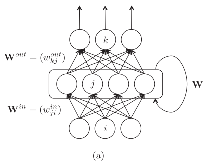
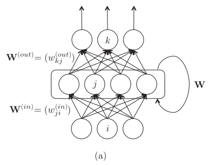
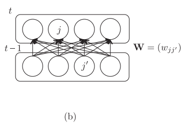
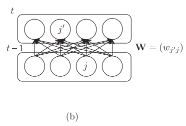
# 『深層学習』第 1,2 刷正誤表

この度は，標記書籍をお買い求めいただき誠にありがとうございました。
標記書籍に誤りがありました。訂正し，深くお詫び申し上げます。

| ページ数 | 位置 | 誤 | 正 |
|---|---|---|---|
| 1 | 下から2行目〜 | 入力層から逆に伝播させ，各層の重みの勾配を計算するという方法です．この計算の際，入力層から離れた | 出力層から逆に伝播させ，各層の重みの勾配を計算するという方法です．この計算の際，出力層から離れた |
| 2 | 2 行目 | いわゆる勾配消失問題という呼ばれる現象が | いわゆる勾配消失問題と呼ばれる現象が |
| 3 | 8 行目 | 異なる原理に基づいて行わるものの， | 異なる原理に基づいて行われるものの， |
| 11 | 3 行目 | retified linear function | rectified linear function |
| 15 | 下から1行目 | 正接双曲線関数 | 双曲線正接関数 |
| 18 | 5〜6行目 | その場合は出力層はユニット 2 つとなり， | その場合は出力層はユニットが 2 つとなり， |
| 34 | 2 行目 | 次に平均を差し引いた後のデータの各成分 $x_{ni}$ を，その標準偏差 | またデータの各成分を平均を差し引いた上で，標準偏差 |
| 35 | 下から5行目 | model avaraging | model averaging |
| 37 | 下から7行目 | $\Delta\mathbf{w}^{(t-1)} \equiv \mathbf{w}^{(t-1)} - \mathbf{w}^{(t-2)}$ | $\Delta\mathbf{w}^{(t-1)} \equiv \mathbf{w}^{(t)} - \mathbf{w}^{(t-1)}$ |
| 39 | 19 行目 | $\sigma = \sigma_u/M^{-2}$ | $\sigma = \sigma_u/M^{1/2}$ |
| 49 | 10 行目 | $E_n = d\log y + (1-d)\log(1-y)$ | $E_n = -d\log y - (1-d)\log(1-y)$ |
| 49 | 13 行目 | $\delta^{(L)} = \frac{d}{y}\cdot\frac{\mathrm{d}y}{\mathrm{d}u} + \frac{1-d}{1-y}\left(-\frac{\mathrm{d}y}{\mathrm{d}u}\right) = d(1-y) - (1-d)y = d-y$ | $\delta^{(L)} = -\frac{d}{y}\cdot\frac{\mathrm{d}y}{\mathrm{d}u} - \frac{1-d}{1-y}\left(-\frac{\mathrm{d}y}{\mathrm{d}u}\right) = -d(1-y) + (1-d)y = y-d$ |
| 50 | 下から7行目 | 順伝播計算は，$\mathbf{U}^{(1)} \equiv \mathbf{X}$ | 順伝播計算は，$\mathbf{Z}^{(1)} \equiv \mathbf{X}$ |
| 51 | 11 行目 | 逆伝播計算は $\boldsymbol{\Delta}^{(L)} = \mathbf{D} - \mathbf{Y}$ とした後， | 逆伝播計算は $\boldsymbol{\Delta}^{(L)} = \mathbf{Y} - \mathbf{D}$ とした後， |
| 51 | 下から4行目 | $E = \sum_{n=1}^{N} E_n(\mathbf{W})$ | $E = \sum_{n=1}^{N} E_n(\mathbf{W})/N$ |
| 52 | 3 行目 | $E = \sum_{n=1}^{N} E_n(\mathbf{W})$ | $E = \sum_{n=1}^{N} E_n(\mathbf{W})/N$ |
| 52 | 2 番目の式 | $\partial\mathbf{b}^{(l)} = \frac{1}{N}\Delta^{(l)}\mathbf{1}_N^\top$ | $\partial\mathbf{b}^{(l)} = \frac{1}{N}\Delta^{(l)}\mathbf{1}_N$ |
