Projet Informatique Individuel
ENSC – Semestre 8

Perceptron multi-couches et réseau neuronal convolutif

Sommaire

---

1. Théorie de la rétro-propagation du gradient

1.1. Problème

Soit le perceptron multi-couches dont les couches sont liées entre elles par des relations de type :

$$f_B\left(W * B + \beta\right) = A \Leftrightarrow f_B\left(\begin{bmatrix} W_{1,1} & \cdots & W_{1,n_B} \\ \vdots & & \vdots \\ W_{n_A,1} & \cdots & W_{n_A,n_B} \end{bmatrix} * \begin{bmatrix} B_1 \\ \vdots \\ B_{n_B} \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{n_B} \end{bmatrix}\right) = \begin{bmatrix} A_1 \\ \vdots \\ A_{n_A} \end{bmatrix}$$

avec

- n_B le nombre d'entrées dans la couche B
- n_A le nombre de sorties dans la couche A
- f_B la fonction d'activation de la couche B ($f_B : \mathbb{R} \to [-1, 1]$)
- $B \in \mathcal{M}_{n_B,1}$ la matrice contenant les valeurs d'entrée
- $\{B_b, \forall\ b \in [\![1, n_B]\!]\}$ les valeurs des entrées
- $A \in \mathcal{M}_{n_A,1}$ la matrice contenant les valeurs de sortie
- $\left\{A_j, \forall\ a \in [\![1, n_A]\!]\right\}$ les valeurs des sorties
- $W \in \mathcal{M}_{n_A,n_B}$ la matrice contenant les poids
- $\{W_{a,b}, \forall\ (a,b) \in [\![1, n_A]\!] \times [\![1, n_B]\!]\}$ les poids entre les B_b et les A_a
- $\beta \in \mathcal{M}_{n_B,1}$ la matrice contenant les biais
- $\{\beta_b, \forall\ b \in [\![1, n_B]\!]\}$ les valeurs des biais

Dans le reste du document, $*$ désignera la multiplication matricielle et $\times$ la multiplication terme à terme.

On notera ${}^{\mathbf{t}}M$ la transposée de la matrice M.

1.2. Correction des poids de la dernière couche

Supposons que pour tout $a \in [\![1, n_A]\!]$ la sortie théorique soit α_a et non A_a.

L'erreur quadratique $\mathcal{E}$ commise sur un tel réseau est : $\mathcal{E} = \frac{1}{2} \sum\limits_{\ell=1}^{n_A} (\alpha_\ell - A_\ell)^2$.

On s'intéresse aux variations qu'apporteraient les modifications de $W_{a,b}$ ($(a,b) \in [\![1, n_A]\!] \times [\![1, n_B]\!]$) sur l'erreur quadratique.

Il s'agit donc de faire varier les $W_{a,b}$ pour chaque couple $(a,b) \in [\![1, n_A]\!] \times [\![1, n_B]\!]$ de telle façon à ce que réduire au maximum $\mathcal{E}$.

On va donc calculer pour tout $(a,b) \in [\![1, n_A]\!] \times [\![1, n_B]\!]$ les dérivées partielles $\varepsilon_{a,b}$ de $\mathcal{E}$ par rapport aux $W_{a,b}$.

Pour un a et un b donnés :

$$
\begin{aligned}
\varepsilon_{a,b} &= \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} \mathcal{E} \\
&= \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} \frac{1}{2} \sum_{\ell=1}^{n_A} (\alpha_\ell - A_\ell)^2 \\
&= \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} \frac{1}{2} \sum_{\ell=1}^{n_A} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} B_k + \beta_k \right) \right)^2 \\
&= \frac{1}{2} \sum_{\ell=1}^{n_A} \left[\frac{\mathrm{d}}{\mathrm{d}W_{a,b}} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} B_k + \beta_k \right) \right)^2 \right] \\
&= \sum_{\ell=1}^{n_A} \left[\left(\alpha_\ell - f_B \left(\sum_{j=1}^{n_B} W_{\ell,j} B_j + \beta_j \right) \right) \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} B_k + \beta_k \right) \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[\left(\alpha_\ell - f_B \left(\sum_{j=1}^{n_B} W_{\ell,j} B_j + \beta_j \right) \right) \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} B_k + \beta_k \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[\left(\alpha_\ell - f_B \left(\sum_{j=1}^{n_B} W_{\ell,j} B_j + \beta_j \right) \right) f'_B \left(\sum_{i=1}^{n_B} W_{\ell,i} B_i + \beta_i \right) \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} \sum_{k=1}^{n_B} W_{\ell,k} B_k + \beta_k \right] \\
&= -\sum_{\ell=1}^{n_A} \left[\left(\alpha_\ell - f_B \left(\sum_{j=1}^{n_B} W_{\ell,j} B_j + \beta_j \right) \right) f'_B \left(\sum_{i=1}^{n_B} W_{\ell,i} B_i + \beta_i \right) \sum_{k=1}^{n_B} \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} W_{\ell,k} B_k + \beta_k \right] \\
&= -\left(\alpha_a - f_B \left(\sum_{j=1}^{n_B} W_{a,j} B_j + \beta_j \right) \right) f'_B \left(\sum_{i=1}^{n_B} W_{a,i} B_i + \beta_i \right) \frac{\mathrm{d}}{\mathrm{d}W_{a,b}} W_{a,b} B_b + \beta_b \\
&= -\left(\alpha_a - f_B \left(\sum_{j=1}^{n_B} W_{a,j} B_j + \beta_j \right) \right) f'_B \left(\sum_{i=1}^{n_B} W_{a,i} B_i + \beta_i \right) B_b \\
&= -(\alpha_a - A_a)\, f'_B \left(\sum_{i=1}^{n_B} W_{a,i} B_i + \beta_i \right) B_b
\end{aligned}
$$

