データサイエンス・シリーズ 6

データ学習アルゴリズム

(渡辺澄夫著, 共立出版 2001)

発表日: 平成 15 年 6 月 25 日
担当者: 岩崎 唯史
担当箇所: 3.1 関数近似モデル
3.1.4 球形基底関数, RBF

3.1.4 球形基底関数, RBF

■球形基底関数 (Radial Basis Function)

$R^M \to R^N$ への関数 $f(x,w) = \{f_i(x,w)\}\ (1 \le i \le N)$
($x \in R^M$ は入力, $w = \{w_{ij}\}$ はパラメータ, H は中間ユニット数)

$$
\begin{aligned}
f_i(x,w) &= \sum_{j=1}^{H} w_{ij} o_j, \\
o_j &= \exp\left(-\frac{\|x-\xi_j\|^2}{2\sigma_j^2}\right), \\
\|x-\xi_j\|^2 &= \sum_{k=1}^{M} (x^{(k)} - \xi_j^{(k)})^2.
\end{aligned}
$$

球形基底関数中のパラメータ

$$
W = \{(\xi_j, \sigma_j, w_{ij}); \xi_j \in R^M, \sigma_j > 0, w_{ij} \in R\}.
$$

■3 層パーセプトロンとの対比

- 中間層の応答が「球形」である.

  cf. 3 層パーセプトロンの中間層 (p.50)

$$o_j = \sigma\left(\sum_k w_{jk}x^{(k)} + \theta_j\right), \qquad \sigma(x) = \frac{1}{1+\exp(-x)}.$$

- 3 層パーセプトロンより多くのパラメータが必要 (短所)

- 3 層パーセプトロンより学習データのないところでの振舞が分かりやすい (長所)

■球形基底関数を用いた学習アルゴリズム

1 組のサンプル (x, y) **の学習誤差:** $\mathcal{E}(x, y, w) \stackrel{\text{def}}{=} \frac{1}{2}\|y - f(x, w)\|^2.$

$$\frac{\partial \mathcal{E}(x, y, w)}{\partial w_{ij}} = (f_i(x, w) - y^{(i)}) \exp\left(-\frac{\|x - \xi_j\|^2}{2\sigma_j^2}\right),$$

$$\frac{\partial \mathcal{E}(x, y, w)}{\partial \xi_j} = \sum_{i=1}^{N}(f_i(x, w) - y^{(i)})w_{ij}\frac{(x - \xi_j)}{\sigma_j^2} \exp\left(-\frac{\|x - \xi_j\|^2}{2\sigma_j^2}\right),$$

$$\frac{\partial \mathcal{E}(x, y, w)}{\partial \sigma_j} = \sum_{i=1}^{N}(f_i(x, w) - y^{(i)})w_{ij}\frac{\|x - \xi_j\|^2}{\sigma_j^3} \exp\left(-\frac{\|x - \xi_j\|^2}{2\sigma_j^2}\right).$$

$\Downarrow$ **最急降下法 (3 層パーセプトロンと同じ手続き)**

$$w(t + 1) = w(t) - \eta(t)\nabla_w \mathcal{E}(x, y, w) + \alpha(t)(w(t) - w(t - 1)),$$

$$\nabla_w \mathcal{E}(x, y, w) \stackrel{\text{def}}{=} \left(\frac{\partial \mathcal{E}(x, y, w)}{\partial \xi_j}, \frac{\partial \mathcal{E}(x, y, w)}{\partial \sigma_j}, \frac{\partial \mathcal{E}(x, y, w)}{\partial w_{ij}}\right) \ \textbf{at}\ w = w(t).$$

■各パラメータの初期設定

(1) ξ_j はサンプル $\{x_i\}$ の周辺に設定 (必要ならクラスタリング).

(2) σ_j はサンプルのばらつきより大きめに設定 (各クラスの分散より大きめ).

(3) w_{ij} の絶対値は y の絶対値より十分小さな乱数で設定.

※うまく機能させるには, サンプル $\{x_i\}$ に関しての事前知識が必要.

◆ Green 関数 ([注 23] の準備)

変数 x, **関数** $u(x)$ **に対する微分演算子** L_x, B_x,

$$\text{微分方程式: } L_x u(x) = f(x), \qquad \text{境界条件: } B_x u(x) = 0 \tag{1}$$

⬇ **微分方程式 (1) の解**

$$L_x G(x,y) = \delta(x-y), \qquad B_x G(x,y) = 0, \tag{2}$$

を満たす関数 (Green 関数)$G(x,y)$ **を導入すると,**

$$u(x) = \int G(x,y) f(y) dy. \tag{3}$$

- $G(x,y)$ **は** L_x^{-1} **の積分核.**
- **「微分方程式 (1) を解く」** $\equiv$ **「Green 関数 (式 (2)) を求める」**
- $G(x,y)$ **は,** y **に外力** $f(y)$ **を加えたときの** x **での応答を表す.**
(式 (3) は重ね合わせの原理)

◆ [注 23] 球形基底関数が提案された背景

n 個の入出力の組 $\{(x_i, y_i); n = 1, 2, \ldots, n\}$ $(x_i, y_i \in R^M)$ が与えられたとき, $R^M \to R$ への関数 f に関する正則化された誤差

$$L(f) = \frac{1}{2}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int (Pf)^2(x)dx,$$

を定義する．ただし, $\lambda > 0$ は定数, P はある微分作用素である．以下, 誤差 $L(f)$ を最小にする関数 f を変分法により求める．

$f(x)$ の変分を $f(x) + (Df)(x)$, P の共役作用素を P^* とすると, $\delta L(f) = 0$ より

$$(P^* P f)(x) = \frac{1}{\lambda}\sum_{i=1}^{n}(y_i - f(x))\delta(x - x_i).$$

ここで, $P^* P$ に対する Green 関数 $G(x, y)$ を導入する．

$$P^* P G(x, y) = \delta(x - y).$$

$$f(x) = \sum_{i=1}^{n} c_i G(x - x_i),$$

$$\sum_{j=1}^{n} (G(x_i, x_j) + \lambda) c_j = y_i \quad (i = 1, 2, \ldots, n).$$

特に微分作用素が

$$(P^* P f)(x) = \sum_{k=0}^{\infty} \frac{(-1)^k \sigma^{2k}}{k!\, 2^k} (\nabla^{2k} f)(x),$$

のとき, $G(x, y) \propto \exp(-(x-y)^2/(2\sigma^2))$ となるので (補足資料参照), 誤差 $L(f)$ を最小にする関数は,

$$f(x) = \sum_{i=1}^{n} c_i \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right).$$

これは全データに関する和になっているが, いくつかのデータ毎の代表に置き換えると球形基底関数になる.