

协方差函数

黄政宇

北京大学北京国际数学研究中心

北京大学国际机器学习研究中心



本堂课大纲

- 核函数性质
- 各向同性的核函数
 - 平方指数(squared exponential)类核函数
 - Matern类核函数
 - 具有紧致支集的分片多项式核函数的
- 非平稳核函数
- 混合核函数
- 函数的特征函数分析



监督学习

➤ 监督学习

- 训练数据 $\{(x_i, f(x_i)) | i = 1, 2, \dots, n\}$
- 测试点 $\{x_i^* | i = 1, 2, \dots, n_*\}$

通常情况下，当 x 和 x' 接近，那么 $f(x)$ 也和 $f(x')$ 接近。
 x 和 x' 的接近程度，由核函数 $\kappa(x, x')$ 刻画。



核函数

➤ 核函数 $\kappa: R^N \times R^N \rightarrow R$

线性回归 $\kappa(x, x') = \sigma_0^2 + x \cdot x'$

➤ 平稳的核函数

$$\kappa(x, x') = \kappa(x - x')$$

➤ 对称的核函数

$$\kappa(x, x') = \kappa(x', x)$$

➤ 各向同性的核函数

$$\kappa(x, x') = \kappa(\|x - x'\|)$$

➤ 正定的核函数

$$\int \kappa(x, x') f(x) f(x') d\mu(x) d\mu(x') \geq 0 \quad \forall f \in L_2(\mu)$$



平稳的核函数

➤ 弱平稳(weakly stationary)过程

期望是常数，协方差函数有平移不变性

➤ 均方连续(mean square continuity)

当 $\|x_k - x_*\| \rightarrow 0$, 那么 $\mathbb{E}\|f(x_k) - f(x_*)\|^2 \rightarrow 0$

$$\Leftrightarrow \kappa(x_k, x_k) - 2\kappa(x_k, x_*) + \kappa(x_*, x_*) \rightarrow 0$$

$$\Leftrightarrow 2\kappa(0) - 2\kappa(x_k - x_*) \rightarrow 0$$

➤ 均方可微(mean square differentiability)

$$\frac{\partial f(x)}{\partial x_i} = l.i.m._{h \rightarrow 0} \frac{f(x+he_i) - f(x)}{h} \quad (\text{均方收敛})$$

k 阶均方可微 \Leftrightarrow

$2k$ 阶偏导数 $\frac{\partial^{2k} \kappa(x)}{\partial^2 x_{i_1} \partial^2 x_{i_2} \cdots \partial^2 x_{i_k}}$ 在 $x = 0$ 存在且有限



平稳的核函数

Bochner定理

一个定义在 \mathbb{R}^N 中的复值核函数，它能定义一个复值的弱平稳的均方连续过程，当且仅当 κ 能表示为

$$\kappa(\tau) = \int_{\mathbb{R}^N} e^{2\pi i s \cdot \tau} d\mu(s)$$

其中 μ 是有限正测度。如果 μ 有密度 ρ

$$\kappa(\tau) = \int_{\mathbb{R}^N} e^{2\pi i s \cdot \tau} \rho(s) ds$$

我们有

$$\text{Fourier特征: } \kappa(x, x') = \mathbb{E}_{s \sim \rho(s)} [e^{2\pi i s \cdot x} e^{-2\pi i s \cdot x'}]$$

$$\rho(s) = \int_{\mathbb{R}^N} e^{-2\pi i s \cdot \tau} \kappa(\tau) d\tau \implies \kappa(\tau) \text{ 的光滑程度}$$



本堂课大纲

- 核函数性质
- 各向同性的核函数
 - 平方指数(squared exponential)类核函数
 - Matern类核函数
 - 具有紧致支集的分片多项式核函数的
- 非平稳核函数
- 混合核函数
- 函数的特征函数分析