En notant $\varepsilon_a = \begin{bmatrix} \varepsilon_{a,1} & \cdots & \varepsilon_{a,n_B} \end{bmatrix} \in \mathcal{M}_{1,n_B}$, on a donc :

$$
\varepsilon_a = -(\alpha_a - A_a)\, f'_B \left(\sum_{i=1}^{n_B} W_{a,i} B_i + \beta_i \right) \begin{bmatrix} B_1 & \cdots & B_{n_B} \end{bmatrix} = -(\alpha_a - A_a)\, f'_B \left(\sum_{i=1}^{n_B} W_{a,i} B_i + \beta_i \right) {}^{\mathbf{t}}B
$$

Puis en notant $\varepsilon = \begin{bmatrix} \varepsilon_{1,1} & \cdots & \varepsilon_{1,n_B} \\ \vdots & & \vdots \\ \varepsilon_{n_A,1} & \cdots & \varepsilon_{n_A,n_B} \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n_A} \end{bmatrix} \in \mathcal{M}_{n_A,n_B}$, on en déduit que :

$$
\begin{aligned}
\varepsilon &= \begin{bmatrix} -(\alpha_1 - A_1)\, f_B'\left(\sum\limits_{i=1}^{n_B} W_{1,i}B_i + \beta_i\right) {}^{\mathbf{t}}B \\ \vdots \\ -(\alpha_{n_A} - A_{n_A})\, f_B'\left(\sum\limits_{i=1}^{n_B} W_{n_A,i}B_i + \beta_i\right) {}^{\mathbf{t}}B \end{bmatrix} \\
&= -\begin{bmatrix} (\alpha_1 - A_1)\, f_B'\left(\sum\limits_{i=1}^{n_B} W_{1,i}B_i + \beta_i\right) \\ \vdots \\ (\alpha_{n_A} - A_{n_A})\, f_B'\left(\sum\limits_{i=1}^{n_B} W_{n_A,i}B_i + \beta_i\right) \end{bmatrix} * {}^{\mathbf{t}}B \\
&= -\left(\begin{bmatrix} (\alpha_1 - A_1) \\ \vdots \\ (\alpha_{n_A} - A_{n_A}) \end{bmatrix} \times \begin{bmatrix} f_B'\left(\sum\limits_{i=1}^{n_B} W_{1,i}B_i + \beta_i\right) \\ \vdots \\ f_B'\left(\sum\limits_{i=1}^{n_B} W_{n_A,1}B_i + \beta_i\right) \end{bmatrix}\right) * {}^{\mathbf{t}}B \\
&= -\left[\left(\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{n_A} \end{bmatrix} - \begin{bmatrix} A_1 \\ \vdots \\ A_{n_A} \end{bmatrix}\right) \times f_B'\left(\begin{bmatrix} \sum\limits_{i=1}^{n_B} W_{1,i}B_i + \beta_i \\ \vdots \\ \sum\limits_{i=1}^{n_B} W_{n_A,i}B_i + \beta_i \end{bmatrix}\right)\right] * {}^{\mathbf{t}}B \\
&= -\left[(\alpha - A) \times f_B'(W * B + \beta)\right] * {}^{\mathbf{t}}B
\end{aligned}
$$

On posant $\delta_B = \left[(\alpha - A) \times f_B'(W * B + \beta)\right]$, on a $\varepsilon = -\delta_B * {}^{\mathbf{t}}B$.

