

自己組織化マップ Self-Organizing Map; SOM

中本 敬子
kenakamoto@nifty.com



自己組織化マップ(SOM)とは

- 人工ニューラルネットワークの一つ
- 2層からなるネットワーク
 - 入力層
 - 出力層 n次元(通常2次元)に配列される。
- 教師無し学習
- 競合学習
 - 勝ちニューロン(とその近傍のニューロン)の結合係数のみが学習される。

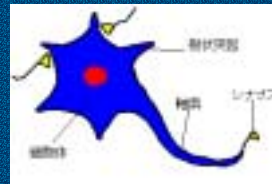


ニューラルネットワーク復習(1)

- 神経回路網をまねた数理モデル
 - 心理学では、コネクショニスト・モデルと言われることの方が多いかもしれません。
- 学習可能性
- 色々な種類がある。
 - 階層型 vs. 相互結合型
 - 教師あり vs. 教師無し



ニューラルネットワーク復習(2)



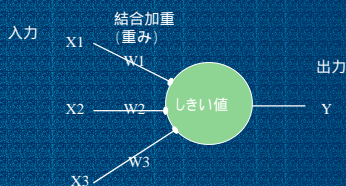
- ニューロン
 - 他のニューロンから入力信号を受け取る
 - 膜電位が上昇
 - しきい値を越えると瞬間的に0vを越える。= ニューロンの発火
 - 軸索を通じて、他のニューロンに刺激を及ぼす。

(図は<http://mars.elcom.nitech.ac.jp/java-cai/neuro/neuro1.html>より引用)



ニューラルネットワーク復習(3)

- ニューロンの性質を模式的に表してみると。

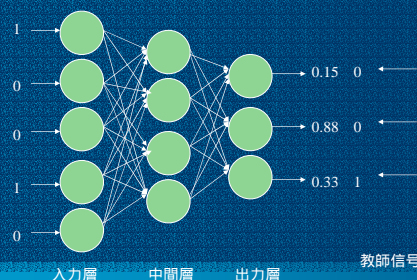


入力 $\sum W_i X_i$ が しきい値 を越えると1、
越えなければ0を出力(階段関数の場合)



階層型 - 教師ありネットワーク(1)

- 例えば、誤差逆伝播法(Back Propagation)による学習



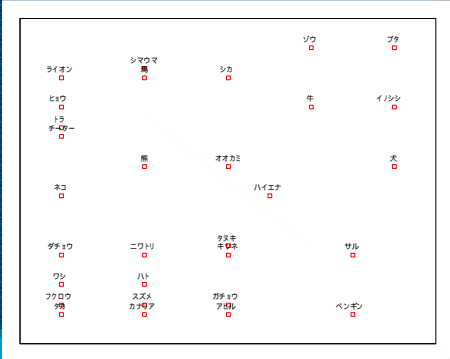
階層型 - 教師ありネットワーク(2)

- 非線形な問題にも対応可能
 - ユニットの出力に連続関数(多くの場合、シグモイド関数)を導入。
 - ネットワークの出力と教師信号との誤差情報からさかのぼって重み付けを修正していく。
- しかし、次のような場合が問題
 - そもそも教師信号が用意できない。
 - 入力データの統計的性質から、カテゴリーを特定していく。
 - 入力データに対して、出力ユニットが1つだけ発火するとは限らない。
 - 競合学習の導入
- 自己組織化マップ

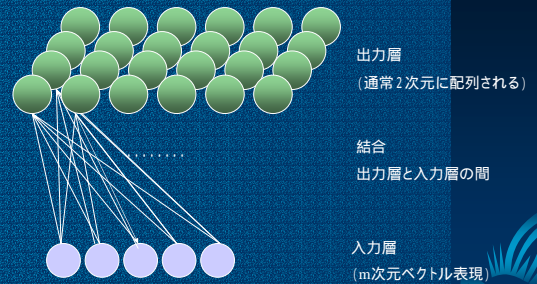
SOMって何に使うの？

- データの視覚化 (visualization)
 - 多変量からなるデータの統計的性質を学習し、
 - 類似したデータが近接するように配列。
 - 単語の使用状況
 - 商品のポジショニング etc...
- ノイズを含む高次元データのパターン認識
 - LVQ (学習ベクトル量子化; Learning Vector Quantization) と組み合わせて行われる。
 - 音声認識
 - 医学での診断 (画像、症状等)

