

付録 誤差逆伝播法

多層パーセプトロンの代表的な学習アルゴリズムである誤差逆伝播法について説明する。

(L-1)層の隠れ層をもつ多層パーセプトロンの動作式は、

$$u_i^l(\xi^\lambda) = 1 \quad (l=0, \dots, L-1, i=0, \lambda=1, \dots, \Lambda) \quad (\text{数式1})$$

$$u_i^l(\xi^\lambda) = \xi_i^\lambda \quad (l=0, i=1, \dots, n_l, \lambda=1, \dots, \Lambda) \quad (\text{数式2})$$

$$u_i^l(\xi^\lambda) = \sigma \left(\sum_{j=0}^{n_{l-1}} w_{ij}^l u_j^{l-1}(\xi^\lambda) \right) \quad (l=1, \dots, L, i=1, \dots, n_l, \lambda=1, \dots, \Lambda) \quad (\text{数式3})$$

である。ここで、 n_l は第 l 層（入力層を第0層、隠れ層を第1層から第(L-1)層、出力層を第L層とする）の素子数、 Λ は教師データ数、 $u_i^l(\xi^\lambda)$ は第 λ 教師データの入力パターン ξ^λ を入力したときの第 l 層の第 i 素子の出力、 w_{ij}^l は第(l-1)層の第 j 素子から第 l 層の第 i 素子への結合係数、 ξ_i^λ は ξ^λ の第 i 成分である。 $\sigma()$ はシグモイド (sigmoid ; S字) 関数：

$$\sigma(x) = \tanh x \quad (\text{数式4})$$

と定義する。

多層パーセプトロンの出力と目標出力との誤差 E を

$$E = \frac{1}{2\Lambda} \sum_{\lambda=1}^{\Lambda} \sum_{i=1}^{n_L} \left(u_i^L(\xi^\lambda) - \xi_i^\lambda \right)^2 \quad (\text{数式5})$$

と定義する。 ξ_i^λ は第 λ 教師データの目標出力 ξ^λ の第 i 成分である。

誤差 E を弛緩的に極小化する。ここでは最急降下法：

$$\frac{du_{ij}^l}{dt} = -\frac{\partial E}{\partial u_{ij}^l} \quad (l=1, \dots, L, i=1, \dots, n_l, j=0, \dots, n_{l-1}) \quad (\text{数式6})$$

による。 t は時間変数である。数式6の右辺を数式5から実際に求めると、次式が得られる：

$$\frac{\partial E}{\partial u_{ij}^l} = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} \delta_i^l(\xi^\lambda) u_j^{l-1}(\xi^\lambda) \quad (l=1, \dots, L, i=1, \dots, n_l, j=0, \dots, n_{l-1}) \quad (\text{数式7})$$

ただし、

$$\delta_i^l(\xi^\lambda) = \left(u_i^l(\xi^\lambda) - \xi_i^\lambda \right) \left[1 - \left\{ u_i^l(\xi^\lambda) \right\}^2 \right] \quad (l = L, \quad i = 1, \dots, n_l) \quad (\text{数式8})$$

$$\delta_i^l(\xi^\lambda) = \left(\sum_{j=1}^{n_{l+1}} \delta_j^{l+1}(\xi^\lambda) w_{ji}^{l+1} \right) \left[1 - \left\{ u_i^l(\xi^\lambda) \right\}^2 \right] \quad (l = 1, \dots, L-1, \quad i = 1, \dots, n_l) \quad (\text{数式9})$$

である。したがって、 w_{ij}^l を更新する学習則は

$$w_{ij}^l \leftarrow w_{ij}^l - \frac{\Delta}{\Lambda} \sum_{\lambda=1}^{\Lambda} \delta_i^l(\xi^\lambda) u_j^{l-1}(\xi^\lambda) \quad (l = 1, \dots, L, \quad i = 1, \dots, n_l, \quad j = 0, \dots, n_{l-1}) \quad (\text{数式10})$$

となる。ここで、 Δ は学習の速さを決める正定数であり、学習係数である。数式10を、無作為に選択された一つの教師データ λ で結合係数を更新するように書き換えると、

$$w_{ij}^l \leftarrow w_{ij}^l - \Delta \delta_i^l(\xi^\lambda) u_j^{l-1}(\xi^\lambda) \quad (l = 1, \dots, L, \quad i = 1, \dots, n_l, \quad j = 0, \dots, n_{l-1}) \quad (\text{数式11})$$

となる。ただし、改めて

$$\Delta = \frac{\Delta}{\Lambda} \quad (\text{数式12})$$

とおいた。