C-1} \\ & & \end{bmatrix}}_{C-1} \Big\} m$$

$C-1$ projections: $y_1, \ldots, y_{C-1}$

$$y_i = w_i^T x$$

$$\Downarrow$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_{C-1} \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

$$x \in \mathbb{R}^m$$

$$y \in \mathbb{R}^{C-1}$$

$$y = W^T x$$

n data points $x^i \in \mathbb{R}^m$

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ & & & \vdots \\ x_m^1 & x_m^2 & & x_m^n \end{pmatrix}$$

$$Y = \begin{pmatrix} y_1^1 & y_1^2 & & y_1^n \\ \vdots & \vdots & & \vdots \\ y_{C-1}^1 & y_{C-1}^2 & & y_{C-1}^n \end{pmatrix}$$

$$Y = W^T X$$

$$S = \sum_{i=1}^{c} S_i$$

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T$$

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_i} x$$

N_i : # of data in class i

$$n = N = \sum_{i=1}^{c} N_i$$

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

$$\mu = \frac{1}{N} \sum_{i=1}^{c} \sum_{x \in \omega_i} x = \frac{1}{N} \sum_{i=1}^{c} N_i \mu_i$$

$$\tilde{S} = \mathrm{Tr}[W^T S W]$$

$$\tilde{S}_B = \mathrm{Tr}[W^T S_B W]$$

$$J(W) = \frac{\tilde{S}_B}{\tilde{S}} = \frac{Tr[W^T S W]}{Tr[W^T S_B W]}$$

need to find $w_1, \dots, w_{c-1}$ orthonormal $\Leftrightarrow$ $\left(w_j S w_i = 0 \quad i \neq j \right)$

that max $J(W)$

$\Leftrightarrow \quad S_B w_i = \lambda_i S w_i$

↳ $C-1$ largest eigenvalues

w_i : $(C-1)$ eigenvectors corresponding to the $C-1$ largest eigenvalues

$$w_j^T S_B w_i = \lambda_i w_j^T S w_i \Rightarrow w_j^T S w_i = 0$$

$$= \lambda_j w_j^T S w_i$$

Kernel LDA

2 classes

data in class 1: $\{x_1^1, \dots, x_{N_1}^1\} = \omega_1$

2: $\{x_1^2, \dots, x_{N_2}^2\} = \omega_2$

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

ϕ ↓ feature map

$\mathcal{H}$ ↓ feature space

$$m_i^\phi = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(x_j^i)$$

$$S_B^\phi = (m_1^\phi - m_2^\phi)(m_1^\phi - m_2^\phi)^T$$

$$S^\phi = \sum_{i=1}^{2} \sum_{x \in \omega_i} (\phi(x) - m_i^\phi)(\phi(x) - m_i^\phi)^T$$

Need to max.

$$J(w) = \frac{w^T S_B^\phi w}{w^T S^\phi w} \quad = \langle w, S^\phi w \rangle_{\mathcal{H}}$$

We can find an expansion for w of the form

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i)$$ (N data points $\{x_1, \dots, x_N\}$)

kernel $k(x, x') = \langle \phi(x), \phi(x') \rangle_H$

$$w^T m_i^\phi = \frac{1}{N_i} \sum_{j=1}^{N} \sum_{l=1}^{N_i} \alpha_j \, k(x_j, x_l^i)$$

$$= \alpha^T M_i$$

$$(M_i)_j = \frac{1}{N_i} \sum_{l=1}^{N_i} k(x_j, x_l^i)$$

$$w^T S_B^\phi w = \alpha^T M^B \alpha$$

$$M^B = (M_1 - M_2)(M_1 - M_2)^T$$

Similarly

$$w^T S^\phi w = \alpha^T M \alpha$$

$$M = \sum_{j=1}^{2} K_j \left(I - \frac{1}{N_j}\right) K_j^T$$

I: id. matrix $(N_j \times N_j)$

K_j: $N \times N_j$ matrix

$$(K_j)_{n,m} = k(x_n, x_m^j)$$

We can find Fisher's linear discrim. in H by maximizing

$$J(\alpha) = \frac{\alpha^T M^B \alpha}{\alpha^T M \alpha}$$

$\Rightarrow$ solved by finding leading eigenvector of

$$M^B \alpha = \lambda M \alpha$$

new data point x

$$\langle w, \phi(x) \rangle_H = w^T \phi(x) = \sum_{i=1}^{N} \alpha_i k(x_i, x)$$