各向同性的核函数

➤ 平方指数(squared exponential)类核函数

$$\kappa(r) = \exp\left(-\frac{r^2}{2l^2}\right)$$

l : 特征长度尺度(characteristic length-scale)

平方指数类核函数对应的过程 $f(x)$ 无穷阶可微。

密度函数: $\rho(s) = (2\pi l^2)^{N/2} \exp(-2\pi^2 l^2 s^2)$

基底函数:

$$\phi_c(x) = \exp\left(-\frac{(x-c)^2}{2l^2}\right)$$

$$\kappa(x, x') = \sigma^2 \int_{-\infty}^{\infty} \exp\left(-\frac{(x-c)^2}{2l^2}\right) \exp\left(-\frac{(x'-c)^2}{2l^2}\right) dc$$

$$= \sqrt{\pi} l \sigma^2 \exp\left(-\frac{(x-x')^2}{4l^2}\right)$$



各向同性的核函数

➤ Matern类核函数

$$\kappa(r; \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l} \right)^\nu B_\nu \left(\frac{\sqrt{2\nu}r}{l} \right)$$

l : 特征长度尺度(characteristic length-scale)

Γ : Gamma 函数

B_ν : 改版的Bessel 函数(不震荡)

Matern类核函数对应的过程 $f(x)$ 是 k 次均方可微, 当且仅当 $\nu > k$

密度函数:

$$\rho(s) = \frac{2^d \pi^{\frac{d}{2}} \Gamma\left(\nu + \frac{d}{2}\right) (2\nu)^\nu}{\Gamma(\nu) l^{2\nu}} \left(\frac{2\nu}{l^2} + 4\pi^2 s^2 \right)^{-(\nu + \frac{d}{2})}$$



各向同性的核函数

➤ Matern类核函数

$$\kappa(r; \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l} \right)^\nu B_\nu \left(\frac{\sqrt{2\nu}r}{l} \right)$$

当 $\nu \rightarrow \infty$

密度函数：

$$\begin{aligned} \rho(s) &= \frac{2^d \pi^{\frac{d}{2}} \Gamma\left(\nu + \frac{d}{2}\right) (2\nu)^\nu}{\Gamma(\nu) l^{2\nu}} \left(\frac{2\nu}{l^2} + 4\pi^2 s^2 \right)^{-(\nu + \frac{d}{2})} \\ &\sim \exp(-2\pi^2 s^2 l^2) \end{aligned}$$

$$\kappa(r; \nu \rightarrow \infty) \sim \exp\left(-\frac{r^2}{2l^2}\right)$$



各向同性的核函数

➤ Matern类核函数

$$\begin{aligned} & \kappa\left(r; \nu = p + \frac{1}{2} \quad p \in \mathbb{N}\right) \\ &= \exp\left(-\frac{\sqrt{2\nu}r}{l}\right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^p \frac{(p+i)!}{i!(p-i)!} \left(\frac{\sqrt{8\nu}r}{l}\right)^{p-i} \end{aligned}$$

机器学习中常用：

$$\kappa\left(r; \frac{3}{2}\right) = \exp\left(-\frac{\sqrt{3}r}{l}\right) \left(1 + \frac{\sqrt{3}r}{l}\right)$$

$$\kappa\left(r; \frac{5}{2}\right) = \exp\left(-\frac{\sqrt{5}r}{l}\right) \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right)$$

.....



