高階微分を学習するニューラル・ネットワーク

益岡 竜介 †‡　　山田 道夫†

†東京大学大学院数理科学研究科

‡富士通研究所

微分情報を学習する目的

特に観測点を増やすことが難しく，入力に関する出力の微分情報が値の情報とは独立に得られる場合に多少計算コストがかかってもより精密な学習を行なう．

例

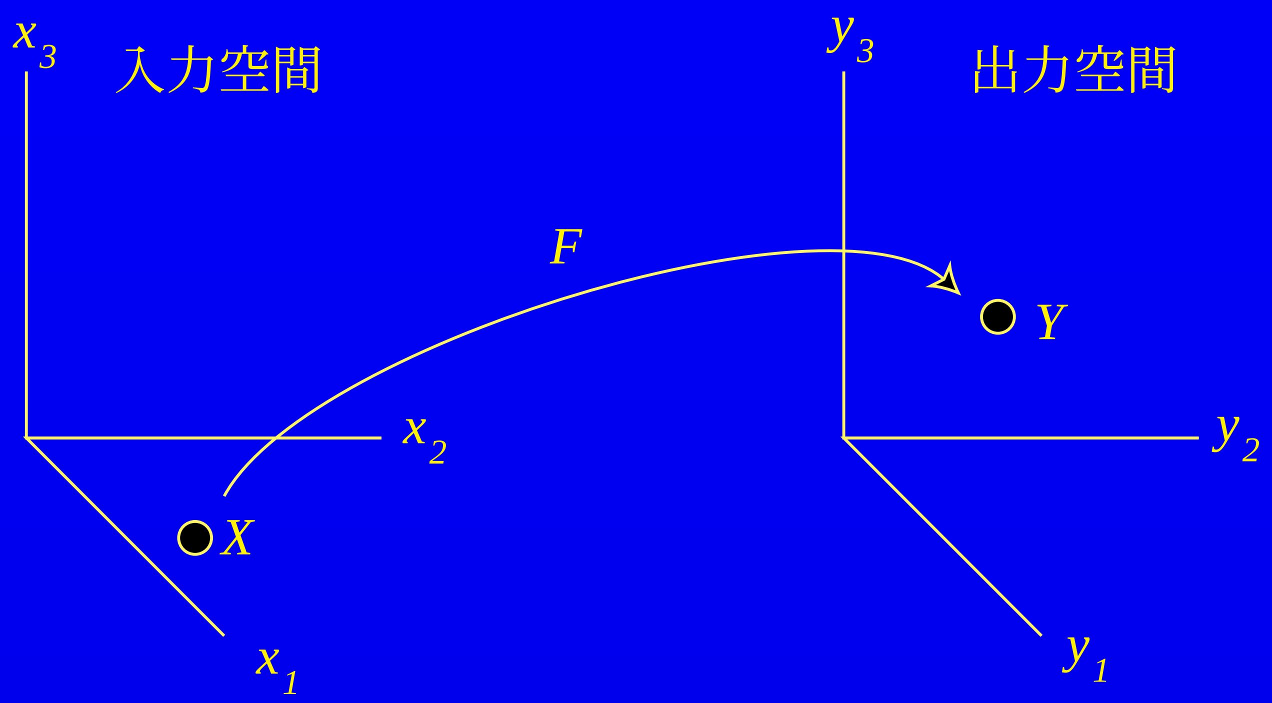
- EBNN (Explanation Based Neural Network)
- パターン認識の回転，移動不変性
- 偏微分方程式を満たす量

微分情報を学習する手段

通常の multilayer perceptron 型のニューラルネットワークをその構造を拡張して, 一般の微分情報を学習できるようにした.

一階の微分の multilayer perceptron 型のネットワークによる学習に関しては, すでに Simard らの tangent prop の研究があり, それを一般の階数の微分に拡張した形になっている.

問題の枠組 1. (定式化)



$$Y = F(X),\ Y = (y_1, y_2, y_3, \ldots),\ X = (x_1, x_2, x_3, \ldots)$$

問題の枠組 2. (定式化)

ネットワークに与えられる一つのデータは以下のような形をしている.

