2015 年 5 月 28 日

正誤表（岡谷貴之著，機械学習プロフェッショナルシリーズ「深層学習」第 1,2 刷）

箇所	誤	正
50 頁最後のパラグラフ	順伝播計算は，$\mathbf{U}^{(1)} \equiv \mathbf{X}$	順伝播計算は，$\mathbf{Z}^{(1)} \equiv \mathbf{X}$
52 頁の 2 番めの式	$$\partial \mathbf{b}^{(l)} = \frac{1}{N}\boldsymbol{\Delta}^{(l)}\mathbf{1}_N^\top$$	$$\partial \mathbf{b}^{(l)} = \frac{1}{N}\boldsymbol{\Delta}^{(l)}\mathbf{1}_N$$
63 頁第 1 行	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 頁式 (5.4)	$$\delta_j^{(l)} = \left\{ \cdots + \left(-\frac{\rho}{\hat{\rho}_j} + \frac{1-\rho}{1-\hat{\rho}_j} \right) \right\} f'(u_j^{(l)})$$	$$\delta_j^{(l)} = \left\{ \cdots + \beta \left(-\frac{\rho}{\hat{\rho}_j} + \frac{1-\rho}{1-\hat{\rho}_j} \right) \right\} f'(u_j^{(l)})$$
96 頁の 1 番目と 2 番めの式	$$\mathbf{u}^{(l)} = \mathbf{W}^{(l)}\mathbf{u}^{(l-1)} + \mathbf{b}^{(l)}$$ $$\mathbf{z}^{(l)} = f^{(l)}(\mathbf{u})$$	$$\mathbf{u}^{(l)} = \mathbf{W}^{(l)}\mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}$$ $$\mathbf{z}^{(l)} = f^{(l)}(\mathbf{u}^{(l)})$$
99 頁の 10 行目	約 200,000 ミニバッチ，すなわち 2 万 × 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 × 128/(総学習サンプル=約百万)=約 25 エポックほど
115 頁の式 (7.1)	$$E(\mathbf{w}) = \sum_n \cdots$$	$$E(\mathbf{w}) = -\sum_n \cdots$$
119 頁の式 (7.7) のカッコ内の第 2 項	$$\sum_{j'} w_{jj'}\delta_{j'}^{t+1}$$	$$\sum_{j'} w_{j'j}\delta_{j'}^{t+1}$$
122 頁下から 2 行目	内部のメモリセルは	内部のメモリセル（図 7.7a）は

