LH-002

H_2学習における新たな重み空間方程式の導出

Derivation of New Weight-Space Equations in H_2-Learning

西山 清*

Kiyoshi NISHIYAMA *

1 はじめに

バックプロパゲーション (backpropagation: BP) 法 [1] が提案されて以来、人工ニューラルネットワークは様々な分野に爆発的に普及した。しかし、その普及と共に BP の学習速度の遅さが大きな問題となった。この問題を解決する代表的なものの一つにカルマンフィルタ [2] に基づく学習アルゴリズム群 [3]-[4] がある。これらを称して H_2学習と呼ぶことにする。

本研究では、H_2学習に属するすべての学習アルゴリズムの挙動を支配する新たな学習方程式を導出する。

2 ニューラルネットワークの学習

多層ニューラルネットワークの学習問題とは、所望の写像 ($\boldsymbol{z}_k = \boldsymbol{\mathcal{F}}(\boldsymbol{x}_k)$) と同じ入出力をもつように入出力の例 $\{\boldsymbol{x}_k, \boldsymbol{z}_k\}$ から出力誤差 $\{\boldsymbol{e}_k\}$ を用いてネットワーク$\boldsymbol{\mathcal{N}}$の結合重み $\{w^1_{m,j}\}$、$\{w^2_{j,i}\}$ としきい値 $\{\theta^1_j\}$、$\{\theta^2_i\}$ を決定する、すなわち、$\boldsymbol{\mathcal{N}}(\boldsymbol{x}_k;\boldsymbol{w}) = \boldsymbol{\mathcal{F}}(\boldsymbol{x}_k)$ を満たす重みベクトル$\boldsymbol{w}$を例から求める問題である。

3 H_2学習

3.1 H_2学習の原理

H_2学習では、多層ニューラルネットワーク$\boldsymbol{\mathcal{N}}$を状態空間モデル

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k, \quad \boldsymbol{y}_k = \boldsymbol{h}_k(\boldsymbol{w}_k) + \boldsymbol{v}_k \tag{1}$$

で表し、この重みベクトル$\boldsymbol{w}_k$の推定値$\hat{\boldsymbol{w}}_k$の平均２乗誤差

$$J_{\mathrm{KF}} = \mathrm{E}\{\|\boldsymbol{w}_k - \hat{\boldsymbol{w}}_k\|^2 | \boldsymbol{y}_0, \cdots, \boldsymbol{y}_k\} \tag{2}$$

を最小にするように$\hat{\boldsymbol{w}}_k$を逐次更新する。ここで、$\mathrm{E}\{\cdot\}$ は$\boldsymbol{w}_k$に関する期待値、 $\|\cdot\|$ はユークリッドノルムをそれぞれ表す。

*岩手大学 工学部 情報システム工学科，〒020-8551 盛岡市上田 4-3-5，e-mail: nisiyama@cis.iwate-u.ac.jp

3.2 H_2学習アルゴリズム

入力層、中間層、出力層がそれぞれ N_1、N_2、N_3 個のニューロンからなる３層の人工ニューラルネットワーク (N_1-N_2-N_3ネットワーク) において、$N_3 = 1$ のときの出力層のニューロンの出力は次式で表される。

$$\begin{aligned} z_k &= f\left(\sum_{j=1}^{N_2} w^2_{j,1} f\left(\sum_{i=1}^{N_1} w^1_{i,j} x_k(i) + \theta^1_j\right) + \theta^2_1\right), \\ & \qquad k = p + lN_p, p = 1, 2, \cdots, N_p \end{aligned} \tag{3}$$

ただし、$f(x)$ はシグモイド関数と呼ばれ、

$$f(x) = \frac{1}{1+\exp(-\alpha x)} \text{ 又は } f(x) = \frac{1-\exp(-\alpha x)}{1+\exp(-\alpha x)}$$

