誤差逆伝播法

迫田真太郎

平成 32 年 2 月 28 日

1 パーセプトロンの学習則とデルタ則

ニューラルネットの起源は**人工ニューロン**であった. これは活性化関数が階段関数であり, 入出力値が 0,1 の 2 値であるニューラルネットの一例であるとみなせる. 信号が離散値であるため勾配が取れない. 従ってパーセプトロンでは専用の学習則がいくつか提案されている.

中でも代表的なものは 1958 年にローゼンブラットが提唱したものである. 今, 中間層なしの古典パーセプトロンで出力ユニットが 1 つであるものを考えることとする.

一つの訓練サンプル $\{\boldsymbol{x}, y\}$ について, 入力 $\boldsymbol{x}$ に対してパーセプトロンの出力が $\hat{y}(\boldsymbol{x})$ であったとき, 出力が間違っているパターンは次の 2 通りになる.

1. 目標出力が $y = 1$ であったのに実際の出力が $\hat{y}(\boldsymbol{x}) = 0$ であった場合
   出力層にやってくるシグナルの量が足りなかったということなので, 発火したユニットとの結合パラメータを大きくすればよい. したがってパラメータの更新式は η を学習率として
   $$w_i \longleftarrow w_i + \eta x_i \tag{1}$$
   となる. $x_i = 0$ となるような i については w_i は不変である.
2. 目標出力が $y = 0$ であったのに実際の出力が $\hat{y}(\boldsymbol{x}) = 1$ であった場合
   出力層にやってくるシグナルが強すぎたということなので, 発火したユニットとの結合パラメータを小さくすればよい. したがって
   $$w_i \longleftarrow w_i - \eta x_i \tag{2}$$

これらの更新式は次のようにまとめることができる.

$$w_i \longleftarrow w_i - \eta(\hat{y}(\boldsymbol{x}) - y)x_i \tag{3}$$

これを**パーセプトロンの学習則**といい, 線形分離可能な問題に対しては訓練データを十分に与えることでパラメータが有限回で最適値に収束していくことが証明されている.

パラメータが有限回の更新で最適値に収束することの証明

考えているパーセプトロンは $\hat{y}(\boldsymbol{x}) = \mathrm{sgn}(\boldsymbol{w}^t\boldsymbol{x})$ として与えられる.$\boldsymbol{w}^t\boldsymbol{x}$ の符号を見ているだけなので $\boldsymbol{w}$ をすべて正の定数倍しても出力は変わらない. 線形分離可能な訓練データ $\{\boldsymbol{x}_1, y_1\}, \{\boldsymbol{x}_2, y_2\}, \ldots, \{\boldsymbol{x}_N, y_N\}$ について, 簡単のため $y_i < 0$ であるものはすべて $\boldsymbol{x}_i = -\boldsymbol{x}_i$, $y_i = -y_i$ と変換しておくこととする. 訓練データを繰り返し学習していく中で, 誤識別された学習パターンを重複を許して順番に並べて $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \ldots$ とする. 重みパラメータの初期値を 0 とした場合, 正の定数倍に対しての不変性から学習率はどのようにとっても変わらないため 1 として, k 回目の誤認識をもとにパラメータを $\boldsymbol{w}^k \to \boldsymbol{w}^{k+1}$ へと更新する式は

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \boldsymbol{x}^k \tag{4}$$

である. ここで解となる重みベクトルの一つを $\hat{\boldsymbol{w}}$ とすると

$$\hat{\boldsymbol{w}}^t\boldsymbol{x}_i > 0 \qquad (i = 1, 2, \ldots, N) \tag{5}$$

であり, 式 (**??**) の両辺からこれを引くと

$$(\boldsymbol{w}^{k+1} - \hat{\boldsymbol{w}}) = (\boldsymbol{w}^k - \hat{\boldsymbol{w}}) + \boldsymbol{x}^k \tag{6}$$