各向同性的核函数

➤ Matern类核函数

$$\kappa\left(r; \frac{1}{2}\right) = \exp\left(-\frac{r}{l}\right)$$

此核函数对应的过程 $f(x)$ 均方连续但是不均方可微

当 $d = 1$ ，对应平稳Ornstein-Uhlenbeck 过程

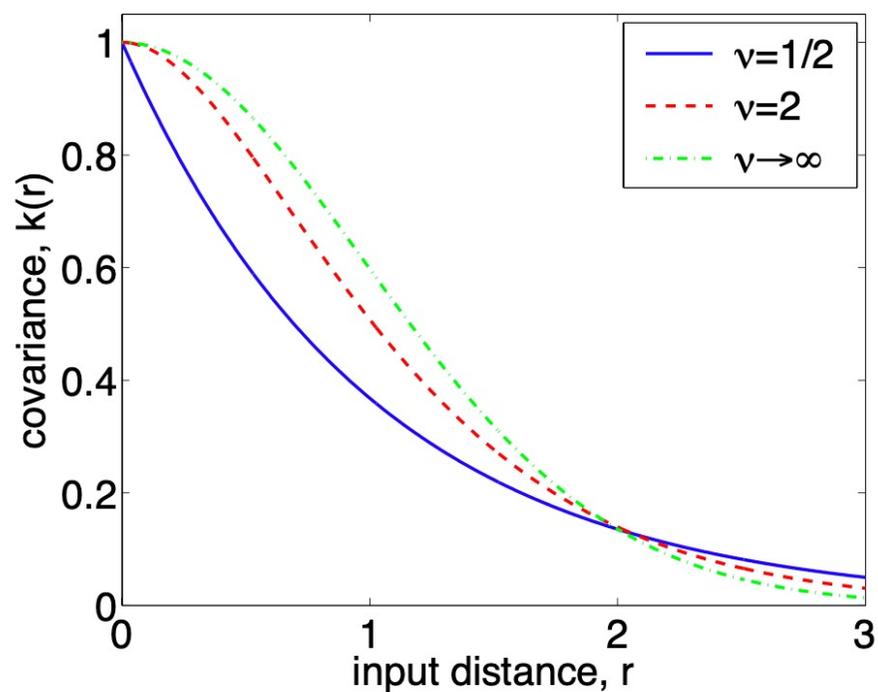
$$dx_t = -x_t dt + dW_t \quad x_0 \sim \mathcal{N}(0,1)$$

$$\frac{1}{2} \longleftarrow \nu \longrightarrow \infty$$
$$\exp\left(-\frac{r^2}{2l^2}\right) \longleftarrow \kappa(r; \nu) \longrightarrow \exp\left(-\frac{r^2}{2l^2}\right)$$

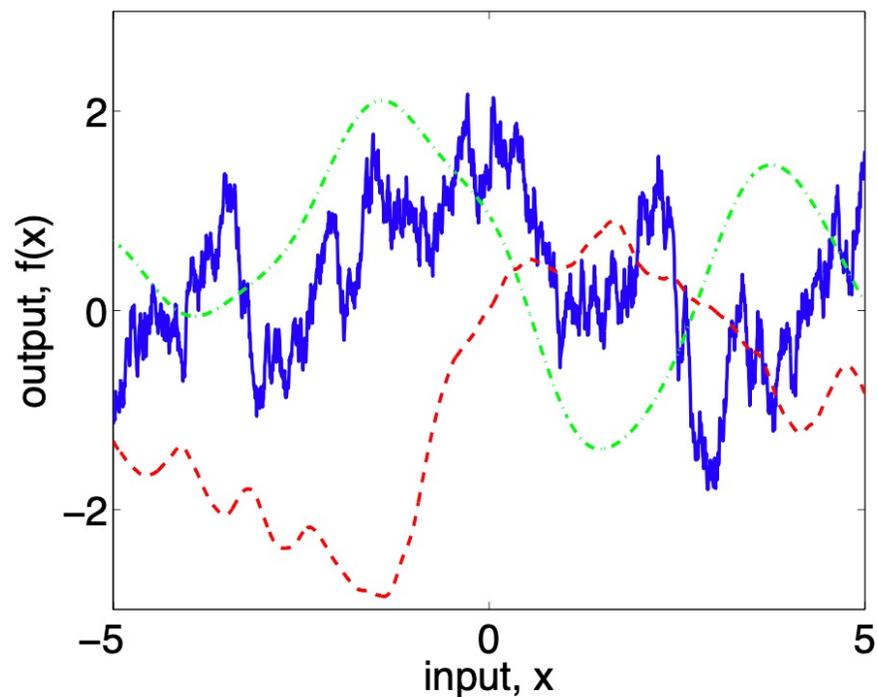


各向同性的核函数

➤ Matern类核函数



(a)