の関数である ($\alpha > 0$ は $f(x)$ の傾きを表す)。

これより、人工ニューラルネットワークの学習とは、式 (3) が所望な入出力関係を満たす重み $\{w^2_{j,1}\}$,$\{w^1_{i,j}\}$ としきい値 $\{\theta^1_j\}$,θ^2_1を入出力の例から決定する問題となる。そこで、重みベクトル

$$\begin{aligned} \boldsymbol{w} &= [\theta^1_1, w^1_{1,1}, \cdots, w^1_{N_1,1}, \theta^1_2, w^1_{1,2}, \cdots, w^1_{N_1,2}, \\ & \qquad \cdots, \theta^2_1, w^2_{1,1}, \cdots, w^2_{N_2,1}]^T \end{aligned} \tag{4}$$

を定義すれば、式 (3) のニューラルネットワークの出力 $z_k = z^3_1[k]$ は次のように重みベクトルの関数として表すことができる。

$$z_k = h_k(\boldsymbol{w}_k) = \boldsymbol{\mathcal{N}}(\boldsymbol{x}_k;\boldsymbol{w}_k) \tag{5}$$

このとき、重みベクトル$\boldsymbol{w}_k$を状態ベクトルとし、雑音雑音 v_kを考慮すれば、この問題における状態空間モデルが得られる。

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k \qquad \text{(状態方程式)} \tag{6}$$

$$y_k = h_k(\boldsymbol{w}_k) + v_k \qquad \text{(観測方程式)} \tag{7}$$

これに拡張カルマンフィルタ [2] を適用すれば、重みベクトルの平均２乗誤差を最小とする H_2学習の基本アルゴリズムが得られる。

$$\hat{\boldsymbol{w}}_k = \hat{\boldsymbol{w}}_{k-1} + \boldsymbol{K}_k(y_k - h_k(\hat{\boldsymbol{w}}_{k-1})) \tag{8}$$

$$\begin{aligned} \boldsymbol{K}_k &= \hat{\boldsymbol{P}}_{k|k-1}\boldsymbol{H}^T_k(\boldsymbol{H}_k\hat{\boldsymbol{P}}_{k|k-1}\boldsymbol{H}^T_k + 1)^{-1} \\ \hat{\boldsymbol{P}}_{k+1|k} &= \hat{\boldsymbol{P}}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{H}_k\hat{\boldsymbol{P}}_{k|k-1} \end{aligned} \tag{9}$$

ただし、

$$\boldsymbol{H}_k = \left.\frac{\partial h_k(\boldsymbol{w})}{\partial \boldsymbol{w}}\right|_{w=\hat{w}_{k-1}} \qquad (10)$$

4 H_2学習の解析

H_2学習において重要な二つの定理を以下に示す。

[補題０]

$\mu > 0$ のとき、任意の行列$\boldsymbol{H}_i$に対して$\mu^{-1}\boldsymbol{I} + \sum_{i=1}^{k} \boldsymbol{H}_i^T \boldsymbol{H}_i$ は正定対称行列となる。ただし、$\boldsymbol{I}$は単位行列、^{T}は行列の転置を表す。

(証明) 省略。

[補題１]

$\hat{\boldsymbol{P}}_{k+1|k} = (\mu^{-1}\boldsymbol{I} + \sum_{i=1}^{k} \boldsymbol{H}_i^T \boldsymbol{H}_i)^{-1}$ はH_2学習のリカッチ方程式を満たす。ただし、$\hat{\boldsymbol{P}}_{1|0} = \mu\boldsymbol{I}$, ^{T}は転置、$^{-1}$は逆行列をそれぞれ表す。

(証明) 省略。

[定理１]

H_2学習における重みベクトルの更新式は次式と等価である。

$$\begin{aligned}\hat{\boldsymbol{w}}_k &= \hat{\boldsymbol{w}}_{k-1} + \mu\beta_k \boldsymbol{H}_k^T (y_k - h_k(\hat{\boldsymbol{w}}_{k-1})) \\ &\quad -\mu \sum_{i=1}^{k-1} \boldsymbol{H}_i^T \boldsymbol{H}_i (\hat{\boldsymbol{w}}_k - \hat{\boldsymbol{w}}_{k-1}) \qquad (11)\end{aligned}$$

