学籍番号：

2017 年 6 月 29 日

数理脳科学　　　　名前：　　　　得点：

小テスト 【バックプロパゲーション】

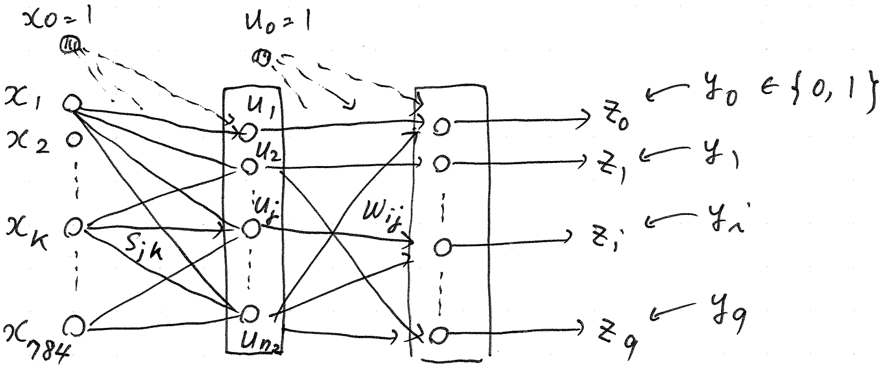

図 1: 3 層神経回路モデル（実質は 2 層）

1. ニューロンの出力関数が

$$f(u) \quad = \quad \frac{1}{1-\mathrm{e}^{-u/T}} \tag{1}$$

の場合を考える．

$$f'(u) \quad = \quad \frac{1}{T}f(u)(1-f(u)) \tag{2}$$

がであることを確かめよ．

(a) $1-f(u)$ を求めよ．

(b) $f'(u)$ を求めよ．　微分の公式：$\left\{\dfrac{1}{g(x)}\right\}' = -\dfrac{g'(x)}{\{g(x)\}^2}$

数理脳科学
2017 年 6 月 29 日

課題 4: 教師あり学習

3 層回路（図 2）を使い，手書き数字データ（28×28）を学習してみよう．この教師あり学習の例を通して，誤差逆伝搬法（バックプロパゲーション）を説明する．

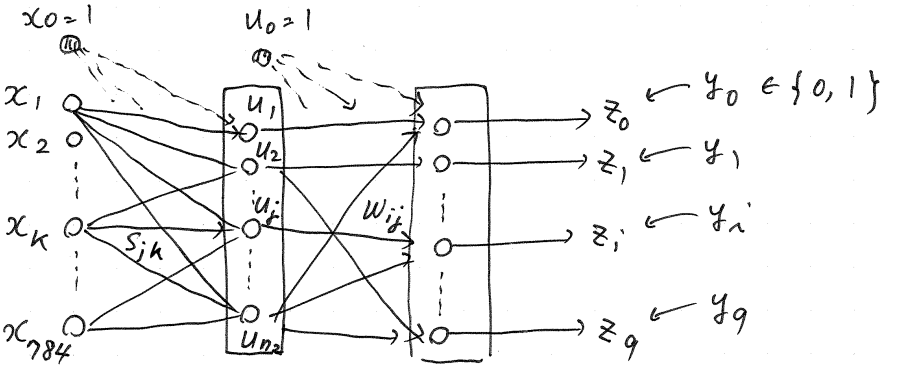

図 2: 3 層神経回路モデル（実質は 2 層）

1　教師あり学習

この課題では，第 2 層として，$n_2 = 100$ 個の素子，第 3 層として，10 個の出力素子を用意する．入力 $\boldsymbol{x}$ に対する望ましい出力を $\boldsymbol{y}(\boldsymbol{x})$ と書く．第 3 層の素子の出力は線形としよう．つまり，第 3 層の i 番目の素子の出力は

$$z_i \quad = \quad \sum_{j=1}^{n_2} w_{ij} u_j - \theta_i \tag{3}$$

とする．ここで，u_j は第 2 層の活動（$j = 1, \cdots, n_2$），θ_i は出力層の i 番目の素子のしきい値（$i = 0, \cdots, 9$），w_{ij} は，第 2 層 j 番目の素子から第 3 層 i 番目の素子への結合の重みである．目的は，ある入力（数字画像）$\boldsymbol{x}$ が与えられた時,

$$E \quad = \quad \frac{1}{2} \sum_{i=0}^{9} \left(z_i(\boldsymbol{x}) - y_i(\boldsymbol{x}) \right)^2 \tag{4}$$

を最小化することだと考えると,

$$\Delta w_{ij} \quad = \quad -\mu \frac{\partial E}{\partial w_{ij}} = -\mu \left(\sum_{j=1}^{n_2} w_{ij} u_j(\boldsymbol{x}) - \theta_i - y_i(\boldsymbol{x}) \right) u_j(\boldsymbol{x}) \tag{5}$$

$$\Delta \theta_i \quad = \quad -\mu \frac{\partial E}{\partial \theta_i} = \mu \left(\sum_{j=1}^{n_2} w_{ij} u_j(\boldsymbol{x}) - \theta_i - y_i(\boldsymbol{x}) \right) \tag{6}$$

