

PRML 輪読会 資料

西本 洋紀

平成 31 年 2 月 13 日

第6章 カーネル法

導入

第3章、第4章で議論した線形回帰、分類問題では、入力 \mathbf{x} から出力 y への写像 $y(\mathbf{x}, \mathbf{w})$ を以下のように定式化した。

$$y(\mathbf{x}, \mathbf{y}) = f(\mathbf{w}^T \phi(\mathbf{x})) \quad (1)$$

但し、 $\phi(\mathbf{x})$ は基底関数である。 $f(\cdot)$ は、回帰問題では恒等関数、分類問題では非線形活性化関数である。学習の際は、訓練データはパラメータ分布の点推定値、あるいは事後分布を得るために利用される。訓練後には、訓練データは捨てられ、新しい入力に対する予測は学習済みのパラメータベクトル \mathbf{w} のみを用いて行われる。このアプローチは、第5章で扱ったニューラルネットワークのような非線形のパラメトリックモデルにおいても同様である。

一方で、訓練データ点の一部ないし全部を、予測時にも利用するようなパターン認識法のクラスが存在する。これはメモリベース法と呼ばれ、全ての訓練データを予測時まで保存しておく必要がある。一般に、「訓練」は早く行うことができる反面、テスト点に対する予測には時間がかかる。

通常、これらのメモリベース法では、入力空間における任意の2つのベクトルの類似度を測る指標が必要になる。この指標にカーネル関数を用いる方法をカーネル法といい、本稿で詳しく議論する。カーネル関数は、入力ベクトル \mathbf{x} からあらかじめ定義された非線形の特徴空間への写像 $\phi(\mathbf{x})$ に基づくモデルにおいて、下式 (2) のように定義される。

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}')^T \phi(\mathbf{x}) \quad (2)$$

多くのカーネル関数は、引数の差に依存したもの、つまり、 $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$ の形をもったもので、いくつかの入力空間の返還に関して不変であることから不変カーネルと呼ばれる。一方、 $k(\mathbf{x}, \mathbf{x}') = k(\|\mathbf{x} - \mathbf{x}'\|)$ のように、2つの入力ベクトルの間の距離 (主にユークリッド距離) にのみ依存するカーネルは、均一カーネルあるいは **Rtextbf**(動径基底関数; radial basis function) と呼ばれる。

本稿の構成

6.1 節では、回帰で用いられる線形モデルを双対表現で表すことで、カーネル関数が自然に現れてくることを確認する。6.2 節では、有効なカーネル関数の様々な構成法について述べる。続いて 6.3 節では、線形回帰モデルの基底関数としてどのようなものを取れば良いかという問題を扱う。一般的に、基底関数としてよく利用されるのは **Rtextbf** である。**Rtextbf** を導入する動機について説明し、その過程で Nadaraya-Watson モデルを導く。このモデルは、カーネル回帰モデルをカーネル密度推定の観点から動機づけることによって導出され、カーネル回帰とも呼ばれる。6.4 節では、6.1 節の発展として、カーネルを確率的識別モデルに対しても適用することで、ガウス過程を導き、ベイズ的な設定においても、自然にカーネルが現れてくることを見る。

6.1 双対表現

線形回帰モデルや線形分類モデルを双対表現で表すことで、カーネル関数が自然に現れてくる。このことは、第7章で扱うサポートベクトルマシンにおいて重要な意味を持つ。本節では特に、回帰のための非確率的線形モデルに対して双対性の概念を用いることでカーネル法を導く。

まず、下式 (3) のように、誤差関数が二乗和誤差と2次の正則化項の和で表せるような線形回帰のモデルを考える。

$$J(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\}^2 + \frac{\lambda}{2} \mathbf{w}^T \mathbf{w} \quad (3)$$

なお、 $\lambda \geq 0$ であるとする。

式 (3) の誤差関数 $J(\mathbf{w})$ の \mathbf{w} についての勾配を零とおくと、 \mathbf{w} は、以下のように、係数が \mathbf{w} の関数であるような、ベクトル集合 $\phi(\mathbf{x}_n)$ の線型結合の和になることがわかる。

$$\begin{aligned} \mathbf{w} &= -\frac{1}{\lambda} \sum_{n=1}^N \{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\} \phi(\mathbf{x}_n) \\ &= \sum_{n=1}^N a_n \phi(\mathbf{x}_n) \\ &= \Phi^T \mathbf{a} \end{aligned} \quad (4)$$