In practice: to increase numerical stability

replace M by $M + \mu I$

$\mu > 0$ small

I id. matrix

Rk Kernel LDA $\Leftrightarrow$ perform Kernel PCA then perform LDA on resulting coordinates

Kernel Canonical Correlation Analysis

Canonical Correlation Analysis
(CCA, Hotelling 1936)

Data $(X_1, Y_1) \ldots (X_N, Y_N)$

$X_i \in \mathbb{R}^m$ iid $X_i \sim X$ (r.v)

$Y_i \in \mathbb{R}^p$ " $Y_i \sim Y$ (r.v.)

Find $a \in \mathbb{R}^m$, $b \in \mathbb{R}^p$ s.t. the correlation between $a^T X$ and $b^T Y$ is maximized

$$\rho = \max_{a,b} \mathrm{Cor}[a^T X, b^T Y]$$

$$= \max_{a,b} \frac{\mathrm{Cov}[a^T X, b^T Y]}{\sqrt{\mathrm{Var}[a^T X]\, \mathrm{Var}[b^T Y]}}$$

Replace unknown distribution of X
by empirical dist $\frac{\delta_{X_1} + \dots + \delta_{X_N}}{N}$

Replace " of Y
by " $\frac{\delta_{Y_1} + \dots + \delta_{Y_N}}{N}$

$$\mu_N^X = \frac{1}{N} \sum_{i=1}^{N} X_i \qquad \mu_N^Y = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

$$C_{XX} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu_N^X)(X_i - \mu_N^X)^T$$

$m \times m$ matrix

$$C_{X,Y} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu_N^X)(Y_i - \mu_N^Y)^T$$

$m \times l$ matrix

$$C_{YY} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \mu_N^Y)(Y_i - \mu_N^Y)^T$$

$$C_{Y,X} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \mu_N^Y)(X_i - \mu_N^X)^T$$

Sol. CCA

$$* \left| \begin{array}{ll} \max\limits_{a,b} & a^T C_{XY} b \\ \text{s.t.} & a^T C_{XX} a = 1 \qquad b^T C_{YY} b = 1 \end{array} \right.$$

Lagrange multiplier method

$\Updownarrow$ generalized eigenpb.

$$\begin{pmatrix} 0 & C_{XY} \\ C_{YX} & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \rho \begin{pmatrix} C_{XX} & 0 \\ 0 & C_{YY} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Solution

$$a = C_{XX}^{1/2} v_1 \qquad b = C_{YY}^{1/2} v_1$$

v_1: left first singular vector of the SVD of $C_{XX}^{-1/2} C_{XY} C_{YY}^{-1/2}$

v_1: right first singular vector of the SVD of $C_{XX}^{-1/2} C_{XY} C_{YY}^{-1/2}$

Kernel CCA

CCA for feature vectors

$\phi_x(X_1), \ldots, \phi_x(X_N) \in \mathcal{H}_X$ (feature space (RKHS))

$\phi_y(Y_1), \ldots, \phi_y(Y_N) \in \mathcal{H}_Y$ (RKHS)

$$\max_{\substack{f \in \mathcal{H}_X \\ g \in \mathcal{H}_Y}} \frac{\mathrm{Cov}[f(X), g(Y)]}{\sqrt{\mathrm{Var}[f(X)]\,\mathrm{Var}[g(Y)]}}$$

$$\mathrm{Cov}[f(X), g(Y)]$$
$$= \mathbb{E}_{X,Y}\Big[\big(f(X) - \mathbb{E}[f(X)]\big) \big(g(Y) - \mathbb{E}[g(Y)]\big) \Big]$$

$$\approx \frac{1}{N}\sum_{i=1}^{N}\left(f(X_i)-\frac{1}{N}\sum_{j=1}^{N} f(X_j)\right)$$
$$\left(g(Y_i)-\frac{1}{N}\sum_{j=1}^{N} g(Y_j)\right)$$