(b)



各向同性的核函数

➤ 具有紧致支集的分片多项式核函数

$$\kappa_{d,q}(r; j) \quad \text{其中 } j = \lfloor d/2 \rfloor + q + 1$$

$$\kappa_{d,0}(r) = (1 - r)_+^j$$

$$\kappa_{d,1}(r) = (1 - r)_+^{j+1} ((j + 1)r + 1)$$

$$\kappa_{d,2}(r) = (1 - r)_+^{j+2} \frac{((j^2 + 4j + 3)r^2 + (3j + 6)r + 3)}{3}$$

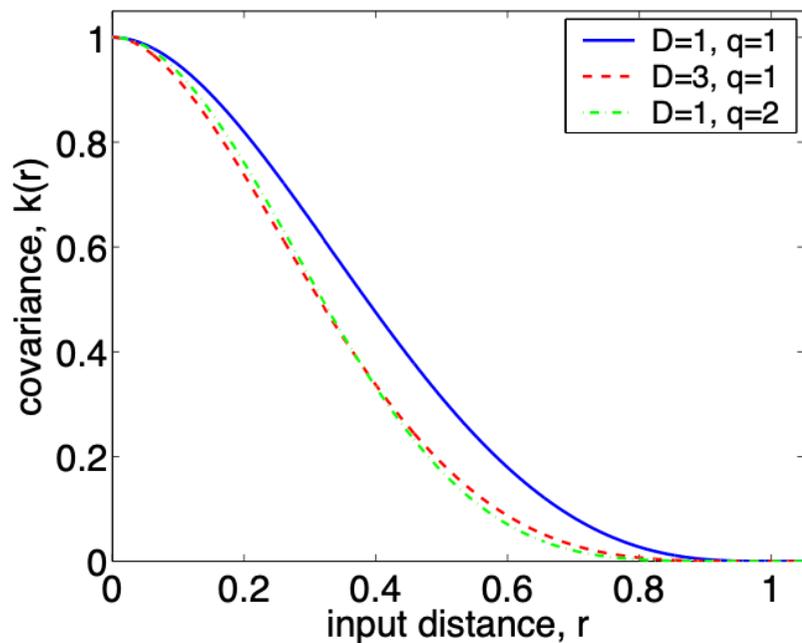
这些核函数是正定核函数，对应的过程 $f(x)$ q 次均方可微

计算优势：协方差矩阵稀疏

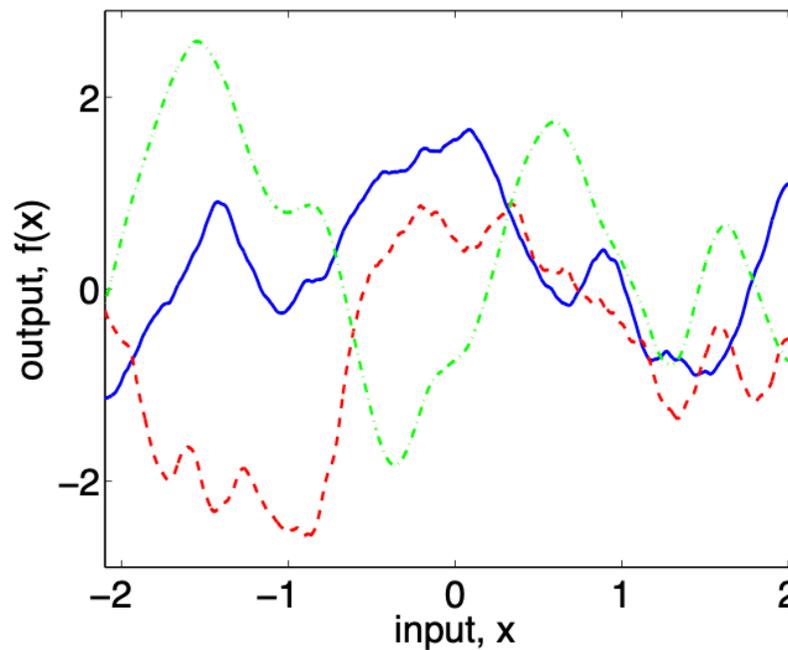


各向同性的核函数

➤ 具有紧致支集的分片多项式核函数



(a)



(b)



非平稳核函数

多项式核函数

线性回归： $\kappa(x, x') = \sigma_0^2 + x \cdot x'$
 $\kappa(x, x') = (\sigma_0^2 + x \Sigma_0 x')^p$

基底函数：

$$\phi_c(x) = x_{i_1} x_{i_2} \cdots x_{i_p} \text{ (} p \text{次多项式)}$$

$$\begin{aligned} \kappa(x, x') &= (x \cdot x')^p = \left(\sum_{i=1}^d x_i x'_i \right)^p \\ &= \left(\sum_{i_1=1}^d x_{i_1} x'_{i_1} \right) \left(\sum_{i_2=1}^d x_{i_2} x'_{i_2} \right) \cdots \cdots \left(\sum_{i_p=1}^d x_{i_p} x'_{i_p} \right) \\ &= \sum_{i_1=1}^d \cdots \sum_{i_p=1}^d \left(x_{i_1} x_{i_2} \cdots x_{i_p} \right) \left(x'_{i_1} x'_{i_2} \cdots x'_{i_p} \right) \\ &= \phi(x)^T \phi(x') \end{aligned}$$



非平稳核函数

➤ 神经网络

$$f(x) = b + \sum_{j=1}^D v_j h(x, \theta_j)$$

