Lecture 4

Connection with Kernel Ridge Regression

$\hat{f}$ (t_i, y_i) $i \in \{1, \dots, n\}$

$$\hat{f} = \underset{f}{\operatorname{argmin}} \Big\{ \underbrace{\|y - f\|_{2,n}^2}_{\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2} + \xi \underbrace{\|f\|_K^2}_{\sum_{m=0}^{\infty} \frac{\langle f, \varphi_m\rangle^2}{\lambda_m^2}} \Big\}$$

We have in mind application to

$t_\ell = T \cdot \frac{\ell - 1}{n}$, K periodic kernel

$f : [0, T] \to \mathbb{C}$ periodic.

→ Much of what we will say is actually more general.

→ Recall that in our example instead

of $m \in \mathbb{N}$, it is convenient to index eigenfunction by $q \in \mathbb{Z}$.
Of course the two are equiv.

Writing $f(t) = \sum_{m=1}^{\infty} \theta_m \varphi_m(t)$

$$\cancel{\hat{\theta} = \operatorname{argmin} \left\{ \frac{1}{n} \| y - \Phi \theta \|^2 \right.}$$

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \left\{ \frac{1}{n} \left\| y - \sum_{m=1}^{\infty} \theta_m \varphi_m \right\|_2^2 + \xi \sum_{m=1}^{\infty} \frac{|\theta_m|^2}{\lambda_m^2} \right\}$$

$$\Phi = \begin{bmatrix} \varphi_1(t_1) & \varphi_2(t_1) & \\ \vdots & \vdots & \cdots \\ \varphi_1(t_m) & \varphi_2(t_m) & \end{bmatrix} \in \mathbb{R}^{n \times \infty}$$

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \left\{ \frac{1}{n} \| y - \Phi \theta \|^2 + \xi \langle \theta, \Lambda^{-2} \theta \rangle \right\}$$

$$\hat{\theta} = \left(\xi\Lambda^{-2} + \frac{1}{n}\Phi^*\Phi\right)^{-1}\frac{1}{n}\Phi^* Y$$

$$= \frac{1}{n}\Lambda^2\Phi^*\left(\xi I + \frac{1}{n}\Phi\Lambda^2\Phi^*\right)^{-1} Y$$

$$\cancel{\hat{f}(x) = \sum_{m=1}^{\infty}}$$

$$\left(\Phi\Lambda^2\Phi^*\right)_{ij} = \sum_{m=1}^{\infty}\lambda_m^2\varphi_m(t_i)\varphi_m^*(t_j)$$

$$= \cancel{K(t_j, t_i)}\ K(t_i t_j)$$

$$\hat{\mathbb{K}}_n := \left(K(t_i t_j)\right)_{ij \le n}$$

$$\hat{f}(t) = \frac{1}{n}\hat{h}_n(t)^*\left(\xi I + \frac{\hat{\mathbb{K}}_n}{n}\right)^{-1} Y$$

$$\hat{h}_n(t) = \begin{pmatrix} k(t_1, t) \\ \vdots \\ k(t_n, t) \end{pmatrix}$$

[Parenthesis not covered in class }

Compute test error / estimation error

$$y_i = f(t_i) + \sigma z_i \qquad z_i \sim N(0,1)$$

$$\hat{f}(t) = \frac{1}{n}\hat{h}_n(t)^* \left(\xi I_n + \frac{1}{n}\hat{K}_n\right)^{-1} f_n$$

$$+ \frac{\sigma}{n}\hat{h}_n(t)^* \left(\xi I_n + \frac{1}{n}\hat{K}_n\right)^{-1} z$$

$$\mathbb{E}\{(\hat{f}(t) - f(t))^2\} = \left[f(t) - \frac{1}{n}\hat{h}_n^* \left(\xi I + \frac{1}{n}\hat{K}_n\right)^{-1} f_n\right]^2$$

$$+ \frac{\sigma^2}{n^2}\hat{h}_n^*(t)\left(\xi I_n + \frac{1}{n}\hat{K}_n\right)^{-2}\hat{h}_n(t)$$

$$K^{2}(t,s) := \int_0^T k(t,r)K(r,s)\,dr$$

$$E\{\|\hat{f}-f\|_{L^2}^2\} = \int_0^T \left[f(t) - \frac{1}{n}K(t,\cdot)(\xi I + \frac{\hat{K}_n}{n})^{-1} f_n\right]^2 dt$$

$$+ \frac{\sigma^2}{n^2} \mathrm{Tr}\left(\hat{K}_n^2 (\xi I + \frac{1}{n}\hat{K}_n)^{-2}\right) \Big]$$

~~Intuitive picture~~

Eigenvalues / eigenvectors of

: $\hat{K}_n$ $\cancel{\varphi_m^{(n)} = \frac{1}{\sqrt{n}} \varphi_m(t) \ldots}$