As in kernel PCA we can assume

$$f=\sum_{i=1}^{N}\alpha_i\,\tilde{\phi}_x(X_i) \qquad g=\sum_{i=1}^{N}\beta_i\,\tilde{\phi}_y(Y_i)$$

$$\tilde{\phi}_x(X_i)=\phi_x(X_i)-\frac{1}{N}\sum_{j=1}^{N}\phi_x(X_j)$$
$$\tilde{\phi}_y(Y_i)=\phi_y(Y_i)-\frac{1}{N}\sum_{j=1}^{N}\phi_y(Y_j)$$

$$\max_{\substack{f\in H_x\\ g\in H_y}} \frac{\mathrm{Cov}[f(X),g(Y)]}{\sqrt{\mathrm{Var}[f(X)]\,\mathrm{Var}[g(Y)]}}$$

$$\approx \max_{\substack{f\in H_x\\ g\in H_y}} \frac{\sum_{i=1}^{N}\langle f,\tilde{\phi}_x(X_i)\rangle\langle g,\tilde{\phi}_y(Y_i)\rangle}{\sqrt{\sum_{i=1}^{N}\langle f,\tilde{\phi}_x(X_i)\rangle^2\sum_{i=1}^{N}\langle g,\tilde{\phi}_y(Y_i)\rangle^2}}$$

As in kernel PCA we can assume

$$f = \sum_{i=1}^{N} \alpha_i \tilde{\phi}_X(X_i) \qquad g = \sum_{i=1}^{N} \beta_i \tilde{\phi}_Y(Y_i)$$

N

$\tilde{K}_X$ and $\tilde{K}_Y$: centered Gram matrices

$$\tilde{K}_{X,ij} = \langle \tilde{\phi}_X(X_i), \tilde{\phi}_X(X_j) \rangle$$

$$\tilde{K}_{Y,ij} = \langle \tilde{\phi}_Y(Y_i), \tilde{\phi}_Y(Y_j) \rangle$$

$$\Leftrightarrow \max_{\substack{\alpha \in \mathbb{R}^N \\ \beta \in \mathbb{R}^N}} \frac{\alpha^T \tilde{K}_X \tilde{K}_Y \beta}{\sqrt{\alpha^T \tilde{K}_X \alpha \; \beta^T \tilde{K}_Y \beta}}$$

- Regularization

$$\simeq \max_{\substack{f \in H_x \\ g \in H_y}} \frac{\sum_{i=1}^{N} \langle f, \tilde{\phi}_x(x_i)\rangle \langle g, \tilde{\phi}_y(y_i)\rangle}{\sqrt{\sum_{i=1}^{N} \langle f, \tilde{\phi}_x(x_i)\rangle^2 + \varepsilon_N \|f\|_{H_x}^2} \times \sqrt{\sum_{i=1}^{N} \langle g, \tilde{\phi}_y(y_i)\rangle^2 + \varepsilon_N \|g\|_{H_y}^2}}$$

converges as $N \to \infty$ (Fukumizu, 2007)

Solution: Generalized eigenproblem

$$\begin{pmatrix} 0 & \tilde{K}_x \tilde{K}_y \\ \tilde{K}_y \tilde{K}_x & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} & \\ & \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\begin{pmatrix} \tilde{K}_x^2 + \varepsilon_N \tilde{K}_x & 0 \\ 0 & \tilde{K}_y^2 + \varepsilon_N \tilde{K}_y \end{pmatrix}$$

Application: to image retrieval
(Hardon, 2004)

X_i: image
Y_i: corresponding text

Idea use d eigenvectors $f_1, \dots, f_d$
$g_1, \dots, g_d$

as feature spaces which contain the depend. between X & Y

Given a text Y_{new} compute its feature vector and the image X_i whose feature maximizes the following inner product

$$X_i \longrightarrow \tilde{\phi}_x(X_i) \longrightarrow \begin{pmatrix} \langle f_1, \tilde{\phi}_x(X_i) \rangle \\ \vdots \\ \langle f_d, \tilde{\phi}_x(X_i) \rangle \end{pmatrix} \cdot V$$

$$V = \begin{pmatrix} \langle g_1, \tilde{\phi}_Y(Y_{new}) \rangle \\ \vdots \\ \langle g_d, \tilde{\phi}_Y(Y_{new}) \rangle \end{pmatrix}$$

Ex $K_X \Leftrightarrow$ Gaussian kernel for images

$K_Y \Leftrightarrow$ Bag of words kernel (frequency of words) for texts

Support Vector Machines

Margin classifier

Training data $(X_1, Y_1), \ldots (X_N, Y_N)$

$X_i \in \mathbb{R}^d$

$Y_i \in \{-1, 1\}$ (binary labels)