と学習すればよい．ここで学習係数 μ は $\mu = 0.05$ などと設定する．ちなみに $u_0 = 1$ という常に興奮する 0 番素子を考えると，$w_{i0} = -\theta_i$ で

$$
\begin{aligned}
z_i &= \sum_{j=0}^{n_2} w_{ij} u_j & (7) \\
\Delta w_{ij} &= -\mu \frac{\partial E}{\partial w_{ij}} = -\mu \left(\sum_{j=0}^{n_2} w_{ij} u_j(\boldsymbol{x}) - y_i(\boldsymbol{x}) \right) u_j(\boldsymbol{x}) & (8)
\end{aligned}
$$

とすっきり書ける．入力 $\boldsymbol{x}$ に対し，出力層の 10 個の素子のうち，もっとも大きな出力値をもつものを回路の出力とする．

学習用データは 60000 枚用意されているが，全部使って学習すると時間がかかるので，そのうち，最初の 1000 枚だけを学習に使って，性能を比較してみよう．正解率は，学習に使ったデータ 1000 枚と，学習に使っていない 1000 枚のテスト用データの両方を使って調べてみる．これらを訓練誤差と汎化誤差とよぼう．例えば 100 回の学習につき 1 回，この統計量を計算すれば，学習がどのように進んでいるのか，把握できる．これが一つの方法．

もう一つの方法は，10 個の素子で，別々に，ROC カーブを描くこともできる（各素子に対し，学習データ，テストデータで，2 本の曲線が描ける）．この場合，出力が 0.5 以上だと，その数字が出現したとする．他の評価方法もありうるので，いろいろ試してみるのがよい．

とくに事前学習せず，3 層回路を，バックプロパゲーションで学習する．ここでは値の設定例を記述しておく．中間層と出力層の間の学習係数は $\mu = 0.01$.

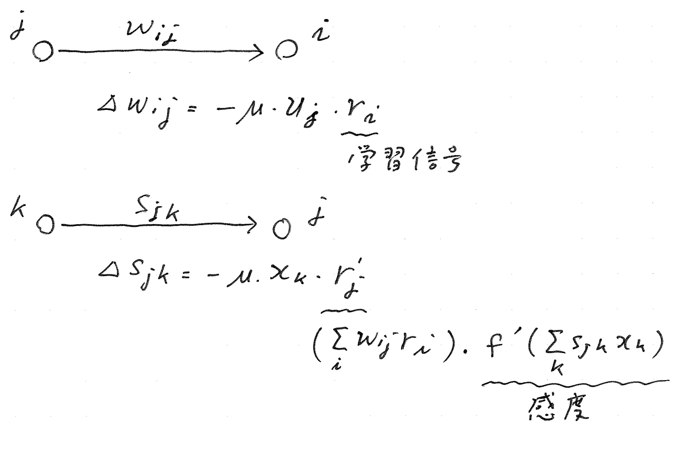

図 3: 入力信号と出力素子の学習信号との Hebb 学習

バックプロパゲーションは，アルゴリズムだけ書いておこう．これは Hebb 学習を一般化したもであると考えれば，わかりやすい（図 3）．第 2 層の j 番目の素子と，第 3 層の i 番

目の素子間の結合の学習は

$$\Delta w_{ij} = -\mu \frac{\partial E}{\partial w_{ij}} = -\mu \left(\sum_{j=0}^{n_2} w_{ij} u_j(\boldsymbol{x}) - y_i(\boldsymbol{x}) \right) u_j(\boldsymbol{x}) \tag{9}$$

$$= -\mu r_i u_j \tag{10}$$

と見ると，これは第 3 層 i 番目の素子の学習信号 r_i と第 2 層 j 番目の素子の出力との Hebb 学習である（$i = 1, \cdots, 9, j = 0, \cdots, n_2$）．ここでは天下りに，アルゴリズムだけを示そう．第 1 層 k 番目の素子と第 2 層 j 番目の素子の結合係数を s_{jk} とすると

$$\Delta s_{jk} = -\mu \frac{\partial E}{\partial s_{jk}} = -\mu \left(\sum_{i=1}^{10} w_{ij} r_i \right) f' \left(\sum_{k=0}^{n_1} s_{jk} x_k \right) x_k \tag{11}$$

$$= -\mu r'_j x_k \tag{12}$$

と書ける（$k = 0, \cdots, n_1, j = 1, \cdots, n_2$）．ここで $f'(u)$ は，第 2 層の素子の入出力関数 $f(u)$ の微分で，

$$f(u) = \frac{1}{1 - \mathrm{e}^{-u/T}} \tag{13}$$

の場合

$$f'(u) = \frac{1}{T} f(u)(1 - f(u)) \tag{14}$$

であり（実際に微分し確認してみよ），出力が a なら，$f' = a(1 - a)$ と簡単に計算できる．