

カーネル法による条件付き確率分布の表現について

清水瑛貴（総合研究大学院大学 統計科学専攻 5年一貫制博士課程3年）

導入：カーネル法

正定値カーネルを用いることで、データの非線形性を統計モデルや推論に取り込むことができる。代表例としてはカーネルPCAやサポートベクトルマシンが挙げられるが、以下で紹介するように確率分布の表現にも用いられ多くの応用がある (Muandet+, 2016)。

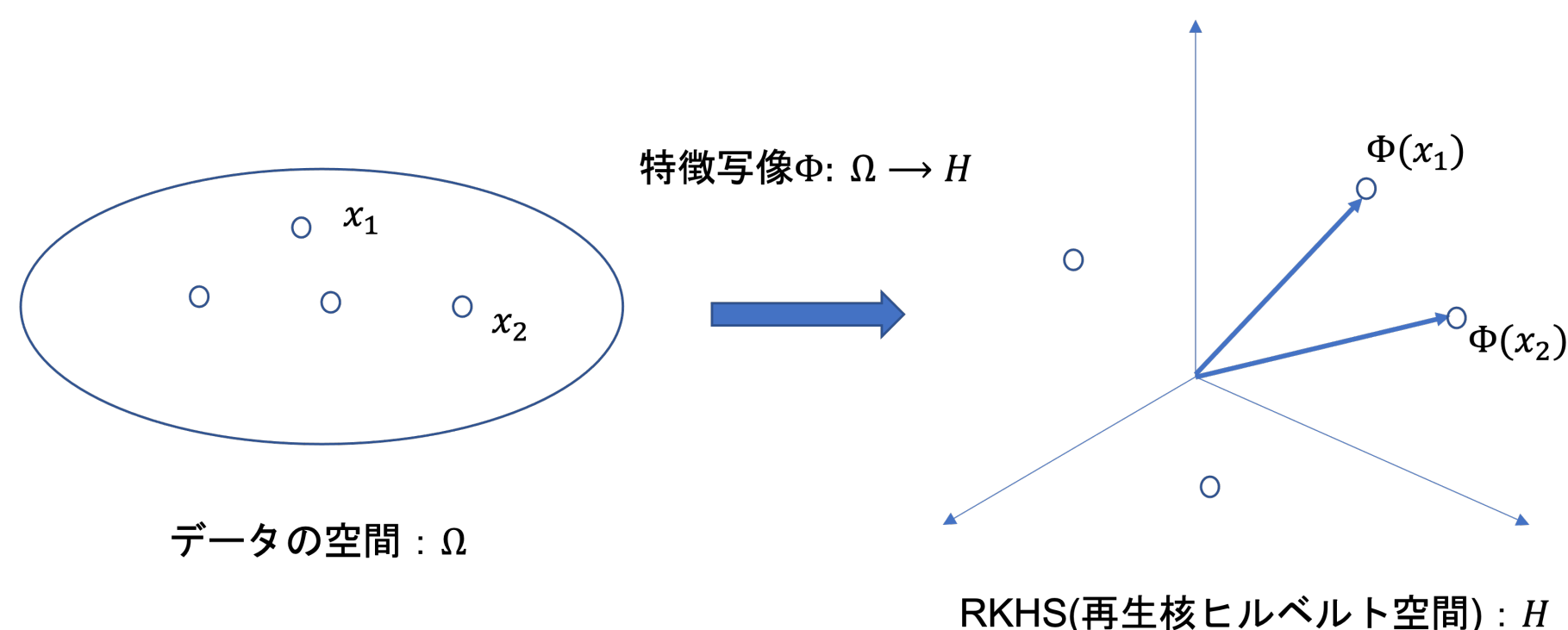


Figure 1. 正定値カーネルによるデータの特徴写像

カーネルにより定められる RKHS の性質

Ω 上の正定値カーネル k に対して、 Ω 上の関数からなる H で以下を満たすものが定まる

- $k(\cdot, x) \in H, x \in \Omega$
- $\langle f, k(\cdot, x) \rangle = f(x)$ for any $f \in H, x \in \Omega$

よって上図では

$$\langle \Phi(x_1), \Phi(x_2) \rangle = \langle k(\cdot, x_1), k(\cdot, x_2) \rangle = k(x_1, x_2)$$