Linear classifier

$$h(x) = \mathrm{sign}(w^T x + b)$$

$$w \in \mathbb{R}^d, \quad b \in \mathbb{R}$$

Select w, b so that a new data x is correctly classified

Hard margin

$$h(x) = \mathrm{sign}(w^T x + b)$$

Assumption: data is linearly separable

(w, b) and (cw, cb) $(c>0)$

give the same classifier

Fix the scale
$$\begin{cases} \min (w^T X_i + b) = 1 & \text{if } Y_i = 1 \\ \max (w^T X_i + b) = -1 & \text{if } Y_i = -1 \end{cases}$$

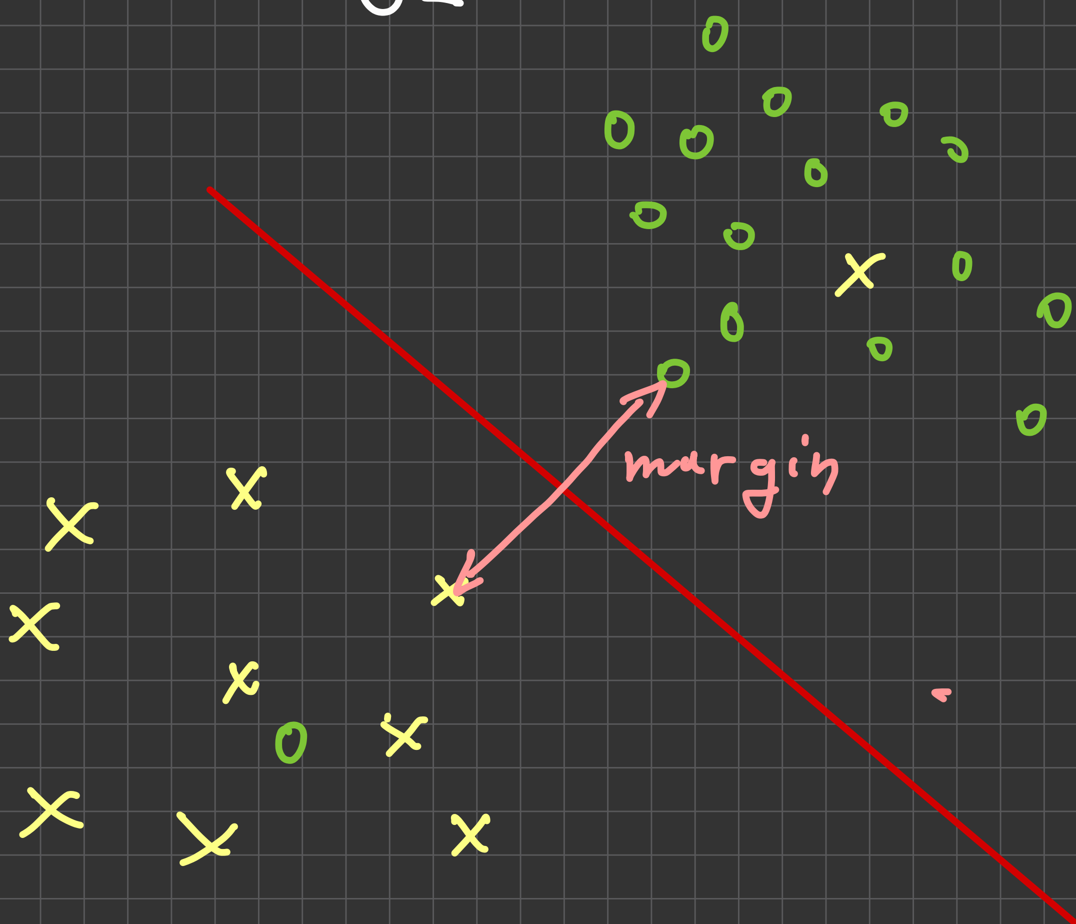
$$\Rightarrow \quad \text{margin} = \frac{2}{\|w\|}$$

$$\begin{cases} \max \dfrac{2}{\|w\|} & \\ \text{s.t.} \quad w^T X_i + b \geq 1 & \text{if } Y_i = 1 \\ \qquad w^T X_i + b \leq -1 & \text{if } Y_i = -1 \end{cases}$$

$$\begin{cases} \min \dfrac{\|w\|}{2} \\ \text{s.t.} \quad Y_i (w^T X_i + b) \geq 1 \quad \forall i \end{cases}$$