SOMで作られたマップの例



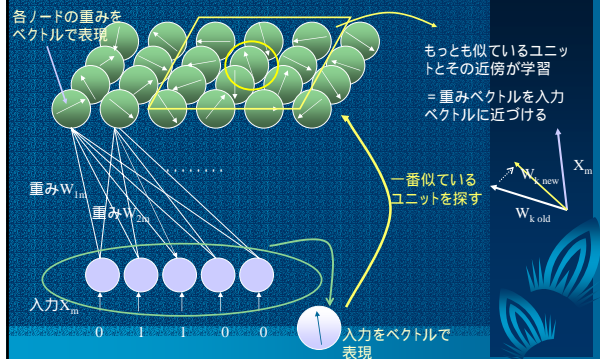
自己組織化マップー模式図



SOMの学習原理

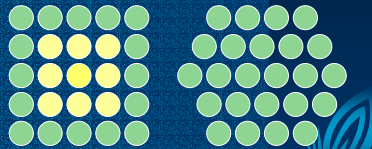
- 入力データベクトルと結合重みベクトルの差を取る。
- もっとも差の小さい出力ユニットを勝者ユニットとする。
- 勝者ユニットとその近傍のユニットが、より入力データベクトルに近づくように学習を行う。

SOMの学習原理ー模式図



出力層の構造

- 2次元であることは本質ではないが、結果の見やすさを考えると2次元になる。
- ユニットの配列
 - 正方形
 - 蜂の巣
- 近傍の取り方
 - 四角形
 - 六角形



数式で書くと・・・

- 勝者ユニット W_c の決定
 $|x - W_c| = \min |x - W_i|$
- 学習 = ベクトルの更新
 - 勝者ユニット近傍 N_c に対してのみ
 $W_i(t+1) = W_i(t) + \alpha(t)[x(t) - W_i(t)]; i \in N_c$
- 学習率 ($0 < \alpha(t) < 1$) と近傍領域 ($N_c = N_c(t)$) は時間とともに減少
 $\alpha(t) = \alpha_0(1 - t/T)$
 $N_c(t) = N_c(0)(1 - t/T)$

日本語で言うと・・・

- 勝者ユニット W_c の決定
 - 入力ベクトルに最も近い重みづけを持つユニットを探す。
- 学習 = ベクトルの更新
 - 勝者ユニットが、もっと入力ベクトルに似るようにする。
- 勝者ユニット近傍 N_c に対してのみ
 - 勝者ユニットとその周りのユニットだけが学習する。
 - ただし、
- 学習率 ($0 < \alpha(t) < 1$) と近傍領域 ($N_c = N_c(t)$) は時間とともに減少
 - はじめは、おおざっぱに大きく学習する (近傍と考える範囲が広く、重み付けの修正の仕方も大きい)。
 - だんだん、小さく小幅に学習し、精密なマップを作っていく。

SOMを使った比喩文の分析

- Clementine7.1をちょっと使ってみる。

・配付資料参照

ちなみにClementineとは、

- SPSS社から販売されているデータマイニング・ツール
- 下記を色々組み合わせてデータを分析できる、らしい。
 - ニューラルネットワーク
 - 自己組織化マップ
 - K-means
 - ロジスティック回帰
 - 主成分・因子分析
 - C & R Tree 等々・・・
- とても楽しそうなツールです。

心理学ではどう使えそうか。

- 認知過程のモデル化
 - 音楽の調性の学習 (Tillman, B., et al, 2000)
 - 音声言語の学習 等・・・
- データ解析手法としては
 - SOMによるデータ解析結果によって仮説を検証した論文はおそらくなさそう。今後も無いような気がする。
 - 大量かつノイズが多いが、直感的に何となく分類できそうな場合に利用可能。・・・仮説発見のための、マイニングツールとして利用

参考文献

- コホネン著 徳高他訳 1996 自己組織化マップ シュプリング
ガー・フェアラーク
 - 自己組織化マップの生みの親によるもの。
 - ヘルシンキ大学工学部 Neural Network Research Center
<http://www.cis.hut.fi/research/> SOM, LVQの研究・学習用パッケージ
(SOM-PAK, LVQ-PAK)が公開されている。
- 徳高平蔵・岸田悟・藤村喜久郎 1999 自己組織化マップの
応用—多次元情報の2次元可視化 海文堂
 - 第1章から第3章までSOMの基本がごく分かりやすく書いてある。
- 岩田彰・松原俊之 1996 ニューラルネットワーク入門
<http://mars.elcom.nitech.ac.jp/java-cai/neuro/menu.html>
 - ニューラルネット全般の入門編。シミュレーションのデモ等もあり、とても
易しく分かりやすい。