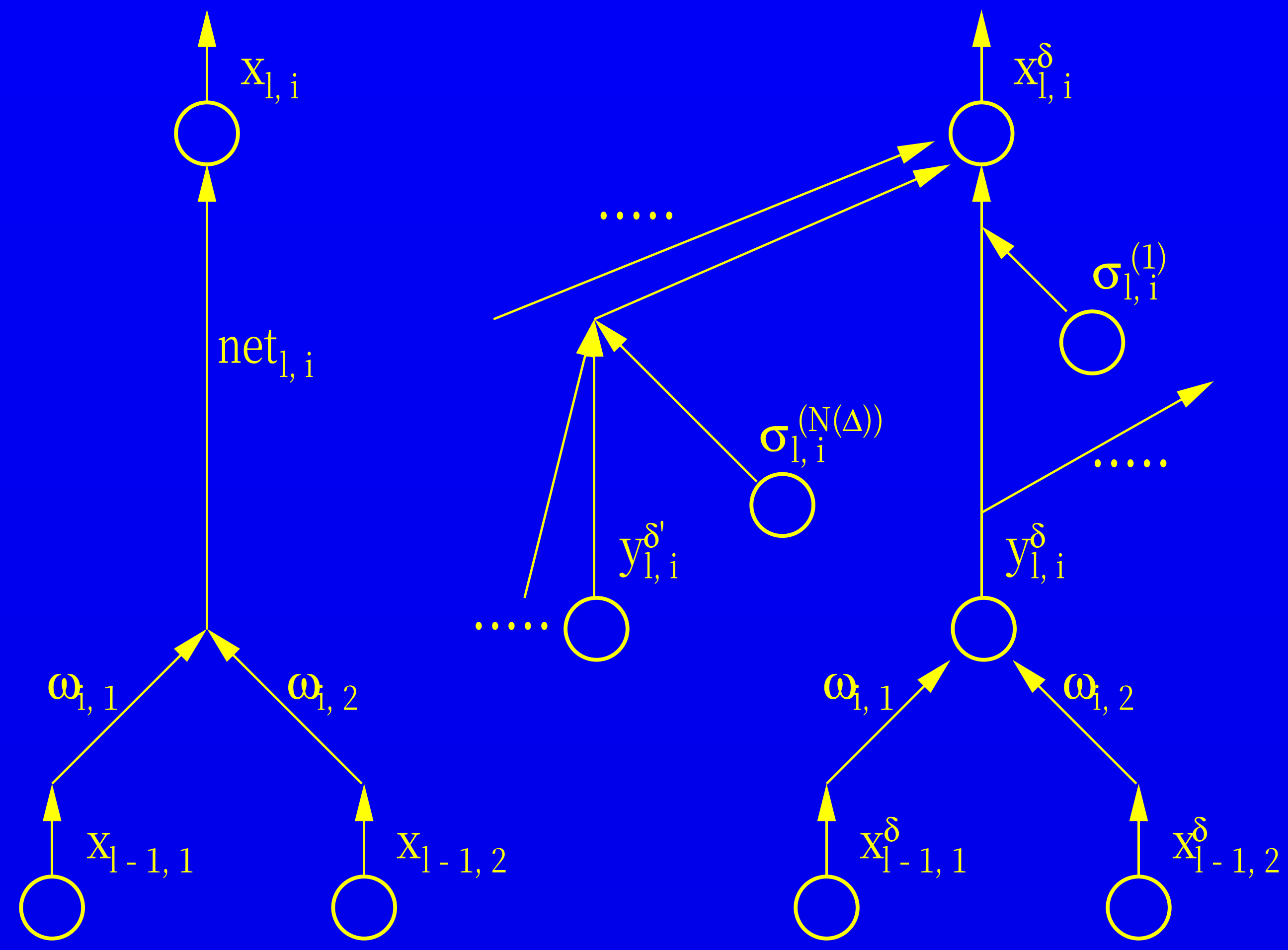
入力 $\Longrightarrow$ X (入力空間内のデータの座標)

出力 $\Longrightarrow$ $\left(Y, \frac{\partial F}{\partial x_1}\Big|_X, \ldots, \frac{\partial^2 F}{\partial x_1 \partial x_2}\Big|_X, \ldots\right)$

(出力空間内のデータの座標と入力に関する微分)

ネットワークの構造

左が元のネットワーク, Value Net で, 右がそれから構成した δ Net である.



Forward Propagation

value net に関しては通常の multilayer perceptron の時と全く同様である.
以下は δ Net のユニットに関する式である.

$$
\begin{aligned}
\sigma_{l,i}^{(m)} &= \sigma^{(m)}(net_{l,i}) \\
y_{l,i}^{\delta} &= net \triangleright y_{l,i}^{\delta} = \sum_j w_{i,j} x_{l-1,j}^{\delta} \\
x_{l,i}^{\delta} &= net \triangleright x_{l,i}^{\delta} = \sum_{(d_{\delta,\Delta},\Delta) \in V(\delta)} d_{\delta,\Delta} \left\{ \sigma_{l,i}^{(N(\Delta))} \prod_{(c,\delta') \in \Delta} \left(y_{l,i}^{\delta'} \right)^{c} \right\}
\end{aligned}
$$

エネルギー関数と BP の対象

以下はエネルギー関数と，重みの更新の式と，重みの更新を計算するために back propagate する ϵ の定義である.