→ quadratic program

Soft margin



$$\min \frac{\|w\|^2}{2} + C \sum_{i=1}^{N} \xi_i$$

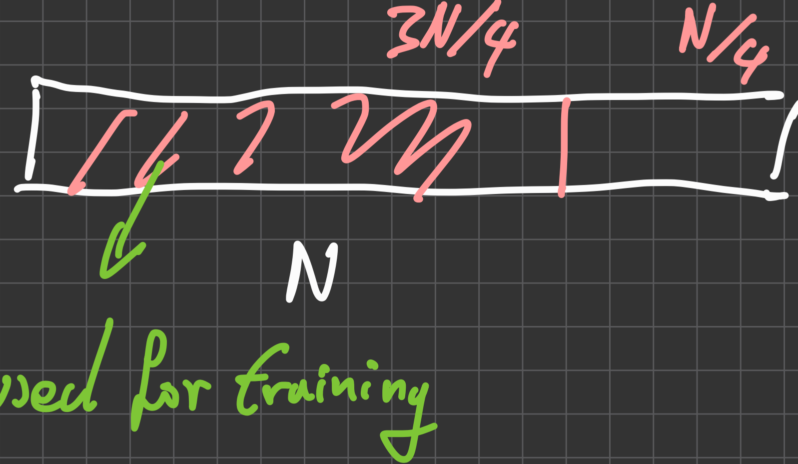
$$\text{s.t.} \quad y_i (w^T x_i + b) \geq 1 - \xi_i \quad \forall i$$

$$\xi_i \geq 0$$

Quadratic program

C → found via cross-validation

training data :



Regularization

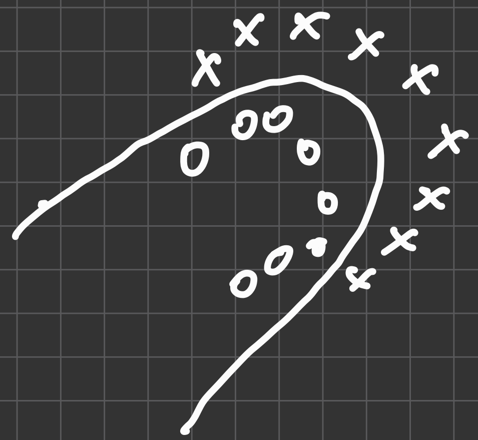
$$(*) \quad \min_{w,b} \underbrace{\sum_{i=1}^{n} \left(1 - y_i (w^T x_i + b)\right)_+}_{\text{Hinge loss}} + \lambda \underbrace{\|w\|^2}_{\text{regul.}}$$

proof The cost $\frac{\|w\|^2}{2} + C \sum_i \xi_i$

can only increase with $\xi_i \nearrow$

$\Rightarrow$ opt sol $\xi_i^* = \left(1 - y_i (w^T x_i + b)\right)_+$

$\Leftrightarrow (*)$



Kernelization of SVM

Training data $(X_1, Y_1), \ldots, (X_n, Y_n)$

$X_i \in \mathcal{X}$ (arbitrary)

$Y_i \in \{-1, 1\}$

Feature space H (RKHS)

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

kernel $k(x, x') = \langle \phi(x), \phi(x') \rangle_H$

Linear classifier on $H \iff$ nonlinear classifier on $\mathcal{X}$

$$h(x) = \operatorname{sign}\left(\langle f, \phi(x) \rangle_H + b \right)$$

$$f \in H, \quad b \in \mathbb{R}$$

$$h(x) = \text{sign}\,(f(x) + b)$$

Nonlinear SVM

$$\min_{f, b, \xi_i} \frac{\|f\|_H^2}{2} + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad Y_i\left(\langle f, \phi(X_i)\rangle_H + b\right) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

$$\Updownarrow$$

$$\min_{f, b} \sum_{i=1}^{n} \left(1 - Y_i\left(f(X_i) + b\right)\right)_+ + \lambda \|f\|_H^2$$