v_j : 权重

$h(x, \theta_j)$: 隐层单元的传递函数，取决于输入到隐层的权重 θ_j ，比如 $h(x, \theta_j) = \tanh(x^T \theta_j)$ ， $(x^T \theta_j)_+$

这种架构之所以重要，是因为Hornik在1993年的研究中已经证明，对于一大类传递函数（但不包括多项式），当隐层单元的数量趋于无穷大时，具有一个隐层的网络是通用逼近器。



非平稳核函数

➤ 神经网络

$$f(x) = b + \sum_{j=1}^D v_j h(x, \theta_j)$$

假设 b ， v_j 独立，期望为 0，方差分别为 σ_b^2 和 σ_v^2 ， θ_j 也是独立同分布，记录所有参数为 w ，那么

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

$$\begin{aligned} \mathbb{E}_w[f(x)f(x')] &= \sigma_b^2 + \sigma_v^2 \sum_j \mathbb{E}_w[h(x, \theta_j)h(x', \theta_j)] \\ &= \sigma_b^2 + \sigma_v^2 D \mathbb{E}_\theta[h(x, \theta)h(x', \theta)] \quad (\sigma_v^2 D \rightarrow \omega^2) \\ &= \sigma_b^2 + \omega^2 \mathbb{E}_\theta[h(x, \theta)h(x', \theta)] \end{aligned}$$

当 D 足够大， $\sigma_v^2 \sim \frac{\omega^2}{D}$ ， $f(x)$ 将趋于高斯过程。



混合核函数

- 两个核函数之和还是核函数

$$f(x) = f_1(x) + f_2(x)$$
$$\kappa(x, x') = \kappa_1(x, x') + \kappa_2(x, x')$$

- 两个核函数之积还是核函数

$$f(x) = f_1(x)f_2(x)$$
$$\kappa(x, x') = \kappa_1(x, x')\kappa_2(x, x')$$

- 核函数归一化

$$\tilde{f}(x) = \kappa(x, x)^{-1/2} f(x)$$
$$\tilde{\kappa}(x, x') = \frac{\kappa(x, x')}{\sqrt{\kappa(x, x)\kappa(x', x')}}}$$

