カーネル法によるノンパラメトリックな統計的推論

福水　健次　　数理・推論研究系教授 ／ 統計的機械学習研究センター長

■ カーネル法

正定値カーネルを用いた，データの非線形性，高次モーメントを取り込むための新しい方法論．効率的な計算を重視．

- 定義：正定値カーネル

  Ω：集合．$k:\Omega\times\Omega\to\mathbf{R}$が正定値であるとは，$k$ が対称で，任意の$n\in\mathbf{N}$, $x_1,\ldots x_n\in\Omega$ に対し，

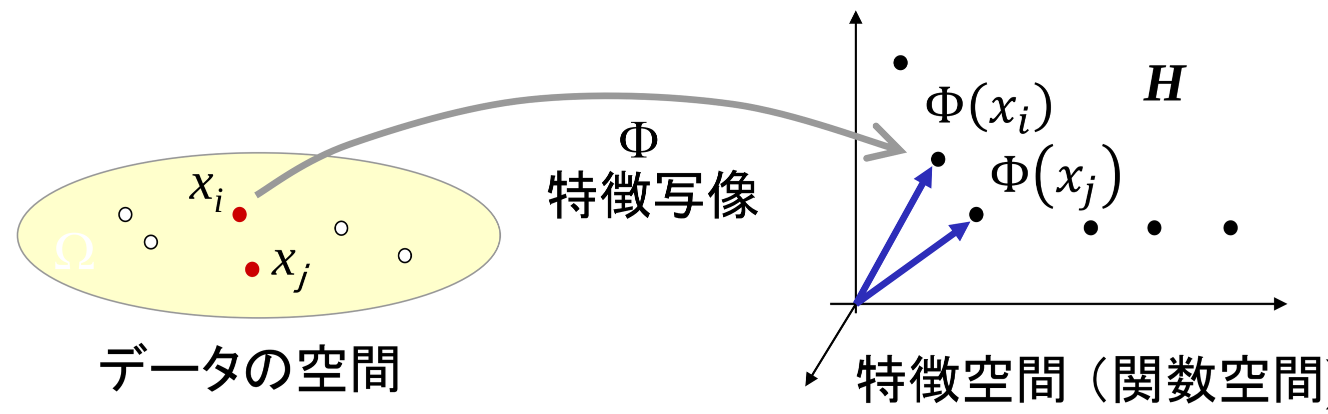
  グラム行列 $\begin{pmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{pmatrix}$ が半正定値．

  例：ガウスカーネル　$\exp(-||x-y||^2/(2\sigma^2))$

- 再生核ヒルベルト空間（RKHS）
  - 正定値カーネルに対しヒルベルト空間が一意に定まる．
  - Ω 上の関数からなる関数空間．一般に無限次元．
  - 特殊な内積を持つ．

    $\langle k(\cdot,x), f\rangle_{\mathcal{H}} = f(x)\ \forall f\in\mathcal{H}, x\in\Omega.$ （再生性）

- 正定値カーネルとRKHSを用いたデータ解析



  - 特徴写像と特徴ベクトル

    $\Phi:\Omega\to H,\qquad x\mapsto\Phi(x):=k(\cdot,x).$
  - カーネルトリック：容易な内積計算

    $\langle\Phi(x),\Phi(y)\rangle=\langle k(\cdot,x),k(\cdot,y)\rangle=k(x,y).$
  - 特徴空間上で線形の解析手法を適用

    （注：多くの手法は内積計算が本質的）

    → 線形手法の非線形化

    カーネル主成分分析，カーネル判別分析

    サポートベクターマシン
  - 非線形カーネル → 非線形／高次情報の抽出

■ カーネル平均による分布の表現

- カーネル平均：特徴ベクトル $\Phi(X)$ の期待値

  $m_X := E[\Phi(X)] = E[k(\cdot,X)]\in H$
  - X の高次モーメントの情報を持つ．

    例) $k(u,x)=c_0+c_1(xu)+c_2(xu)^2+\cdots$　と書ければ

    $E[k(u,X)]=c_0+c_1E[X]u+c_2E[X^2]u^2+\cdots$

- 特性的なカーネル

  $P\mapsto m_P$　が単射，すなわち

  $E_{X\sim P}[k(u,X)]=E_{X\sim Q}[k(u,X)]\quad\Leftrightarrow\quad P=Q.$

  例．ガウスカーネル．c.f. 特性関数 $E[e^{iu^Tx}]$.

- カーネル平均による統計的推論

  原理：特性的なカーネルを用いると，

  分布に関する推論　→　カーネル平均に関する推論

  例：　2標本問題　$m_X = m_Y$? (Gretton et al. NIPS'06),

  独立性検定　$m_{(X,Y)} = m_X\otimes m_Y$?　(Gretton et al. NIPS'07)

■ カーネル平均によるベイズ推論

「行列計算」に基づく新しいノンパラメトリックなベイズ推論法．

分布はカーネル平均として，変数間の関係は共分散作用素として表現．

- 共分散作用素　$C_{YX} := E[\Phi_Y(Y)\Phi_X(X)^T] : H_X\to H_Y$

- 推定量（標本平均，標本共分散作用素）

  $\hat{m}_X := \frac{1}{n}\sum_{i=1}^n\Phi_X(X_i),\quad \hat{C}_{YX} := \frac{1}{n}\sum_{i=1}^n\Phi_Y(X_i)\otimes\Phi_X(X_i)^T$