Representer thm

$$\Rightarrow$$

$$f = \sum_{i=1}^{n} w_i \phi(X_i) = \sum_{i=1}^{n} w_i K(\cdot, X_i)$$

$K(X,X)$: $n\times n$ matrix

$$\big(K(X,X)\big)_{i,j} = k(X_i, X_j) = \langle \phi(X_i), \phi(X_j) \rangle_H$$

$$\Leftrightarrow \min_{w,b,\xi} \frac{1}{2} w^T k(X,X) w + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t. } Y_i \big((k(X,X)w)_i + b \big) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

Lagrange multiplier method

the dual pb is

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j Y_i Y_j \big(k(X,X)\big)_{i,j}$$

$$\text{s.t. } 0 \leq \alpha_i \leq C$$

$$\sum_{i=1}^{n} y_i \alpha_i = 0$$

$\alpha^* = \arg\max$

only data with $0 < \alpha_i^* \leq C$

The classifier

$$f^*(x) + b^* = \sum_{i=1}^{n} \alpha_i^* y_i \, k(x, X_i) + b^*$$

$$b^* = y_i - \sum_{j=1}^{n} \alpha_j^* y_j \, k(X_i, X_j) \quad \forall i \,/\, 0 < \alpha_i^*$$

Indeed the Lagrangian associated with

$$\min_{w, b, \xi} \frac{1}{2} w^T k(X, X) w + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i \big((k(X, X) w)_i + b \big) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

is

$\Lambda(w, b, \xi, \alpha, \beta) =$

$$\frac{1}{2} w^T k(X, X) w + C \sum_{i=1}^{n} \xi_i$$

$$+ \sum_i \alpha_i \left(1 - \xi_i - y_i\left((k(X,X)w)_i + b\right)\right)$$

$$+ \sum_i \beta_i (-\xi_i)$$

The optimal w

$$\nabla_w \Lambda = 0 \Leftrightarrow k(X,X)w - \sum_i \alpha_i y_i \, k(X,X)_{:,i} = 0$$

$$\Rightarrow w_i = \alpha_i y_i$$

$$\nabla_b \Lambda = 0 \Rightarrow \sum_i \alpha_i y_i = 0$$

$$\nabla_{\xi_i} \Lambda = 0 \Rightarrow C - \alpha_i - \beta_i = 0 \Leftrightarrow 0 \leq \alpha_i \leq C$$

at the opt. in w, b, ξ_i

$$\min_{w,b,\xi} \Lambda = -\frac{1}{2}\sum_{i,j} (\alpha_i Y_i k(X_i, X_j) \alpha_j Y_j) + \sum_i \alpha_i$$

(KKT equations)

<u>Thm</u> The sol. of primal and dual svm is given by

(1) $1 - Y_i (f^*(X_i) + b^*) - \xi_i^* \leq 0 \quad \forall i$

(2) $\xi_i^* \geq 0 \quad \forall i$

(3) $0 \leq \alpha_i^* \leq C \quad \forall i$

(4) $\alpha_i^* (1 - Y_i (f^*(X_i) + b^*) - \xi_i^*) = 0 \quad \forall i$

(5) $\xi_i^* (C - \alpha_i^*) = 0$

(6) $$\sum_{j=1}^{n} k(X_i, X_j) w_j^* - \sum_{j=1}^{n} \alpha_j^* Y_j k(X_i, X_j) = 0$$