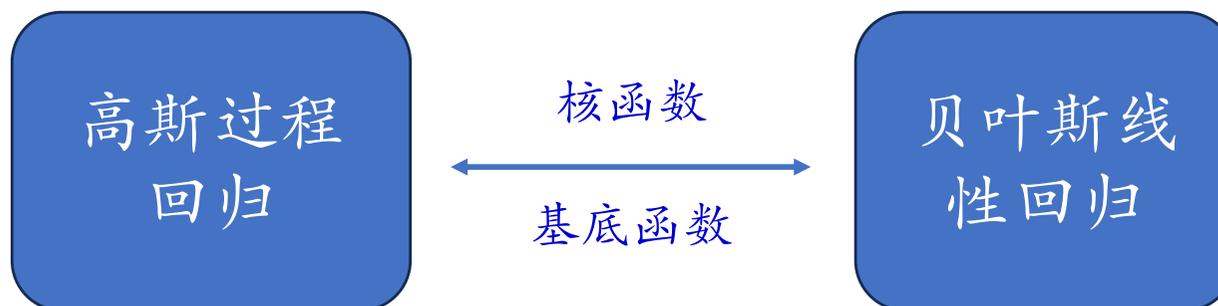


本堂课大纲

- 核函数性质
- 各向同性的核函数
 - 平方指数(squared exponential)类核函数
 - Matern类核函数
 - 具有紧致支集的分片多项式核函数的
- 非平稳核函数
- 混合核函数
- 函数的特征函数分析



核函数的特征函数分析





核函数的特征函数分析

➤ 特征函数

$$\int \kappa(x, x') \phi(x) d\mu(x) = \lambda \phi(x')$$

$\phi(x)$ 是核函数 $\kappa(x, x')$ 对应测度 $\mu(x)$ 的特征函数。通常情况，我们有无数个特征函数 $\phi_1(x), \phi_2(x), \dots$ 它们满足

$$\int \phi_i(x) \phi_j(x) d\mu(x) = \delta_{ij}$$

它们对应的特征值为 $\lambda_1 \geq \lambda_2 \geq \dots$

通常我们感兴趣的测度是 1) 紧致支集上的 2) $d\mu(x) = \rho(x)dx$ 。



核函数的特征函数分析

Mercer定理

(R^N, μ) 是有限测度空间，对称正定核函数 κ 定义

$$T_\kappa f(x) = \int_{R^N} \kappa(x, x') f(x') d\mu(x')$$

$T_\kappa : L_2(R^N, \mu) \rightarrow L_2(R^N, \mu)$ 。 $\phi_i \in L_2(R^N, \mu)$ 是算子 T_κ 关于正特征值 $\lambda_i > 0$ 的归一化特征函数，那么

- 特征值 $\{\lambda_i\}_{i=1}^\infty$ 是绝对可加的
- $\kappa(x, x') = \sum_{i=1}^\infty \lambda_i \phi_i(x) \phi_i^*(x')$ 在 μ^2 意义下几乎处处成立，其中级数几乎处处绝对收敛且一致收敛。

$\phi_1(x) \rightarrow \phi_\infty(x)$ 从低频到高频

κ 的光滑程度 $\leftrightarrow \lambda_i$ 的衰减率



核函数的特征函数分析

退化的核

一个退化的核仅有有限个非零特征值。

➤ 例子

在特征空间中的 D 维线性回归模型会产生一个退化核，该核至多有 D 个非零特征值。

如果测度仅在特征空间 R^N 中的有限 n 个点上赋予权重，那么特征分解就简单地变成了一个 $n \times n$ 的矩阵的特征分解，即使该核是非退化的。



平稳的核函数

➤ 例子

Bochner 定理

$$\kappa(x - x') = \int_{\mathbb{R}^N} e^{2\pi i s \cdot x} (e^{2\pi i s \cdot x'})^* \rho(s) ds$$