となり、特徴空間上での内積計算はカーネルの値に還元される。

Kernel Mean Embedding による確率分布の表現

正定値カーネルを用いることで、確率分布を RKHS 上の平均として特徴づけることができる。

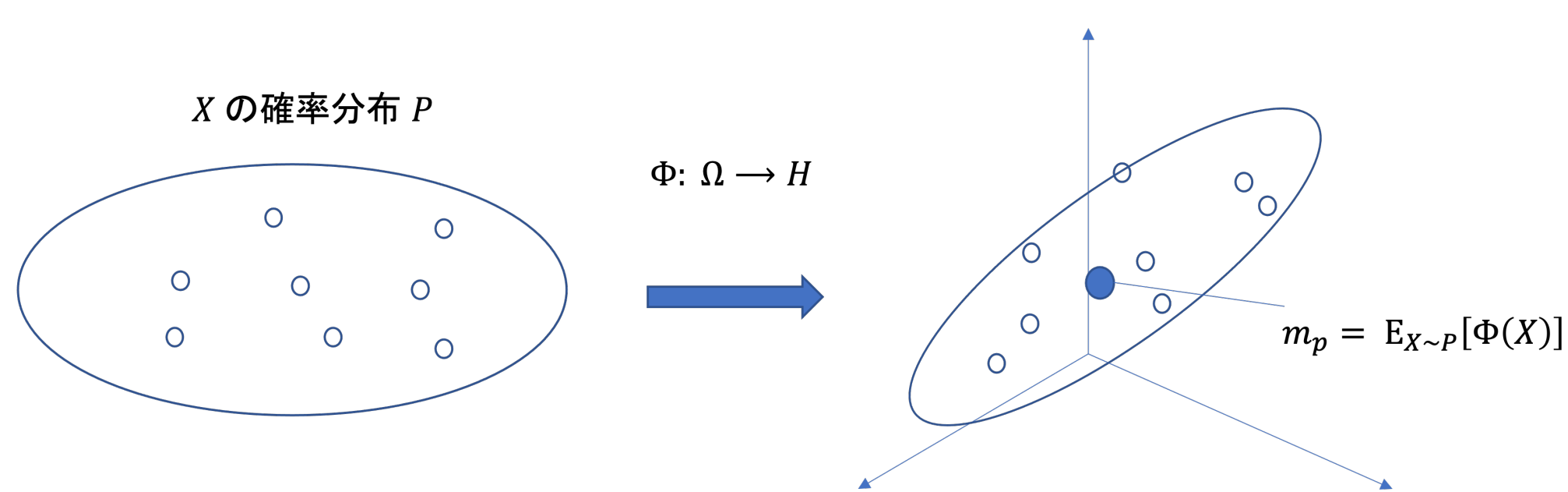


Figure 2. Kernel Mean Embedding (KME)

特に、ガウスカーネルなどの「特性的なカーネル」(Sriperumbudur+, 2010) は、 m_p により P を一意に定めることが知られている。また上述の RKHS の性質に関連して、 m_p は以下の式を満たす。

$$\langle f, m_p \rangle = E[\langle f, k(\cdot, x) \rangle] = E[f(x)]$$

P に従う i.i.d. サンプルが得られるとき、以下のように推定できる。

KME の推定

$$\hat{m}_p = \frac{1}{M} \sum_{i=1}^M \Phi(x_i)$$

また一般に

$$\hat{m}_p = \frac{1}{M} \sum_{i=1}^M w_i \Phi(x_i)$$

のように w_i によって重み付けされたサンプルによる表現ともみなせる。ここで、 w は負の値をとりうる。

条件付き確率分布への応用

KME を $P(Y|X)$ の推定に応用することができ、Conditional Kernel Mean Embedding (Song+, 2013) と呼ばれる。