| 52 | 下から1行目 | $\frac{\partial E}{\partial w_{ji}^{(l)}} = \frac{E(\ldots, w_{ij}^{(l)}+\varepsilon, \ldots) - E(\ldots, w_{ij}^{(l)}, \ldots)}{\varepsilon}$ | $\frac{\partial E}{\partial w_{ji}^{(l)}} = \frac{E(\ldots, w_{ji}^{(l)}+\varepsilon, \ldots) - E(\ldots, w_{ji}^{(l)}, \ldots)}{\varepsilon}$ |
| 53 | 1 行目 | ここで $E(\ldots, w_{ij}^{(l)}+\varepsilon, \ldots)$ という表記は，$E(\mathbf{w})$ の変数 $\mathbf{w}$ のうち $w_{ij}^{(l)}$ だけに $\varepsilon$ を加算するという意味です．$\varepsilon \to 0$ の極限をとると，上の式は偏微分 $\partial E(\mathbf{w})/\partial w_{ij}^{(l)}$ の定義に一致します． | ここで $E(\ldots, w_{ji}^{(l)}+\varepsilon, \ldots)$ という表記は，$E(\mathbf{w})$ の変数 $\mathbf{w}$ のうち $w_{ji}^{(l)}$ だけに $\varepsilon$ を加算するという意味です．$\varepsilon \to 0$ の極限をとると，上の式は偏微分 $\partial E(\mathbf{w})/\partial w_{ji}^{(l)}$ の定義に一致します． |

| | | | |
|---|---|---|---|
| 62 | 10 行目 | 任意の正則な $D_y \times D_y$ 行列 $\mathbf{Q}$ に対し, | 任意の $D_y \times D_y$ の直交行列 $\mathbf{Q}$ に対し, |
| 63 | 1 行目 | い $(D_y \leq D_x)$ 場合であっても, | い $(D_y > D_x)$ 場合であっても, |
| 65 | 12 行目 | $\dfrac{\partial \hat{\rho}_j}{\partial u_j^{(l)}} = f'(u_j^{(l)})$ | $\dfrac{\partial \hat{\rho}_j}{\partial u_j^{(l)}} = \dfrac{1}{N} f'(u_j^{(l)})$ |
| 65 | 式 (5.4) | $\delta_j^{(l)} = \left\{ \cdots + \left( -\dfrac{\rho}{\hat{\rho}_j} + \dfrac{1-\rho}{1-\hat{\rho}_j} \right) \right\} f'(u_j^{(l)})$ | $\delta_j^{(l)} = \left\{ \cdots + \beta \left( -\dfrac{\rho}{\hat{\rho}_j} + \dfrac{1-\rho}{1-\hat{\rho}_j} \right) \right\} f'(u_j^{(l)})$ |
| 65 | 脚注 | *6　2 層のネットワークに限れば, スパース正則化の対象となる中間層より下に重みを持つ層はないので, パラメータの更新式を修正するだけで済みます. | 削除 |
| 85 | 図 6.5<br>図 6.6 | 0.01 0.08 0.01<br>0.08 0.62 0.08<br>0.01 0.08 0.01 | 0.01 0.08 0.01<br>0.08 0.64 0.08<br>0.01 0.08 0.01 |
| 95 | 7 行目 | $\bar{x}_{ij} = \dfrac{1}{K} \displaystyle\sum_{k=0}^{K-1} \sum_{(p,q) \in P_{ij}} w_{pq} x_{i+p,j+q,k}$ | $\bar{x}_{ij} = \dfrac{1}{K} \displaystyle\sum_{k=0}^{K-1} \sum_{(p,q) \in P_{ij}} w_{pqk} x_{i+p,j+q,k}$ |