ここで、 Φ は、 n 番目の行が $\phi(\mathbf{x}_n)^T$ で与えられるような計画行列である。また、

$$a_n = \frac{1}{\lambda} \{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\} \quad (5)$$

として、そのベクトル形式を $\mathbf{a} = \{a_1, \dots, a_N\}^T$ とする。

$\mathbf{w} = \Phi^T \mathbf{a}$ を用いて (\mathbf{w}) を $J(\mathbf{a})$ に変換すると、下式のようにになる。

$$J(\mathbf{a}) = \frac{1}{2} \mathbf{a}^T \Phi \Phi^T \Phi \Phi^T \mathbf{a} - \mathbf{a}^T \Phi \Phi^T \mathbf{t} + \frac{1}{2} \mathbf{t}^T \mathbf{t} + \frac{\lambda}{2} \mathbf{a}^T \Phi \Phi^T \mathbf{a} \quad (6)$$

ただし、 $\mathbf{t} = \{t_1, \dots, t_N\}$ であるとする。ここで、 $N \times N$ の対称行列であり、その要素が

$$K_{nm} = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) \quad (7)$$

で定められるグラム行列 $\mathbf{K} = \Phi \Phi^T$ を定義する。なおここで、式 (2) で定義されるカーネル関数 $k(\mathbf{x}, \mathbf{x}')$ を使った。グラム行列を用いると、二乗和誤差関数は

$$J(\mathbf{a}) = \frac{1}{2} \mathbf{a}^T \mathbf{K} \mathbf{K} \mathbf{a} - \mathbf{a}^T \mathbf{K} \mathbf{t} + \frac{1}{2} \mathbf{t}^T \mathbf{t} + \frac{\lambda}{2} \mathbf{a} \mathbf{K} \mathbf{a} \quad (8)$$

と少し簡単に書くことができる。

さて、パラメータベクトル \mathbf{w} を直接扱う代わりに、最小二乗法のアルゴリズムをパラメータベクトル \mathbf{a} で表現し直すことにする。これは双対表現と呼ばれる。 $J(\mathbf{w})$ を $J(\mathbf{a})$ に変換して、 $\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} = 0$ を \mathbf{a} について解いても良いが、式 (3) を利用すればより簡単に双対表現が求められる。すなわち、 $\mathbf{w} = \Phi^T \mathbf{a}$ を式 (3) に代入して、 \mathbf{a} について解けば良い。結果は

$$\mathbf{a} = (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t} \quad (9)$$

となる。式 (9) の双対表現を線形回帰モデルに代入し直すことによって、新しい入力 \mathbf{x} に対する予測は次のように与えられることがわかる。

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) = \mathbf{a} \Phi \phi(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t} \quad (10)$$

ここで、 $\mathbf{k}(\mathbf{x})$ は要素 $k_n = k(\mathbf{x}_n, \mathbf{x})$ を持つベクトルである。

以上より、双対表現に変換することで、最小二乗法の解はカーネル関数 $k(\mathbf{x}, \mathbf{x}')$ のみによって表現できることがわかる。これが双対表現と呼ばれる理由は、 $\phi(\mathbf{x})$ の要素の線型結合によって \mathbf{a} が表現できることから、パラメータベクトル \mathbf{w} を使った、元々の定式化を復元できるためである。

双対表現に変換すると、パラメータベクトル \mathbf{a} を求めるのに $N \times N$ の逆行列を求める必要がある。一方、元々のパラメータ空間における定式化では、 $M \times M$ 行列の逆行列を求めることによって \mathbf{w} が得られる。通常 N は M よりもずっと大きいので、双対表現にすると計算量が大きくなることになり、一見不都合であるかのように見える。しかし、双対表現の意義は、全てがカーネル関数 $k(\mathbf{x}, \mathbf{x}')$ で表現されることにある。これにより、特徴ベクトル $\phi(\mathbf{x})$ を明示的に考えることを避け、高次元の、時には無次元の特徴空間を間接的に扱うことができる。