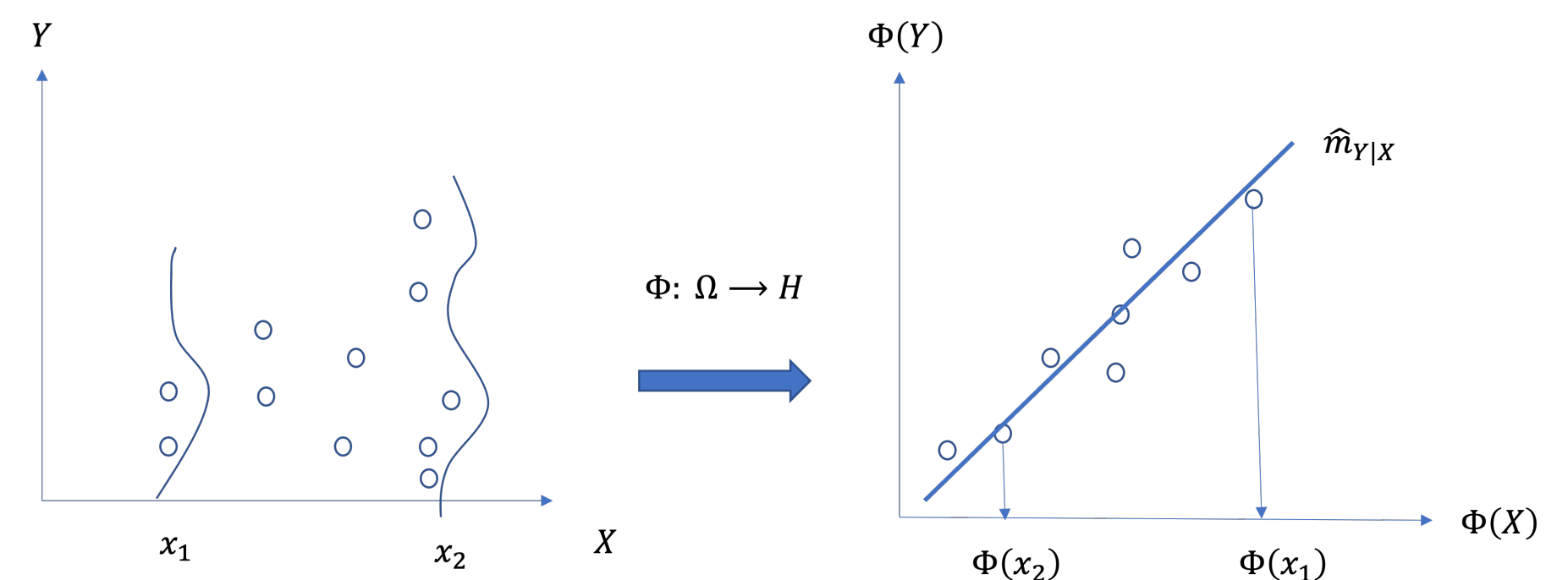


Figure 3. Conditional KME

上図のように、関数空間上で回帰問題を解いているとも解釈をすることができる (Park+, 2020)。

Conditional KME の推定

$$\hat{m}_{Y|X} = \sum_{i=1}^M \beta_i(x) \Phi(y_i) = \Phi \beta(x)$$

$$\beta(x) = (\mathbf{K}_X + \lambda I)^{-1} \mathbf{k}_X$$

ここで、 $[\mathbf{K}_X]_{ij} := k_X(x_i, x_j)$, $\mathbf{k}_X := (k_X(x_1, \cdot), \dots, k_X(x_n, \cdot))^T$ 。

$\beta(x)$ を前述の w と対応させることで、重み付きサンプルによる表現ともみてとれる。

なお、 $\Phi(Y)$ が線形カーネルの場合は、カーネルリッジ回帰と一致する。 Y に関する特性的なカーネルを用いることで複雑な条件付き分布を表現できる。

提案：深層学習と組み合わせた新たな手法に向けて

本研究は unpublished であるため、手法の詳細を述べることは控える。モチベーションとしては、以下に述べる従来の Conditional KME の問題点を改善することである。

Conditional KME の問題点

- グラム行列 \mathbf{K}_X の逆行列の計算コスト
データ数を n とするとき $O(n^3)$ になってしまう
- カーネルの表現能力
深層学習と比べると特徴量の学習を積極的には行わない
- カーネルのハイパーパラメタの決め方
 k_X と k_Y の両方のハイパラが存在しており難しい

上記の問題を解決するために、深層学習を用いて $\hat{m}_{Y|X}$ を推定できるような手法を考案した。

まず、 \mathbf{K}_X の逆行列に関しては深層学習モデルに置き換えることができるため、計算の必要がなくなる。自動微分や確率的勾配法により効率良く推論が可能となる。

次に、表現能力に関する向上させることができ、 X が高次元のデータでも対応できるようになる。

最後に、カーネルのハイパラに関して、 k_X は深層学習モデルに吸収させるため、ある意味では自動的に決定される。一方で、 k_Y のハイパラの決定は容易ではない (Cross Validation などが通用しない)。多少ヒューリスティックではあるが、推論のコストを増やさず、かつ overfit しづらいアプローチを考案した。

今後は、ベイズ推論、強化学習やシミュレータのパラメータ決定問題などへの応用に取り組んでいく予定である。

本研究は福水教授 (統数研) との共同研究である。