| 96 | 1 番目<br>2 番目<br>の式 | $\mathbf{u}^{(l)} = \mathbf{W}^{(l)} \mathbf{u}^{(l-1)} + \mathbf{b}^{(l)}$<br>$\mathbf{z}^{(l)} = f^{(l)}(\mathbf{u})$ | $\mathbf{u}^{(l)} = \mathbf{W}^{(l)} \mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}$<br>$\mathbf{z}^{(l)} = f^{(l)}(\mathbf{u}^{(l)})$ |
| 99 | 10 行目 | 約 200,000 ミニバッチ, すなわち 2 万 $\times$128/(総学習サンプル = 数百万) = 約 25 エポックほど | 約 200,000 ミニバッチ, すなわち 20 万 $\times$128/(総学習サンプル = 約百万) = 約 25 エポックほど |
| 114 | 13 行目 | ネットワーク内部にある帰環路に | ネットワーク内部にある帰還路に |
| 114 | 図 7.3<br>(a) |  |  |
| 114 | 図 7.3<br>(b) |  |  |
| 115 | 式 (7.1) | $E(\mathbf{w}) = \displaystyle\sum_n \cdots$ | $E(\mathbf{w}) = -\displaystyle\sum_n \cdots$ |
| 116 | 10 行目 | をそれぞれ $\mathbf{v}^t = (v_j^t)$ と $\mathbf{y}^t = (y_j^t)$ のように | をそれぞれ $\mathbf{v}^t = (v_k^t)$ と $\mathbf{y}^t = (y_k^t)$ のように |
| 119 | 式 (7.7)<br>第 2 項 | $\displaystyle\sum_{j'} w_{jj'} \delta_{j'}^{t+1}$ | $\displaystyle\sum_{j'} w_{j'j} \delta_{j'}^{t+1}$ |
| 120 | 1 行目 | $\dfrac{\partial E}{\partial w_{jj'}} = \displaystyle\sum_{t=1}^{T} \dfrac{\partial E}{\partial u_j^t} \dfrac{\partial u_j^t}{\partial w_{jj'}} = \sum_{t=1}^{T} \delta_j^t z_j^{t-1}$ | $\dfrac{\partial E}{\partial w_{jj'}} = \displaystyle\sum_{t=1}^{T} \dfrac{\partial E}{\partial u_j^t} \dfrac{\partial u_j^t}{\partial w_{jj'}} = \sum_{t=1}^{T} \delta_j^t z_{j'}^{t-1}$ |

| | | | |
|---|---|---|---|
| 120 | 3 行目 | $\frac{\partial E}{\partial w_{kj}^{out}} = \sum_{t=1}^{T} \frac{\partial E}{\partial v_k^t} \frac{\partial v_k^t}{\partial w_{kj}^{out}} = \sum_{t=1}^{T} \delta_j^t z_j^t$ | $\frac{\partial E}{\partial w_{kj}^{out}} = \sum_{t=1}^{T} \frac{\partial E}{\partial v_k^t} \frac{\partial v_k^t}{\partial w_{kj}^{out}} = \sum_{t=1}^{T} \delta_k^{out,t} z_j^t$ |
| 122 | 下から<br>2 行目 | 内部のメモリセルは | 内部のメモリセル（図 7.7a）は |
| 123 | 2,3,<br>4,6 式 | $u_j^t = \sum_j w_{ji}^{(in)} x_j^t + \sum_{j'} w_{jj'} z_j^{t-1}$<br>$\cdots = f\left(\sum_j w_{ji}^{F,in} x_j^t + \sum_{j'} w_{jj'}^F z_j^{t-1} + s_j^{t-1}\right)$<br>$\cdots = f\left(\sum_j w_{ji}^{I,in} x_j^t + \sum_{j'} w_{jj'}^I z_j^{t-1} + s_j^{t-1}\right)$<br>$\cdots = f\left(\sum_j