<table>
<tr><td>箇所</td><td>誤</td><td>正</td></tr>
<tr><td>123 頁の第 2,3,4,6 式</td><td>$$u_j^t = \sum_j w_{ji}^{(in)} x_j^t + \sum_{j'} w_{jj'} z_j^{t-1}$$ $$\cdots = f\left(\sum_j w_{ji}^{F,in} x_j^t + \sum_{j'} w_{jj'}^F z_j^{t-1} + s_j^{t-1}\right)$$ $$\cdots = f\left(\sum_j w_{ji}^{I,in} x_j^t + \sum_{j'} w_{jj'}^I z_j^{t-1} + s_j^{t-1}\right)$$ $$\cdots = f\left(\sum_j w_{ji}^{O,in} x_j^t + \sum_{j'} w_{jj'}^O z_j^{t-1} + s_j^t\right)$$</td><td>$$u_j^t = \sum_i w_{ji}^{in} x_i^t + \sum_{j'} w_{jj'} z_{j'}^{t-1}$$ $$\cdots = f\left(\sum_i w_{ji}^{F,in} x_i^t + \sum_{j'} w_{jj'}^F z_{j'}^{t-1} + w_j^F s_j^{t-1}\right)$$ $$\cdots = f\left(\sum_i w_{ji}^{I,in} x_i^t + \sum_{j'} w_{jj'}^I z_{j'}^{t-1} + w_j^I s_j^{t-1}\right)$$ $$\cdots = f\left(\sum_i w_{ji}^{O,in} x_i^t + \sum_{j'} w_{jj'}^O z_{j'}^{t-1} + w_j^O s_j^t\right)$$</td></tr>
<tr><td>123 頁の 10 行目に追加</td><td colspan="2">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.</td></tr>
<tr><td>上の 123 頁の訂正および追加についての備考</td><td colspan="2">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. Gers, N. N. Schraudolph, and J. Schmidhuber, Learning Precise Timing with LSTM Recurrent Networks, Journal of Machine Learning Research, 3:115-143, 2002）.</td></tr>
<tr><td>124 頁中央からの 3 つの式</td><td>$u_k^{out,t} = \sum_j w_{kj} z_j^t$ であり， $$\frac{\partial u_k^{out,t}}{\partial u_j^{O,t}} = w_{kj} f'(u_j^{O,t}) f(s_j^t)$$ 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 $$\epsilon_j^t = \sum_k w_{kj} \delta_k^{out,t} + \sum_{j'} w_{j'j} \delta_{j'}^{t+1}$$</td><td>$v_k^t = \sum_j w_{kj}^{out} z_j^t$ であり， $$\frac{\partial v_k^t}{\partial u_j^{O,t}} = w_{kj}^{out} f'(u_j^{O,t}) f(s_j^t)$$ 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 $$\epsilon_j^t = \sum_k w_{kj}^{out} \delta_k^{out,t} + \sum_{j'} w_{j'j} \delta_{j'}^{t+1}$$</td></tr>
<tr><td>125 頁の最初の式</td><td colspan="2">誤： $\delta_j^{cell,t} = \tilde{\delta}_j^t + g_j^{F,t+1} \phi_j^{t+1} + \delta_j^{I,t+1} + \delta_j^{F,t+1} + \delta_j^{O,t}$
正： $\delta_j^{cell,t} = \tilde{\delta}_j^t + g_j^{F,t+1} \delta_j^{cell,t+1} + w_j^I \delta_j^{I,t+1} + w_j^F \delta_j^{F,t+1} + w_j^O \delta_j^{O,t}$</td></tr>
<tr><td>129 頁の 10 行目と 11 行目</td><td>$p(\mathbf{l}|\mathbf{X}) = \alpha_{|L'|,T} + \alpha_{|L'|+1,T}$ 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）.</td><td>$p(\mathbf{l}|\mathbf{X}) = \alpha_{|\mathbf{l}'|,T} + \alpha_{|\mathbf{l}'|-1,T}$ 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）.</td></tr>
<tr><td>129 頁最後の式</td><td>$$-\sum_n \log p(\mathbf{d}|\mathbf{X})$$</td><td>$$-\sum_n \log p(\mathbf{d}_n|\mathbf{X}_n)$$</td></tr>
<tr><td>130 頁下から 4 行目</td><td>$\hat{\mathbf{l}} = \mathcal{B}(\boldsymbol{\pi})$ とする方法です.</td><td>$\hat{\mathbf{l}} = \mathcal{B}(\hat{\boldsymbol{\pi}})$ とする方法です.</td></tr>
<tr><td>134 頁の最初の式中</td><td>$$\Phi(\mathbf{x}_n|\boldsymbol{\theta})$$</td><td>$$\Phi(\mathbf{x}_n, \boldsymbol{\theta})$$</td></tr>
</table>

箇所	誤	正
139 頁の 2 番めの式	$$L(\boldsymbol{\theta}) = \sum_{n=1}^{N} p(\mathbf{v}_n\|\boldsymbol{\theta})$$	$$L(\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\mathbf{v}_n\|\boldsymbol{\theta})$$
141 頁の 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 頁の式 (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 頁の 11 行目と 14 行目の式	$$\sigma(x - i + 5)$$	$$\sigma(x - i + 0.5)$$
152 頁の最初の式	$$p(\mathbf{h}^{(l)}\|\mathbf{h}^{(l+1)}) = \prod_i \sigma(b_i^{(l)} + \sum_j w_{ij}^{(l+1)} h_j^{(l+1)})$$	$$p(h_i^{(l)} = 1\|\mathbf{h}^{(l+1)}) = \sigma(b_i^{(l)} + \sum_j w_{ij}^{(l+1)} h_j^{(l+1)})$$
152 頁の 2 番めの式	$$p(h_j^{(l)}\|\mathbf{h}^{(l-1)}) = \cdots$$	$$p(h_j^{(l)} = 1\|\mathbf{h}^{(l-1)}) = \cdots$$
153 頁の式の最後の項	$$-\sum_j \sum_k w_{ij}^{(2)} h_j^{(1)} h_j^{(2)}$$	$$-\sum_j \sum_k w_{jk}^{(2)} h_j^{(1)} h_k^{(2)}$$
154 頁の最後の式	$$p(v_i\|\mathbf{h}^{(1)}) = \cdots$$	$$p(v_i = 1\|\mathbf{h}^{(1)}) = \cdots$$
155 頁の最初の式	$$p(h_j^{(l)}\|\mathbf{h}^{(1-1)}) = \cdots$$	$$p(h_j^{(l)} = 1\|\mathbf{h}^{(l-1)}) = \cdots$$