ただし、

$$\beta_k = (\boldsymbol{H}_k \hat{\boldsymbol{P}}_{k|k-1} \boldsymbol{H}_k^T + 1)^{-1} \qquad (12)$$

であり、$\hat{\boldsymbol{P}}_{k|k-1}$は$\mathrm{H}_2$学習のリカッチ方程式 (式 (9)) を満たす。

(証明)

補題１を用いて式 (11) を整理すれば次式が得られる。

$$\begin{aligned}\hat{\boldsymbol{w}}_k &= \hat{\boldsymbol{w}}_{k-1} + \hat{\boldsymbol{P}}_{k|k-1} \boldsymbol{H}_k^T (\boldsymbol{H}_k \hat{\boldsymbol{P}}_{k|k-1} \boldsymbol{H}_k^T + 1)^{-1} \\ &\quad \times (y_k - h_k(\hat{\boldsymbol{w}}_{k-1}))\end{aligned}$$

これはH_2学習における重みベクトルの更新式 (式 (8)) と一致する。

定理１の重み空間更新方程式 (式 (11)) より、H_2学習と最急降下法 (BP) や共役勾配法などに基づく他の学習の関係が容易にわかる。

[系１]

過去の軌跡上の瞬時最急方向の張る部分空間に直交する方向に重みベクトルが更新されるとき、H_2学習は最急降下法と一致する。

(証明) 省略。

[定理２]

H_2学習における重みベクトルの更新量$\Delta\hat{\boldsymbol{w}}_k$は次の方程式を満たす。

$$\begin{aligned}&\Delta\hat{\boldsymbol{w}}_k^T \left(\boldsymbol{I} + \mu \sum_{i=1}^{k-1} \boldsymbol{H}_i^T \boldsymbol{H}_i \right) \Delta\hat{\boldsymbol{w}}_k \\ &\quad = < \Delta\hat{\boldsymbol{w}}_k,\ \mu\beta_k \boldsymbol{H}_k^T (y_k - h_k(\hat{\boldsymbol{w}}_{k-1})) > \qquad (13)\end{aligned}$$

ただし、

$$\Delta\hat{\boldsymbol{w}}_k = \hat{\boldsymbol{w}}_k - \hat{\boldsymbol{w}}_{k-1},\quad \beta_k = (\boldsymbol{H}_k \hat{\boldsymbol{P}}_{k|k-1} \boldsymbol{H}_k^T + 1)^{-1}$$

であり、$\hat{\boldsymbol{P}}_{k|k-1}$は$\mathrm{H}_2$学習のリカッチ方程式を満たす。また、$<\cdot,\cdot>$は内積を表す。

(証明)

定理１の式 (11) を整理すると式 (13) が得られる。

定理２の式 (13) をH_2学習の重み空間学習方程式と呼ぶ。

[系２]

H_2学習では重み空間の各点において瞬時出力２乗誤差が減少する方向に必ず更新される。

(証明) 補題０と定理２を用いる。

5 まとめ

H_2学習における重みベクトルの振舞いを規定する全く新しい発想に基づいた二つ方程式を導出した。これらの方程式はH_2学習アルゴリズムの理論的な挙動解析に新たな視点を与えただけでなく、最急降下法 (BP) や共役勾配法などに基づく他の学習アルゴリズムとの関係も明らかにした。

参考文献

[1] D.E. Rumelhart and J.L. McCelland, Parallel Distributed Processing, Vol. 1, The MIT Press, 1986.

[2] 西山 清、最適フィルタリング、培風館、2001.

[3] G.V. Puskorious and L.A. Feldkamp, “Decoupled extended Kalman Filter training of feedforward layered networks,” IEEE Trans. Neural Networks, 2, pp. 771-777, 1991.

[4] Y. Iiguni, H. Sakai and H. Tokumaru, “A real-time learning algorithm for a multilayered neural network based on the extended Kalman filter,” IEEE Trans. Signal Processing, 40, 4, pp. 959-966, 1992.