であり, 両辺を自乗して

$$\begin{aligned}
(\boldsymbol{w}^{k+1} - \hat{\boldsymbol{w}})^2 &= ((\boldsymbol{w}^k - \hat{\boldsymbol{w}}) + \boldsymbol{x}^k)^2 && (7)\\
&= (\boldsymbol{w}^k - \hat{\boldsymbol{w}})^2 \\
&\quad +2(\boldsymbol{w}^k - \hat{\boldsymbol{w}})^t\boldsymbol{x}^k + (\boldsymbol{x}^k)^2 && (8)\\
&= (\boldsymbol{w}^k - \hat{\boldsymbol{w}})^2 + 2(\boldsymbol{w}^k)^t\boldsymbol{x}^k \\
&\quad -2\hat{\boldsymbol{w}}^k\boldsymbol{x}^k + (\boldsymbol{x}^k)^2 && (9)
\end{aligned}$$

となる.$\boldsymbol{x}^k$ は誤認識されたパターンなので $(\boldsymbol{w}^k)^t\boldsymbol{x}^k \leq 0$ が成り立ち, 従って式 (**??**) は

$$(\boldsymbol{w}^{k+1} - \hat{\boldsymbol{w}})^2 \leq (\boldsymbol{w}^k - \hat{\boldsymbol{w}})^2 - 2\hat{\boldsymbol{w}}^k\boldsymbol{x}^k + (\boldsymbol{x}^k)^2 \tag{10}$$

ここで $\hat{\boldsymbol{w}}$ の正の定数倍は同じく解であることを利用すると正の定数 α を用いて $\hat{\boldsymbol{w}} \to \alpha\hat{\boldsymbol{w}}$ と書き換られ

$$(\boldsymbol{w}^{k+1} - \alpha\hat{\boldsymbol{w}})^2 \leq (\boldsymbol{w}^k - \alpha\hat{\boldsymbol{w}})^2 - 2\alpha\hat{\boldsymbol{w}}^k\boldsymbol{x}^k + (\boldsymbol{x}^k)^2 \tag{11}$$

となる. 不等式の条件を変化させないよう定数 β, γ を

$$\beta \stackrel{\text{def}}{=} \max_{p=1,\ldots,n} (\boldsymbol{x}_p)^2 \tag{12}$$

$$\gamma \stackrel{\text{def}}{=} \min_{p=1,\ldots,n} \hat{\boldsymbol{w}}^t\boldsymbol{x}_p > 0 \tag{13}$$

と定義する. すると

$$(\boldsymbol{w}^{k+1} - \alpha\hat{\boldsymbol{w}})^2 \leq (\boldsymbol{w}^k - \alpha\hat{\boldsymbol{w}})^2 - 2\alpha\gamma + \beta \tag{14}$$

であり, $\alpha = \frac{\beta}{\gamma}$ とし添え字 k 順番にずらしていくと

$$\begin{aligned}
(\boldsymbol{w}^{k+1} - \alpha\hat{\boldsymbol{w}})^2 &\leq (\boldsymbol{w}^k - \alpha\hat{\boldsymbol{w}})^2 - \beta && (15)\\
&\leq (\boldsymbol{w}^{k-1} - \alpha\hat{\boldsymbol{w}})^2 - 2\beta && (16)\\
&\cdots \\
&\leq (\boldsymbol{w}^1 - \alpha\hat{\boldsymbol{w}})^2 - k\beta && (17)
\end{aligned}$$

という不等式が成り立つ. この左辺は 0 以上なので

$$0 \leq (\boldsymbol{w}^1 - \alpha\hat{\boldsymbol{w}})^2 - k\beta \tag{18}$$

よって

$$k \leq \frac{(\boldsymbol{w}^1 - \alpha\hat{\boldsymbol{w}})^2}{\beta} \tag{19}$$

k が定数で上から抑えられるため有限回で収束することが示された.

2 誤差逆伝播法とは

3 デルタの意味

4 誤差逆伝播法の利点

5 勾配消失, 勾配爆発問題

5.1 勾配消失, 勾配爆発問題とは

5.2 勾配消失への対策

5.2.1 事前学習

5.2.2 活性化関数の工夫

参考文献

[1] 瀧雅人. これならわかる深層学習入門. p114-131

[2] 鍋谷昴一. 単純パーセプトロンの収束定理と限界. http://starpentagon.net/analytics/simple_perceptron/