6.2 カーネル関数の構成

実際にカーネル置換を行うにはカーネル関数として有効なものを構成する必要があり、本節では、そのための様々なアプローチをみる。

- 特徴空間への写像 $\phi(\mathbf{x})$ をもとに、対応するカーネルを設計する方法。
例えば、一次元の入力空間において、次のように

$$k(x, x') = \phi(\mathbf{x})^T \phi(\mathbf{x}') = \sum_{i=1}^M \phi_i(\mathbf{x}) \phi_i(\mathbf{x}') \quad (11)$$

- カーネル関数を直接定義する方法
与えた関数がカーネル関数として有効であることを保証する2つの方法:

- 与えた関数を $k(\mathbf{x}, \mathbf{z})$ として、式変形により

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{z}) \quad (12)$$

と、特徴空間への非線形写像を明示的に構成する方法。

- 以下の定理を用いる方法

定理

関数 $k(\mathbf{x}, \mathbf{x}')$ が有効なカーネル関数であるための必要十分条件は、任意の $\{\mathbf{x}_n\}$ に対して、要素が $k(\mathbf{x}_n, \mathbf{x}_m)$ で与えられるグラム行列 \mathbf{K} が半正定値であることである。

- 単純なカーネル関数を構成要素として、新たなカーネル関数を構築する方法
次の性質を利用することができる。

新たなカーネルを構成する方法

有効なカーネルとして $k_1(\mathbf{x}, \mathbf{x}')$ と $k_2(\mathbf{x}, \mathbf{x}')$ が与えられたとき, 次の関数もやはりカーネル関数である.

$$\begin{aligned}
 k(\mathbf{x}, \mathbf{x}') &= ck_1(\mathbf{x}, \mathbf{x}') \\
 k(\mathbf{x}, \mathbf{x}') &= f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') \\
 k(\mathbf{x}, \mathbf{x}') &= q(k_1(\mathbf{x}, \mathbf{x}')) \\
 k(\mathbf{x}, \mathbf{x}') &= \exp(k_1(\mathbf{x}, \mathbf{x}')) \\
 k(\mathbf{x}, \mathbf{x}') &= k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') \\
 k(\mathbf{x}, \mathbf{x}') &= k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}') \\
 k(\mathbf{x}, \mathbf{x}') &= k_3(\phi(\mathbf{x}), \phi(\mathbf{x}')) \\
 k(\mathbf{x}, \mathbf{x}') &= \mathbf{x}^T \mathbf{A} \mathbf{x}' \\
 k(\mathbf{x}, \mathbf{x}') &= k_a(\mathbf{x}_a, \mathbf{x}'_a) + k_b(\mathbf{x}_b, \mathbf{x}'_b) \\
 k(\mathbf{x}, \mathbf{x}') &= k_a(\mathbf{x}_a, \mathbf{x}'_a)k_b(\mathbf{x}_b, \mathbf{x}'_b)
 \end{aligned}$$

構成要素としてよく使われるカーネル関数には, 以下の2つがある.

1. 多項式カーネル

$$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}')^M \quad (13)$$

2. ガウスカーネル

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2) \quad (14)$$

• 確率的生成モデルからカーネル関数を構成する方法

次の2つの方法がある.

1. 生成モデル $p(\mathbf{x})$ が与えられた時に, $k(\mathbf{x}, \mathbf{x}') = p(\mathbf{x})p(\mathbf{x}')$ によりカーネル関数を定義する方法

2. フィッシャーカーネル

6.3 RBF ネットワーク

$$E = \frac{1}{2} \sum_{n=1}^N \int \{y(\mathbf{x}_n + \xi) - t_n\}^2 \nu(\xi) d\xi \quad (15)$$

$$y(\mathbf{x}) = \sum_{n=1}^N t_n h(\mathbf{x} - \mathbf{x}_n) \quad (16)$$

$$h(\mathbf{x} - \mathbf{x}_n) = \frac{\nu(\mathbf{x} - \mathbf{x}_n)}{\sum_{n=1}^N \nu(\mathbf{x} - \mathbf{x}_n)} \quad (17)$$

0.0.1 Nadaraya-Watson モデル

6.4 ガウス過程

本節では, 回帰のための確率的な線形モデルと, ガウス過程との双対性を示す.