- ベイズ推論
  - Bayes' rule：$q(x\mid y)=\frac{p(y|x)\pi(x)}{\int p(y|x)\pi(x)dx}$

- Kernel Bayes' Rule: 事後確率のカーネル平均

  $m_{Q_x|y} = C_{XY}^{\pi}{C_{YY}^{\pi}}^{-1}\Phi(y)$

ただし $C_{YX}^{\pi} = C_{(YX)X}C_{XX}^{-1}m_\pi,\quad C_{YY}^{\pi} = C_{(YY)X}C_{XX}^{-1}m_\pi$

グラム行列計算により事後確率（カーネル平均）が得られる

事前確率：$\sum_{i=1}^{\ell}\alpha_i\Phi(\tilde{X}_i)$, P_{XY}: $(X_1,Y_1),\ldots,(X_n,Y_n)$ $i.i.d.$

⇒ 事後確率：　$\hat{m}_{Q_x|y} = \sum_{i=1}^n w_i(y)\Phi(X_i)$,

$w(y) = R_{X|Y}\mathbf{k}(y),\quad \mathbf{k}(y) = \left(k(Y_1,y),\ldots,k(Y_n,y)\right)^T$

$R_{X|Y} = \Lambda G_{YY}\left((\Lambda G_{YY})^2 + \delta_n I_n\right)^{-1}\Lambda\mathbf{k}(y),$

$\Lambda = \mathrm{Diag}[(G_{XX}+n\varepsilon_n I_n)^{-1}G_{X\tilde{X}}\alpha]$

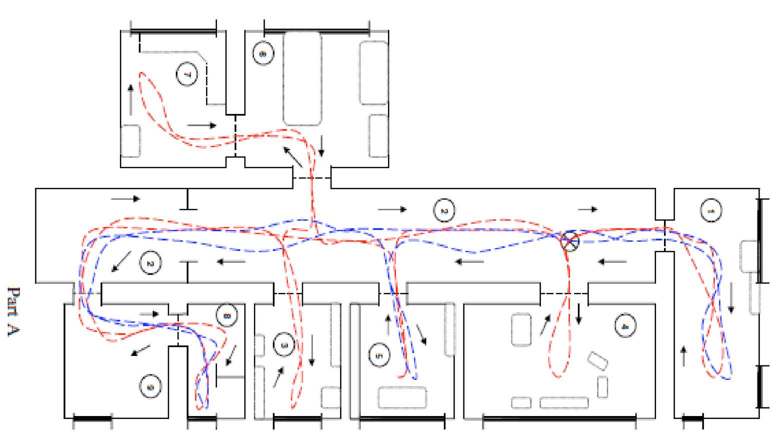
- カーネルベイズ推論則の応用
  - ノンパラメトリックHMM (Fukumizu et al. NIPS 2011)

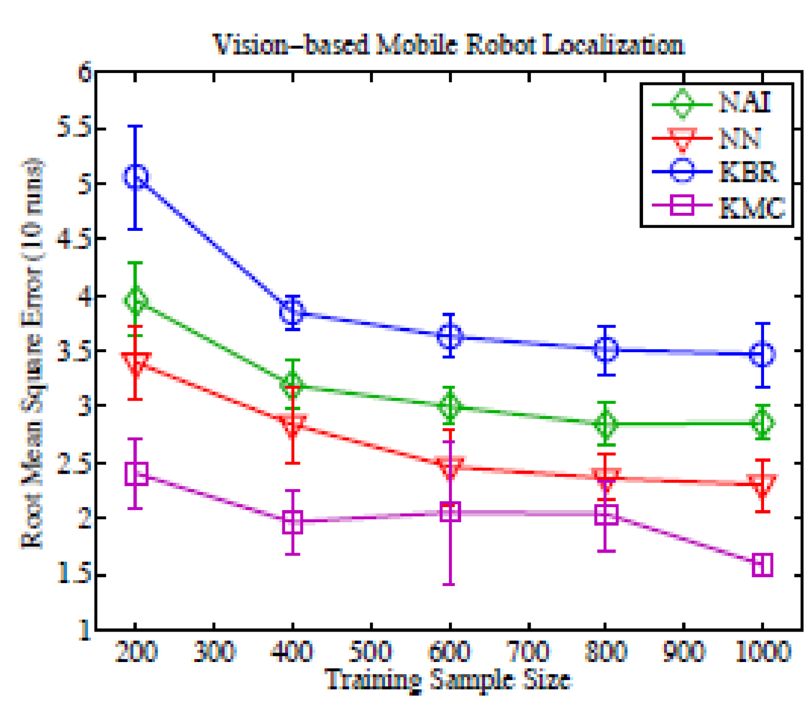
    遷移則や観測過程 をノンパラメトリックにGram行列表現．
  - Kernel Approximate Bayesian Computation (Kernel ABC) (Nakagome , Fukumizu, Mano. 2012)

    尤度関数が陽に書けない場合
  - POMDP: Bellman方程式のカーネル化　(Nishiyama et al UAI2012)
  - Monte Carloとの組み合わせ (カーネル Monte Carlo フィルタ, Kanagawa et al AAAI 2014)

□ カーネルMonte Carloフィルタの応用：ロボット位置推定
- 状態遷移：モデル化容易　→ Monte Carlo Sampling

  観測過程：モデル化困難　→ カーネルベイズ





NAI: naïve method (closest image in training data)

NN: Particle Filter + K-近傍 (Vlassis, Terwijn, Kröse 2002)

KBR: KBR + KBR

KMC: KBR + Monte Carlo