第4章 自己組織化マップ

4.1〜4.5

06T4029T 斉藤優

自己組織化マップ

- SOM（self-organizing map：自己組織化マップ）
- 教師なし学習のアルゴリズムに用いるニューラルネットモデル
- コホーネンネットともいう

多次元データの可視化

- SOM
  - 多次元データを低次元のマップに描くことで圧縮
    $\Longrightarrow$ 多次元データの可視化
  - 各ユニットは重み（入力ベクトルと同次元のベクトル）を持つ
  - 入力ベクトルと一番近い重みを持つ勝者ユニットを特定
    勝者ユニットと周りのユニットの重みを調節
    $\Longrightarrow$ 競合学習

数理モデル（1）

I次元の多変量データがN個あると仮定

1. t番目に取り出した入力ベクトル

$$x(t) = (x_1(t)\, x_2(t)\, \cdots\, x_i(t)\, \cdots\, x_I(t))'$$

時間tにおける各ユニットの重み

$$m_j(t) = (m_{1j}(t)\, m_{2j}(t)\, \cdots\, m_{ij}(t)\, \cdots\, m_{Ij}(t))'$$

2. 入力ベクトルと各重みの差のノルムの最小値を計算

$$\|x(t) - m_c(t)\| = \operatorname*{Min}_j \left\|x(t) - m_j(t)\right\|$$

3. 各ユニットの重みを変更

$$m_j(t+1) = m_j(t) + \alpha(t) \times h_{cj}(t) \times (x(t) - m_j(t))$$

数理モデル（2）

$\alpha(t)$ と $h_{cj}(t)$ は $x(t)$ の影響の強さを表現した係数

$$\alpha(t) = \mathrm{Max}\left(1 - \frac{t}{T}, 0\right) \qquad h_{cj}(t) = \exp\left(\frac{\|r_c - r_j\|^2}{2\sigma^2(t)}\right)$$

4. 重みが T 回変更されるまで, 入力ベクトルを変えて 2 と 3 の処理を繰り返す
5. 入力ベクトルに一番近い重みを持つユニットを対応

線形空間との相違

主成分分析とSOMの比較

- 記述の対象
  - 主成分分析は変数とオブザベーション
  - SOMはオブザベーションのみ
- 軸の有無
  - 主成分分析は軸の解釈有り
  - SOMは軸の解釈無し
- 次元数
  - 主成分分析は表現できない情報は捨てられる
  - SOMはユニットを増やしマップを広くすれば,オブザベーションの相違をいくらでも表現可

カクテルマップ

- カテゴリカル変数
  離散値であり大小関係がない値からなる変数
- 表4.3のような多次元データをSOMで可視化
  - オブザベーションが多いのでマップに布置しにくい
  - 各変数の水準を布置（図4.5〜4.7）

決定木との連動

明確に観測対象を分類

$\Longrightarrow$ SOMと決定木を組み合わせるを効果的（図4.8）

分析結果の不定性

- 決定木の場合
  分岐の基準や分岐に利用する変数により全く異なる決定木となる
- SOM の場合
  重みの初期値によりマップそのものが変化

まとめ

- SOM
  - 同一のデータから同一のマップが得られるとは限らない
  - 探索的に示唆を与えるツール