$\cancel{\hat{K}_n [\varphi_m(t_i)]_{i \le n}}$

$$\varphi_m^{(n)} = \sqrt{\frac{T}{n}} \left[\varphi_m(t_1) \ldots \varphi_m(t_n)\right]^T$$

$$\|\varphi_m^{(n)}\|^2 \approx \frac{T}{n} \sum_{i=1}^n |\varphi_m(t_i)|^2 \approx$$

$$\approx T \frac{1}{T} \int_0^T |\varphi_m(t)|^2 dt \approx 1.$$

exact equality for our example

$$\left(\hat{K}_n \varphi_m^{(n)}\right)_i = \sqrt{\frac{T}{n}} \sum_{\ell=1}^{n} K(t_i,t_\ell)\,\varphi_m(t_\ell)$$

$$\approx \frac{\sqrt{Tn}}{T}\int_0^T K(t_i,t)\,\varphi_m(t)\,dt =$$

ality

ur

e

$$= \sqrt{\frac{n}{T}}\,\lambda_m^2\, \varphi_m(t_i) = \frac{n}{T}\,\lambda_m^2 \left(\varphi_m^{(n)}\right)_i$$

$$\frac{1}{n}\hat{K}_n \approx \frac{1}{T}\sum_{q=1}^{n} \lambda_q^2\, \varphi_q^{(n)}\, \varphi_q^{(n)\,*}$$

equality in our case

set for simplicity $T=1$. (rescaling in ξ)

$$\hat{f}(t) = \sum_{m\in S_n}^{\infty}$$

these are

$\frac{1}{n}\sum_{i=1}^{n} y_i \varphi_m(t_i)$

In general

$$\hat{f}(t) \approx \sum_{q=1}^{n} \frac{\lambda_q^2}{\xi + \lambda_q^2} \varphi_q(t) \cdot \overbrace{\frac{\langle y, \varphi_q^{(n)} \rangle}{\sqrt{n}}}$$

Equality for our case. assuming λ_m decr.

$$\hat{f}(t) = \sum_{q \in S_n} \frac{\lambda_q^2}{\xi + \lambda_q^2} \cancel{\varphi_n}\, \varphi_q(t) \underbrace{\frac{\langle y\ \varphi_q^{(n)} \rangle}{\sqrt{n}}}_{\theta_q + \sigma g_q + \alpha_q}$$

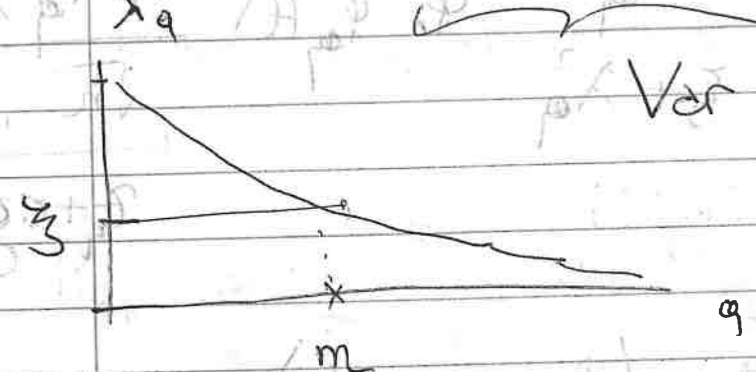
$$\hat{f}_q(t) = \sum_{q \in S_n} \frac{\lambda_q^2}{\xi + \lambda_q^2} \left(\theta_q + \frac{\sigma g_q}{\sqrt{n}} + \alpha_q\right) \cdot \varphi_q(t)$$

$$\cancel{\mathbb{E}\|\hat{f} - f\|_{L^2}^2 = \sum_{q \notin S_n} |\theta_q|^2 + \sum_{q \in S_n} \left(\frac{\xi}{\xi + \lambda_q^2}\right)^2 |\theta_q|^2}$$

$$\cancel{+ \sum_{q \in S_n} \left(\frac{\lambda_q^2}{\xi + \lambda_q^2}\right) |\alpha_q|^2 +}$$

$$E\left(\|\hat{f}-f\|_{L^2}^2\right) = \sum_{q\notin S_n} |\theta_q|^2 + \sum_{q\in S_n}\left(\frac{-\xi}{\xi+\lambda_q^2}\theta_q + \frac{\lambda_q^2}{\xi+\lambda_q^2}\alpha_q\right)^2$$

$$+ \underbrace{\frac{\sigma^2}{n}\sum_{q\in S_n}\left(\frac{\lambda_q^2}{\lambda_q^2+\xi}\right)^2}_{\text{Var}}$$



Take m so that $\lambda_m \geq \xi \geq \lambda_{m+1}$

Further assume $|\lambda_{\ell+1} - \lambda_\ell| \ll \lambda_\ell$

$$\text{Var} \leq \frac{\sigma^2 m}{n} + \frac{\sigma^2}{n}\underbrace{\sum_{q\geq m}\left(\frac{\lambda_q}{\lambda_m}\right)^4}_{(*)}$$

$$(*) \approx \sum_{q=m}^{\infty} \frac{q^{-4s}}{m^{-4s}} \leq Cm \quad \text{for } s \geq 1/4$$

$$\mathrm{Var} \leq \frac{C\sigma^2 m}{n}$$

The bias term can be analyzed similarly

White noise ~~problem~~ model (filtering)

$\|f_*\|_k < \infty \qquad f_* : [0,T] \to \mathbb{C}$

periodic RKHS

Observe

$$dY_t = f_*(t)\,dt + \frac{\sigma}{\sqrt{n}}\,dB_t$$

$B_t = (B_t^1 + iB_t^2)/\sqrt{2}$ complex BM

Will assume wlog $T = 1$.

Reason for the scaling?