誤差逆伝播法

中嶋 慧

April 7, 2024, v3

Abstract

『ディープラーニングと物理学』[1] に基づいて誤差逆伝播法[1] [2] について解説する。

1 逆伝播法

第 0 層が入力で、第 N 層が出力のニューラルネットワークを考える。出力を、

$$|h_0\rangle := {}^t(x_1, x_2, \cdots, x_n) \tag{1.1}$$

と表す。また、

$$(h_l)_i = \sigma_l((J_l|h_{l-1}\rangle)_i + (b_l)_i) \quad (l = 1, 2, \cdots, N, i = 1, 2, \cdots, m_l) \tag{1.2}$$

であり、

$$|h_N\rangle = {}^t((h_N)_1, (h_N)_1, \cdots, (h_N)_m) \tag{1.3}$$

が出力である。σ_l は活性化関数であり、J_l は重みを表す $m_l \times m_{l-1}$ 行列, $(b_l)_i$ はバイアスを表す。$|i\rangle_l$ を i 番目の成分が 1 で他の成分が 0 のベクトルとし、${}_l\langle i|$ をその転置とすると、(1.2) は、

$$|h_l\rangle = \sum_i |i\rangle_l \sigma_l({}_l\langle i|J_l|h_{l-1}\rangle + {}_l\langle i|b_l\rangle) \tag{1.4}$$

となる。さて、

$$(\tilde{h}_l)_i := (h_l)_i \quad (i = 1, 2, \cdots, m_l), \quad (\tilde{h}_l)_0 := 1, \tag{1.5}$$

$$(\mathcal{J}_l)_{ik} := (J_l)_{ik}, \quad (\mathcal{J}_l)_{i0} := (b_l)_i \tag{1.6}$$

とすると、

$$|h_l\rangle = \sum_i |i\rangle_l \sigma_l({}_l\langle i|\mathcal{J}_l|\tilde{h}_{l-1}\rangle), \tag{1.7}$$

$$|\tilde{h}_l\rangle = |\tilde{0}\rangle_l + \sum_{i=1}^{m_l} |\tilde{i}\rangle_l \sigma_l({}_l\langle i|\mathcal{J}_l|\tilde{h}_{l-1}\rangle) \tag{1.8}$$

---

[1] 伝播は「でんぱ」であり、「でんぱん」ではないらしい。

[2] この名前だと誤差が逆方法に伝播していくような印象を受ける。誤差が順方向に伝播するのはイメージしやすい (?) が、逆方法に伝播するとは？。逆伝播法なら変な誤解は生じないと思う。

となる。ここで、$|\tilde{\alpha}\rangle_l$ は α 番目の成分が 1 で他の成分が 0 の (m_l+1) 成分ベクトルである。

さて、(1.8) の変分は、

$$
\begin{aligned}
\delta|\tilde{h}_l\rangle &= \sum_{i=1}^{m_l} |\tilde{i}\rangle_l \sigma_l'({}_l\langle i|\mathcal{J}_l|\tilde{h}_{l-1}\rangle)_l\langle i|\Big(\delta\mathcal{J}_l|\tilde{h}_{l-1}\rangle + \mathcal{J}_l\delta|\tilde{h}_{l-1}\rangle\Big) \\
&= \mathcal{G}_l\Big(\delta\mathcal{J}_l|\tilde{h}_{l-1}\rangle + \mathcal{J}_l\delta|\tilde{h}_{l-1}\rangle\Big) \qquad (1.9)
\end{aligned}
$$

である。ここで、

$$
\mathcal{G}_l := \sum_{i=1}^{m_l} |\tilde{i}\rangle_l \sigma_l'({}_l\langle i|\mathcal{J}_l|\tilde{h}_{l-1}\rangle)_l\langle i| \qquad (1.10)
$$

である。また、

$$
\delta|h_l\rangle = G_l\Big(\delta\mathcal{J}_l|\tilde{h}_{l-1}\rangle + \mathcal{J}_l\delta|\tilde{h}_{l-1}\rangle\Big), \qquad (1.11)
$$

$$
G_l := \sum_{i=1}^{m_l} |i\rangle_l \sigma_l'({}_l\langle i|\mathcal{J}_l|\tilde{h}_{l-1}\rangle)_l\langle i| \qquad (1.12)
$$

