深層学習の補足　　　　　　　　　　　　　　　　　　　　　担当：後藤 滋樹

微分の公式:

$$\frac{d}{dx}e^x = e^x \text{ (指数関数)}, \ \frac{d}{dx}\log(x) = \frac{1}{x} \text{ (対数関数)},$$
$$\frac{d}{dx}f(x)\cdot g(x) = \frac{f'(x)\cdot g(x) - f(x)\cdot g'(x)}{g^2(x)} \text{ (分数)},$$
$$\frac{d(f(g(x)))}{dx} = f'(g(x))\cdot g'(x) \text{ (合成関数)}.$$

準備として次の式を先に計算しておく。

$$\frac{\partial y_k}{\partial u_j^{(L)}} = \frac{\partial}{\partial u_j^{(L)}}\Big(\frac{\exp(u_k^{(L)})}{\sum_i \exp(u_i^{(L)})}\Big)$$ ここで $\exp(u_i^{(L)})$ は純粋に $u_i^{(L)}$ のみに依存する。

よって $\dfrac{\partial \exp(u_k^{(L)})}{\partial u_j^{(L)}}$ は $k=j$ のときは $\exp(u_k^{(L)}) = \exp(u_j^{(L)})$ となるが、$k \neq j$ のときには 0 である。

また $\dfrac{\partial}{\partial u_j^{(L)}}\Big(\sum_i \exp(u_i^{(L)})\Big) = \exp(u_j^{(L)})$ となり j 以外の項は残らない。

このことに注意をして分数の公式を適用する。$k=j$ のとき、分子の第 2 項では j 項だけが残る。

$$\frac{\partial y_k}{\partial u_j^{(L)}} = \frac{\partial}{\partial u_j^{(L)}}\Big(\frac{\exp(u_k^{(L)})}{\sum_i \exp(u_i^{(L)})}\Big) = \frac{\exp(u_j^{(L)})\sum_i \exp(u_i^{(L)}) - \exp(u_j^{(L)})\exp(u_j^{(L)})}{\Big(\sum_i \exp(u_i^{(L)})\Big)^2}$$

次に $k \neq j$ のときは上の分子の第 1 項が 0 となる。

$$\frac{\partial y_k}{\partial u_j^{(L)}} = \frac{\partial}{\partial u_j^{(L)}}\Big(\frac{\exp(u_k^{(L)})}{\sum_i \exp(u_i^{(L)})}\Big) = \frac{0 - \exp(u_k^{(L)})\exp(u_j^{(L)})}{\Big(\sum_i \exp(u_i^{(L)})\Big)^2}$$

テキストの p.49 の式に戻る。E_n の定義の式を用いて、対数の公式と y_k の定義により、

$$\delta_j^{(L)} = \frac{\partial E_n}{\partial u_j^{(L)}} = \frac{\partial}{\partial u_j^{(L)}}\Big(-\sum_k d_k \log y_k\Big) = -\sum_k d_k \frac{1}{y_k}\frac{\partial y_k}{\partial u_j^{(L)}} = -\sum_k d_k \Big(\frac{\sum_i \exp(u_i^{(L)})}{\exp(u_k^{(L)})}\Big)\frac{\partial y_k}{\partial u_j^{(L)}}$$

$j=k$ の場合と $j \neq k$ の場合に分けて計算する。まず $j=k$ の場合に上の準備を使う。

$$-d_j\Big(\frac{\sum_i \exp(u_i^{(L)})}{\exp(u_j^{(L)})}\Big)\frac{\partial y_k}{\partial u_j^{(L)}} = -d_j\Big(\frac{\sum_i \exp(u_i^{(L)})}{\exp(u_j^{(L)})}\Big)\frac{\exp(u_j^{(L)})\sum_i \exp(u_i^{(L)}) - \exp(u_j^{(L)})\exp(u_j^{(L)})}{\Big(\sum_i \exp(u_i^{(L)})\Big)^2}$$
$$= -d_j(1-y_j)$$

次に $j \neq k$ の場合も同様に

$$-d_k\Big(\frac{\sum_i \exp(u_i^{(L)})}{\exp(u_k^{(L)})}\Big)\frac{\partial y_k}{\partial u_j^{(L)}} = -d_j\Big(\frac{\sum_i \exp(u_i^{(L)})}{\exp(u_k^{(L)})}\Big)\frac{0 - \exp(u_k^{(L)})\exp(u_j^{(L)})}{\Big(\sum_i \exp(u_i^{(L)})\Big)^2} = -d_k(-y_j)$$

この後はテキストにあるように $\sum_k d_k = 1$ を用いて（ここでは計算の順序を少しだけ変更）

$$\delta_j^{(L)} = -d_j(1-y_j) - \sum_{k\neq j} d_k(-y_j) = -d_j + \sum_k d_k \cdot y_j = y_j - d_j$$