$$
\begin{aligned}
E &= \sum_{\delta \geq 0} \alpha^{\delta} E^{\delta} \\
E^{\delta} &= \frac{1}{2} \sum_{p} \left(O_p^{\delta} - x_p^{\delta} \right)^2 \\
\Delta w_{i,j} &= -\frac{\partial E}{\partial w_{i,j}} \\
\epsilon \triangleright u &= -\frac{\partial E}{\partial net \triangleright u}
\end{aligned}
$$

Back Propagation (出力層と Value Net)

以下は出力層と Value Net のユニットに関する back propagation の式である.

$$
\begin{aligned}
\epsilon \triangleright x^0 &= \alpha^0 \left(O_p^0 - x_p^0\right) \sigma^{(1)}(net) \\
\epsilon \triangleright x^\delta &= \alpha^\delta \left(O_p^\delta - x_p^\delta\right) \\
\epsilon \triangleright x_{l-1,j} &= \sum_i \left(\epsilon_{l,i} + \sum_{m \geq 1} \epsilon \triangleright \sigma_{l,i}^{(m)} \right) \omega_{i,j} \sigma_{l-1,j}^{(1)}
\end{aligned}
$$

Back Propagation (δ Net)

以下は δ Net のユニットに関する back propagation の式である.

$$
\begin{aligned}
\epsilon^{\delta}_{l-1,j} &= \epsilon \triangleright x^{\delta}_{l-1,j} = \sum_{i} \epsilon \triangleright y^{\delta}_{l,i}\, \omega_{i,j} \\
\epsilon \triangleright \sigma^{(m)}_{l,i} &= \sum_{\delta : N(\delta) \geq m} \epsilon^{\delta}_{l,i} \left(\sum_{\Delta : N(\Delta) = m \,\wedge\, Sum(\Delta) = \delta} p^{\Delta} \right) \sigma^{(m+1)}_{l,i} \\
\epsilon \triangleright y^{\delta}_{l,i} &= \sum_{\delta' : \delta' \geq \delta} \epsilon^{\delta'}_{l,i} \sum_{\Delta : Sum(\Delta) = \delta' \wedge \exists c\, (c,\delta) \in \Delta} d_{\delta',\Delta}\, \sigma^{(N(\Delta))}_{l,i} \\
& \quad \times c\, (y^{\delta}_{l,i})^{c-1} \left(\prod_{(c'',\delta'') \in (\Delta - \{(c,\delta)\})} (y^{\delta''}_{l,i})^{c''} \right)
\end{aligned}
$$

重みの更新ルール

以下が back propagate された ϵ を使って重みを更新する式である.

$$
\begin{aligned}
\Delta w_{i,j} &= -\frac{\partial E}{\partial w_{i,j}} \\
&= \epsilon \triangleright x_{l,i}\ x_{l-1,j} + \sum_{\delta>0} \epsilon \triangleright y^{\delta}_{l,i}\ x^{\delta}_{l-1,j} + \sum_{m\geq 1} \epsilon \triangleright \sigma^{(m)}_{l,i}\ x^{0}_{l-1,j} \\
&= \sum_{\delta\geq 0} \epsilon \triangleright y^{\delta}_{l,i}\ x^{\delta}_{l-1,j} + \sum_{m\geq 1} \epsilon \triangleright \sigma^{(m)}_{l,i}\ x^{0}_{l-1,j}
\end{aligned}
$$

応用の可能性 (MASCON) 1

気象学の方では，風速場の (比較的) 少ない観測点 (数個から数百程度) の情報からそのグローバルな情報を求める MASCON (Mass-Consistent) という方法が使われている．本当は風速場に関する完全な流体の方程式を解く必要があるのだが，この方法では観測点における観測データによる拘束条件とその他の点で以下のような質量保存の法則を満たすことだけを条件に全体の状態を求める．

$$\nabla f(x) = 0$$

応用の可能性 (MASCON) 2

MASCON にこのネットワークを応用するには，以下の式を最小化するような f を multilayer perceptron 型のネットワークで表されるような範囲の関数から求めるという問題に直す.

$$\sum (f(x_i) - y_i)^2 + \int (\nabla f - 0)^2 dx$$

前者の項 ⟹ Value Net へ

後者の項 ⟹ ランダムに点をとり，δ Net へ

今後の課題

- このネットワークとアルゴリズムを実際に計算機にプログラムとしてインプリメントし，具体的な問題に適用して，その効果などを調べる.
- radial basis function (RBF) に関する Poggio and Girosi などの取り扱いを，同様の問題の枠組で，一般の微分の情報を含む場合に拡張する.

まとめ

- 観測点を増やすことは難しいが，一方で微分情報が値の情報とは独立に得られる場合などに計算コストはかかるが，より精密な学習を可能とする，微分情報を使って学習することができる multilayer perceptron 型のネットワークについてそのネットワークの構造やアルゴリズムを報告した．

- 応用の可能性として，気象学の MASCON という手法に応用する際の方法を考察した．