w_{ji}^{O,in} x_j^t + \sum_{j'} w_{jj'}^O z_j^{t-1} + s_j^t\right)$ | $u_j^t = \sum_j w_{ji}^{in} x_i^t + \sum_{j'} w_{jj'} z_{j'}^{t-1}$<br>$\cdots = f\left(\sum_i w_{ji}^{F,in} x_i^t + \sum_{j'} w_{jj'}^F z_{j'}^{t-1} + w_j^F s_j^{t-1}\right)$<br>$\cdots = f\left(\sum_i w_{ji}^{I,in} x_i^t + \sum_{j'} w_{jj'}^I z_{j'}^{t-1} + w_j^I s_j^{t-1}\right)$<br>$\cdots = f\left(\sum_i w_{ji}^{O,in} x_i^t + \sum_{j'} w_{jj'}^O z_{j'}^{t-1} + w_j^O s_j^t\right)$ |
| 123 | 10 行目<br>追加 | $w_j^F$, $w_j^I$ はそれぞれ，メモリセルから忘却ゲートと入力ゲートの値を決めるユニット（図 7.7 の f と c）への結合の $s_j^{t-1}$ の重みです．これらは下で扱う出力ゲートに関する同様の結合とあわせて，「のぞき穴 (peephole)」結合とも呼ばれています． | |
| | 上の<br>123<br>ページ<br>の訂正<br>および<br>追加に<br>ついて<br>の備考 | 最初に提案された LSTM（$w_j^F = w_j^I = w_j^O = 0$ に相当，文献 [30]）では，出力ゲートが閉じていると各ゲートはセルの状態を知ることができず，タスクによってはこれが問題となる場合がありますが，「のぞき穴」結合はこれを解決します（参考：F.A. Gers, N. N. Schraudolph, and J. Schmidhuber. Learning Precise Timing with LSTM Recurrent Networks. *Journal of Machine Learning Research*, 3:115-143, 2002）. | |
| 124 | 12 行目 | $u_k^{out,t} = \sum_j w_{kj} z_j^t$ であり，<br>$\frac{\partial u_k^{out,t}}{\partial u_j^{O,t}} = w_{kj} f'(u_j^{O,t}) f(s_j^t)$<br>となります．もう一つの伝播先の，次時刻のメモリユニットへの総入力についても同様に計算でき，したがって $g_j^{O,t}$ を出力するユニットのデルタは<br>$\epsilon_j^t = \sum_k w_{kj} \delta_k^{out,t} + \sum_{j'} w_{j'j} \delta_{j'}^{t+1}$ | $v_k^t = \sum_j w_{kj}^{out} z_j^t$ であり，<br>$\frac{\partial v_k^t}{\partial u_j^{O,t}} = w_{kj}^{out} f'(u_j^{O,t}) f(s_j^t)$<br>となります．もう一つの伝播先の，次時刻のメモリユニットへの総入力についても同様に計算でき，したがって $g_j^{O,t}$ を出力するユニットのデルタは<br>$\epsilon_j^t = \sum_k w_{kj}^{out} \delta_k^{out,t} + \sum_{j'} w_{j'j} \delta_{j'}^{t+1}$ |
| 125 | 最初<br>の式 | $\delta_j^{cell,t} = \tilde{\delta}_j^t + g_j^{F,t+1} \phi_j^{t+1} + \delta_j^{I,t+1} + \delta_j^{F,t+1} + \delta_j^{O,t}$ | $\delta_j^{cell,t} = \tilde{\delta}_j^t + g_j^{F,t+1} \delta_j^{cell,t+1} + w_j^I \delta_j^{I,t+1} + w_j^F \delta_j^{F,t+1} + w_j^O \delta_j^{O,t}$ |
| 127 | 9 行目 | ‘*ccba*’ です． | ‘*cbab*’ です． |