(7) $\sum_{j=1}^{n} \alpha_j^* y_j = 0$

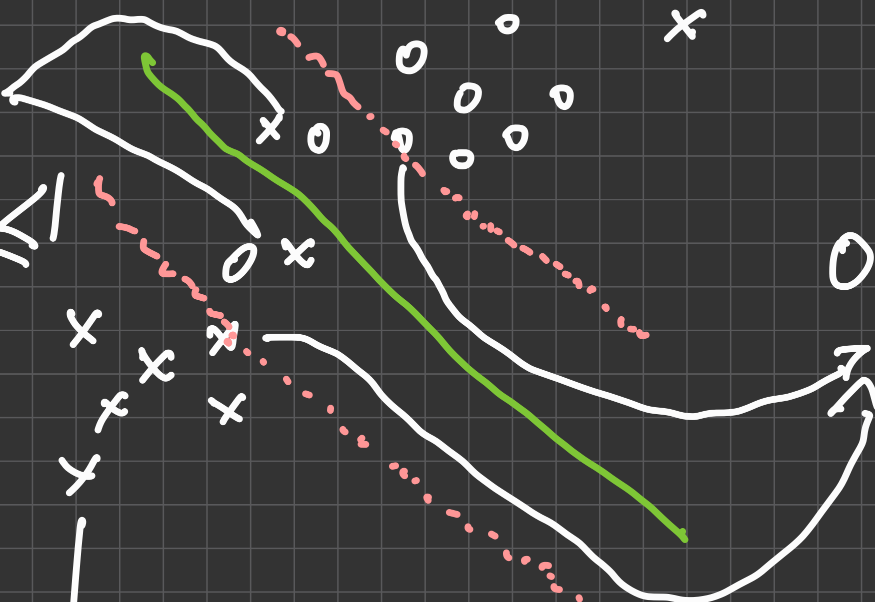
Sparse expression

$$\varphi^*(x) = f^*(x) + b^* = \sum \alpha_i^* y_i k(x, x_i) + b^*$$

x_i: support vectors

$\alpha_i^* = C$

$y_i \varphi^*(x_i) \leq 1$



$0 < \alpha_i < C$

$y_i \varphi^*(x_i) = 1$

$\alpha_i = 0$

Operator Valued Kernels

<u>Pb</u> $f^\dagger : \mathcal{X} \longrightarrow \mathcal{Y}$

$f^\dagger$: unknown

Given $f^\dagger(X) = Y$ approximate $f^\dagger$

$(X, Y) \in \mathcal{X}^N \times \mathcal{Y}^N$

$X = (X_1, \ldots, X_N)$

$Y = (Y_1, \ldots, Y_N)$

$f^\dagger(X) = (f^\dagger(X_1), \ldots, f^\dagger(X_N))$

$\mathcal{X}$: arbitrary

$\mathcal{Y}$: separable Hilbert space

$\dim(\mathcal{Y}) < \infty \rightarrow$ matrix-valued kernel (Alvarez et al 2012)

$\dim(\mathcal{Y}) = \infty \rightarrow$ operator-valued kernel

(Kadri et al., 2016)

$\langle \cdot, \cdot \rangle_{\mathcal{Y}}$: inner product on $\mathcal{Y}$

$\mathcal{L}(\mathcal{Y})$: set of bounded linear operators mapping $\mathcal{Y}$ to $\mathcal{Y}$

Def We call

$$K : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{Y})$$

an operator valued kernel if

(i) K is Hermitian, i.e.

$$K(x, x') = K(x', x)^T \quad \forall x, x' \in \mathcal{X}$$

A^T: adjoint of A w.r.to $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$

$$\langle Ay, y' \rangle_{\mathcal{Y}} = \langle y, A^T y' \rangle_{\mathcal{Y}}$$

(ii) $\sum_{i,j=1}^{m} \langle y_i, K(x_i, x_j) y_j \rangle_{\mathcal{Y}} \geq 0$

$\forall \ (x_i, y_i) \in X \times Y, \ m \in \mathbb{N}$

—

$\dim(Y) = s$

$K(x_i, x_j)$: $s \times s$ matrix

$$\begin{array}{cc} i & \begin{pmatrix} & & \\ & \boxed{} & \\ & & \end{pmatrix} \\ & \quad j \end{array}$$

K non degenerate $\Leftrightarrow \displaystyle\sum_{i,j=1}^{m} \langle y_i, K(x_i, x_j) y_j \rangle_Y = 0$

$\Downarrow$

$y_i = 0 \ \forall i$

if $x_i \neq x_j \ \forall i, j$

<u>Ex</u> $\dim(Y) = s$

$$K(x, x') = k(x, x') \underbrace{I_s}_{A} \quad SP$$

entangled kernels (2021)

RKHS

K non-degenerate, locally bounded separately continuous

$x \to K(x, x')$ continuous $\forall x'$

$\Updownarrow$

RKHS: $\mathcal{H}$

$$\mathcal{H} = \{ f : \mathcal{X} \to \mathcal{Y} \}$$

$$= \text{Closure Span} \left\{ z \to K(z, x) y \;\middle|\; \begin{smallmatrix} x \in \mathcal{X} \\ y \in \mathcal{Y} \end{smallmatrix} \right\}$$

Closure w.r. to inner product identified by reproducing property

$$\langle f, K(\cdot,x)y \rangle_{\mathcal{H}} = \langle f(x), y \rangle_{\mathcal{Y}}$$