Mercer 定理

$$\kappa(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i^*(x')$$

对于平稳核函数，特征函数： $e^{2\pi i s \cdot x}$ ，特征值： $\rho(s)$
 $\rho(s)$ 决定了光滑程度。



核函数的特征函数分析

➤ 例子

考虑核函数 $\kappa(x, x') = \exp\left(-\frac{(x-x')^2}{2l^2}\right)$,
密度函数 $p(x) = \mathcal{N}(x; 0, \sigma^2)$

➤ 特征值特征函数

$$\lambda_k = \sqrt{\frac{2a}{A}} B^k$$

$$\phi_k = \exp(-(c-a)x^2) H_k(\sqrt{2c}x)$$

$$a = \frac{1}{4\sigma^2}, b = \frac{1}{2l^2}, A = a + b + c, c = \sqrt{a^2 + 2ab}, B = \frac{b}{A}$$

$$H_k(\sqrt{2c}x) = (-1)^k \exp(x^2) \frac{d^k}{dx^k} \exp(-x^2)$$



核函数的特征函数分析

➤ 数值计算

$$\begin{aligned}\lambda_k \phi_k(x') &= \int_{\mathbb{R}^N} \kappa(x, x') \rho(x) \phi_k(x) dx \\ &\approx \frac{1}{n} \sum_{i=1}^n \kappa(x_i, x') \phi_k(x_i) \quad x_i \sim \rho(x)\end{aligned}$$

Gram矩阵 $K_{ij} = \kappa(x_i, x_j)$ ，特征值分解得到

$$K u_k = \lambda_k^{\text{mat}} u_k$$

我们有核函数的特征函数特征值估计

$$\lambda_k \approx \frac{\lambda_k^{\text{mat}}}{n} \quad [\phi_k(x_1) \phi_k(x_2) \cdots \phi_k(x_n)]^T \approx \sqrt{n} u_k$$

Nystrom 方法

$$\phi_k(x) = \frac{\sqrt{n}}{\lambda_k^{\text{mat}}} [\kappa(x_1, x) \kappa(x_2, x) \cdots \kappa(x_n, x)] u_k$$



核函数的特征函数分析

➤ 有界区域的核函数

$$f(x) = \sum_{i=1}^{\infty} w_i \sqrt{\lambda_i} \phi_i(x) \quad w_i \sim \mathcal{N}(0,1)$$

$$\int \phi_i(x) \phi_j(x) d\Omega = \delta_{ij}$$

$$\kappa(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i^*(x')$$

$$\rho(x) = \frac{1}{|\Omega|} dx$$

Ω

$$\lambda_i \sim \mathcal{O}(i^{-\alpha})$$



核函数的特征函数分析

➤ 有界区域的核函数

$$T_\kappa : L_2(\Omega, \mu) \rightarrow L_2(\Omega, \mu) \quad T_\kappa \phi_i = \lambda_i \phi_i$$

$$\kappa(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i^*(x')$$

➤ 例子

$\Omega = [0,1]$, $T_\kappa = \sigma^2(-\Delta + \tau^2)^{-\alpha/2}$, Δ 是满足Dirichlet边界条件的Laplace算子

$$\phi_i = \frac{\lambda_i}{\sigma^2} (-\Delta + \tau^2)^{\alpha/2} \phi_i , \quad \phi_i = 0 \text{ 在 } \partial\Omega$$

$$\phi_i(x) = \sin i\pi x , \quad \lambda_i = \frac{\sigma^2}{((i\pi)^2 + \tau^2)^{\alpha/2}} \sim \mathcal{O}(i^{-\alpha})$$



参考文献

➤ 参考文献

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, Chapter 4

Dashti, Masoumeh, and Andrew M. Stuart. "The Bayesian approach to inverse problems." arXiv preprint arXiv:1302.6989 (2013).