| 129 | 10 行目<br>11 行目 | $p(\mathbf{l}\|\mathbf{X}) = \alpha_{\|L'\|,T} + \alpha_{\|L'\|+1,T}$ のように計算できます（図 7.8 にもあるように，有効なパスはいつも，最後の時刻 $t = T$ では $\|L'\|$ か $\|L'\|+1$ のいずれかに到達するはずであることから）. | $p(\mathbf{l}\|\mathbf{X}) = \alpha_{\|\mathbf{l}'\|,T} + \alpha_{\|\mathbf{l}'\|-1,T}$ のように計算できます（図 7.8 にもあるように，有効なパスはいつも，最後の時刻 $t = T$ では $\|\mathbf{l}'\|$ か $\|\mathbf{l}'\| - 1$ のいずれかに到達するはずであることから）. |
| 129 | 最後<br>の式 | $-\sum_n \log p(\mathbf{d}\|\mathbf{X})$ | $-\sum_n \log p(\mathbf{d}_n\|\mathbf{X}_n)$ |

| | | | |
|---|---|---|---|
| 130 | 下から 4 行目 | $\hat{\mathbf{l}} = \mathcal{B}(\boldsymbol{\pi})$ とする方法です. | $\hat{\mathbf{l}} = \mathcal{B}(\hat{\boldsymbol{\pi}})$ とする方法です. |
| 134 | 最初の式 | $\boldsymbol{\Phi}(\mathbf{x}_n \| \boldsymbol{\theta})$ | $\boldsymbol{\Phi}(\mathbf{x}_n, \boldsymbol{\theta})$ |
| 135 | 下から 4 行目 | 例えば $M = 10$ では | 例えば $M = 20$ では |
| 139 | 2 番目の式 | $L(\boldsymbol{\theta}) = \sum_{n=1}^{N} p(\mathbf{v}_n \| \boldsymbol{\theta})$ | $L(\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{v}_n \| \boldsymbol{\theta})$ |
| 141 | 7 行目 | 隠れ変数を指定したときの可視変数の条件付き分布 $p(\mathbf{v} \| \mathbf{h}, \boldsymbol{\theta})$ は，定義により | 可視変数を指定したときの隠れ変数の条件付き分布 $p(\mathbf{h} \| \mathbf{v}, \boldsymbol{\theta})$ は，定義により |
| 149 | 式 (8.20c) | $\Delta b_j = \epsilon \left( h_j^{(0)} - p_j^{(T)} \right)$ | $\Delta b_j = \epsilon \left( p_j^{(0)} - p_j^{(T)} \right)$ |
| 150 | 11 行目 14 行目 | $\sigma(x - i + 5)$ | $\sigma(x - i + 0.5)$ |
| 152 | 最初の式 | $p(\mathbf{h}^{(l)} \| \mathbf{h}^{(l+1)}) = \prod_i \sigma(b_i^{(l)} + \sum_j w_{ij}^{(l+1)} h_j^{(l+1)})$ | $p(h_i^{(l)} = 1 \| \mathbf{h}^{(l+1)}) = \sigma(b_i^{(l)} + \sum_j w_{ij}^{(l+1)} h_j^{(l+1)})$ |
| 152 | 2 番目の式 | $p(h_j^{(l)} \| \mathbf{h}^{(l-1)}) = \cdots$ | $p(h_j^{(l)} = 1 \| \mathbf{h}^{(l-1)}) = \cdots$ |
| 153 | 式の最後の項 | $-\sum_j \sum_k w_{ij}^{(2)} h_j^{(1)} h_j^{(2)}$ | $-\sum_j \sum_k w_{jk}^{(2)} h_j^{(1)} h_k^{(2)}$ |
| 154 | 最後の式 | $p(v_i \| \mathbf{h}^{(1)}) = \cdots$ | $p(v_i = 1 \| \mathbf{h}^{(1)}) = \cdots$ |
| 155 | 最初の式 | $p(h_j^{(l)} \| \mathbf{h}^{(1-1)}) = \cdots$ | $p(h_j^{(l)} = 1 \| \mathbf{h}^{(l-1)}) = \cdots$ |