である。

いま、E を誤差関数とすると、これは $|h_N\rangle$ だけの関数であり、

$$
\delta E = \langle l_0|\delta|h_N\rangle \qquad (1.13)
$$

と書ける[3)] 。ここで、

$$
(\langle l_0|)_j := \frac{\partial E}{\partial(h_N)_j} \qquad (1.14)
$$

である。よって、

$$
\begin{aligned}
\delta E &= \langle l_0|G_N\Big(\delta\mathcal{J}_N|\tilde{h}_{N-1}\rangle + \mathcal{J}_N\delta|\tilde{h}_{N-1}\rangle\Big) \\
&= \langle l_1|\delta\mathcal{J}_N|\tilde{h}_{N-1}\rangle + \langle l_1|\mathcal{J}_N\delta|\tilde{h}_{N-1}\rangle \\
&= \langle l_1|\delta\mathcal{J}_N|\tilde{h}_{N-1}\rangle + \langle l_2|\delta\mathcal{J}_{N-1}|\tilde{h}_{N-2}\rangle + \cdots + \langle l_N|\delta\mathcal{J}_1|\tilde{h}_0\rangle \qquad (1.15)
\end{aligned}
$$

となる。ここで、

$$
\langle l_1| := \langle l_0|G_N, \qquad (1.16)
$$

$$
\begin{aligned}
\langle l_{k+1}| &:= \langle l_k|\mathcal{J}_{N-k+1}\mathcal{G}_{N-k} \\
&= \langle l_k|J_{N-k+1}G_{N-k} \quad (k=1,2,\cdots,N-1) \qquad (1.17)
\end{aligned}
$$

---

3)例えば、多クラス分類では「出力」が、

$$
p_i = e^{(J_N|h_{N-1}\rangle)_i+(b_N)_i} / \sum_j e^{(J_N|h_{N-1}\rangle)_j+(b_N)_j}
$$

である。これは、

$$
p_i = (h_N)_i/\sum_j (h_N)_j, \quad (h_N)_i = \sigma_N((J_N|h_{N-1}\rangle)_i + (b_N)_i), \quad \sigma_N(y) := e^y
$$

と書ける。誤差関数は p_i で表せるので、h_N だけで表せる。

である。まとめると、

$$\begin{aligned}\delta E &= \sum_{k=1}^{N} \langle l_{N-k+1} | \delta \mathcal{J}_k | \tilde{h}_{k-1} \rangle \\ &= \sum_{k=1}^{N} \sum_{a=1}^{m_k} \sum_{b=0}^{m_{k-1}} (\delta \mathcal{J}_k)_{ab} (\langle l_{N-k+1} |)_a (\tilde{h}_{k-1})_b \end{aligned} \tag{1.18}$$

なので、

$$\frac{\partial E}{\partial (\mathcal{J}_k)_{ab}} = (\langle l_{N-k+1} |)_a (\tilde{h}_{k-1})_b \tag{1.19}$$

を得る。ここで、$(\mathcal{J}_k)_{ai} = (J_k)_{ai}$ $(i = 1 \cdots, m_{k-1})$, $(\mathcal{J}_k)_{a0} = (b_k)_a$ であったから、

$$\frac{\partial E}{\partial (J_k)_{ai}} = (\langle l_{N-k+1} |)_a (\tilde{h}_{k-1})_i = (\langle l_{N-k+1} |)_a (h_{k-1})_i, \tag{1.20}$$

$$\frac{\partial E}{\partial (b_k)_a} = (\langle l_{N-k+1} |)_a \tag{1.21}$$

である。

　まとめると、誤差関数の J_k や b_k についての導関数は、(1.20), (1.21) で求めることができ、$|h_k\rangle$ は (1.2), $\langle l_k|$ は (1.16), (1.17) で求められる。(1.16), (1.17) は伝播の方向が逆なので、この方法は逆伝播法と呼ばれる。

References

[1] 田中章詞, 富谷昭夫, 橋本幸士『ディープラーニングと物理学』講談社サイエンティフィク, 2019 年