Feature Maps

$\mathcal{F}$: separable Hilbert space

$(\langle \cdot, \cdot \rangle_{\mathcal{F}})$

$\mathcal{L}(\mathcal{Y}, \mathcal{F})$: set of bounded linear operators from $\mathcal{Y}$ to $\mathcal{F}$

<u>Def</u> We say that $\mathcal{F}$ and

$$\Psi : \mathcal{X} \longrightarrow \mathcal{L}(\mathcal{Y}, \mathcal{F})$$

are a feature space and a feature map for the kernel K if

$\forall\ x, x', y, y'$ $\quad y^T \Psi^T(x) \Psi(x') y'$

$\qquad\qquad\qquad\qquad\qquad \shortparallel$

$$y^T K(x, x') y' = \langle \Psi(x) y, \Psi(x') y' \rangle_{\mathcal{F}}$$

$\Psi^T(x)$: adjoint of $\Psi(x) \in \mathcal{L}(\mathcal{F}, \mathcal{Y})$

$$\langle \Psi(x) y, \alpha \rangle_{\mathcal{F}} = \langle y, \Psi^T(x) \alpha \rangle_{\mathcal{Y}}$$

$$\alpha^T \alpha' = \langle \alpha, \alpha' \rangle_{\mathcal{F}}$$

$$K(x, x') = \Psi^T(x)\ \Psi(x')$$

$\dim(\mathcal{Y}) = s$

$\dim(\mathcal{F}) = s'$

$$\Psi : \mathcal{X} \longrightarrow \mathcal{L}(\mathcal{Y}, \mathcal{F})$$

$\Psi(x)$: $s' \times s$ matrix

$\Psi_{i,j}(x)$ $\quad 1 \leq i \leq s'$

$\qquad\qquad 1 \leq j \leq s$

$$K(x,x') = \underbrace{\underset{s\times s'}{\Psi^T(x)} \quad \underset{s'\times s}{\Psi(x')}}_{s\times s \text{ matrix}}$$

$$K(x,x') = \sum_{i=1}^{s'} \Psi_i^T(x)\Psi_i(x')$$

$s'=1$

$\dim(\mathcal{F})=1$

$$\Psi : \mathcal{X} \to \mathcal{L}(\mathcal{Y}, \mathbb{R})$$

$$\left(K(x,x') = \quad \Psi(\quad \right)$$

- For $\alpha \in \mathcal{F}$

$$\Psi^T\alpha : \mathcal{X} \to \mathcal{Y}$$

$$x \to y \,/$$

$$\langle y'; y \rangle_{\mathcal{Y}} = \langle y'; \Psi^T(x)\alpha \rangle_{\mathcal{Y}} \quad \forall y'$$

<u>Thm</u> The RKHS $\mathcal{H}$ defined by $K(x,x') = \Psi^T(x)\Psi(x')$ is the linear span of $\Psi^T\alpha$ over $\alpha \in \mathcal{F}$ s.t. $\|\alpha\|_{\mathcal{F}} < \infty$

Furthermore

$$\langle \Psi^T(\cdot)\alpha, \Psi^T(\cdot)\alpha' \rangle_{\mathcal{H}} = \langle \alpha, \alpha' \rangle_{\mathcal{F}}$$

$$\|\Psi^T(\cdot)\alpha\|_{\mathcal{H}}^2 = \|\alpha\|_{\mathcal{F}}^2 \quad \forall \alpha, \alpha' \in \mathcal{F}$$

$$\Psi : \mathcal{X} \to \mathcal{L}(\mathcal{Y}, \mathcal{F})$$

$$\Psi^T : \mathcal{X} \to \mathcal{L}(\mathcal{F}, \mathcal{Y})$$

$$\mathcal{H} = \{ x \mapsto \Psi^T(x)\alpha \}$$

$\dim(Y) = s$

$\dim(\mathcal{F}) = s'$

$$\mathcal{H} = \left\{ \begin{pmatrix} | \\ \sum_j \psi_{i,j}^T(x)\,\alpha_j \\ | \end{pmatrix} \rightarrow i \right.$$

$$\begin{pmatrix} \sum_j \psi_{1,j}^T(x)\alpha_j \\ \\ \sum_j \psi_{s,j}^T(x)\alpha_j \end{pmatrix}$$