1.3. Correction des biais de la dernière couche

Pour la correction des biais, c'est exactement la même démarche, sauf que cette fois on dérive par rapport à β_b et non plus par rapport à $W_{a,b}$.
En effectuant, les mêmes calculs, on trouve que :

$$\epsilon_b = \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\beta_b} = -(\alpha_a - A_a)\, f_B'\left(\sum_{i=1}^{n_B} W_{a,i}B_i + \beta_i\right)$$

Puis, en posant $\epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{n_B} \end{bmatrix}$, on trouve que $\epsilon = -\left[(\alpha - A) \times f_B'(W * B + \beta)\right] = -\delta_B$

1.4. Correction des poids de l'avant-dernière couche

Maintenant qu'on a calculé les variations à appliquer sur les poids entre la dernière et l'avant-dernière couche, il s'agit désormais de faire de même pour les poids entre les couches précédentes.

L'erreur quadratique $\mathcal{E}$ commise est toujours : $\mathcal{E} = \frac{1}{2} \sum_{\ell=1}^{n_A} (\alpha_\ell - A_\ell)^2$.

On s'intéresse maintenant aux variations qu'apporteraient les modifications de $W'_{b,c}$ ($(b,c) \in [\![1,n_B]\!] \times [\![1,n_C]\!]$) sur l'erreur quadratique.

Il s'agit donc de faire varier les $W'_{b,c}$ pour chaque couple $(b,c) \in [\![1,n_B]\!] \times [\![1,n_C]\!]$ de telle façon à ce que réduire au maximum $\mathcal{E}$.

On va donc calculer pour tout $(b,c) \in [\![1,n_B]\!] \times [\![1,n_C]\!]$ les dérivées partielles $\varepsilon'_{b,c}$ de $\mathcal{E}$ par rapport aux $W'_{b,c}$.

Pour un b et un c donnés :

$$
\begin{aligned}
\varepsilon'_{b,c} &= \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \mathcal{E} \\
&= \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \frac{1}{2} \sum_{\ell=1}^{n_A} (\alpha_\ell - A_\ell)^2 \\
&= \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \frac{1}{2} \sum_{\ell=1}^{n_A} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} B_k + \beta_k \right) \right)^2 \\
&= \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \frac{1}{2} \sum_{\ell=1}^{n_A} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right)^2 \\
&= \frac{1}{2} \sum_{\ell=1}^{n_A} \left[\frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right)^2 \right] \\
&= \sum_{\ell=1}^{n_A} \left[\left(\alpha_\ell - f_B \left(\sum_{i=1}^{n_B} W_{\ell,i} f_C \left(\sum_{h=1}^{n_C} W'_{i,h} C_h + \beta'_h \right) + \beta_i \right) \right) \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right) \right] \\
&= \sum_{\ell=1}^{n_A} \left[(\alpha_\ell - A_\ell) \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \left(\alpha_\ell - f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[(\alpha_\ell - A_\ell) \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} f_B \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[(\alpha_\ell - A_\ell) f'_B \left(\sum_{i=1}^{n_B} W_{\ell,i} f_C \left(\sum_{h=1}^{n_C} W'_{i,h} C_h + \beta'_h \right) + \beta_i \right) \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[(\alpha_\ell - A_\ell) f'_B \left(\sum_{i=1}^{n_B} W_{\ell,i} B_i + \beta_i \right) \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \left(\sum_{k=1}^{n_B} W_{\ell,k} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) + \beta_k \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[(\alpha_\ell - A_\ell) f'_B \left(\sum_{i=1}^{n_B} W_{\ell,i} B_i + \beta_i \right) \sum_{k=1}^{n_B} \left(W_{\ell,k} \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} f_C \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) \right) \right] \\
&= -\sum_{\ell=1}^{n_A} \left[(\alpha_\ell - A_\ell) f'_B \left(\sum_{i=1}^{n_B} W_{\ell,i} B_i + \beta_i \right) \sum_{k=1}^{n_B} \left(W_{\ell,k}\, f'_C \left(\sum_{h=1}^{n_C} W'_{k,h} C_h + \beta'_h \right) \frac{\mathrm{d}}{\mathrm{d}W'_{b,c}} \left(\sum_{j=1}^{n_C} W'_{k,j} C_j + \beta'_j \right) \right) \right]
\end{aligned}
$$

$$
\begin{aligned}
\varepsilon'_{b,c} &= -\sum_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,b}\ f'_C\left(\sum_{h=1}^{n_C} W'_{b,h}C_h + \beta'_h\right)\frac{\mathrm{d}}{\mathrm{d}W'_{b,c}}\left(W'_{b,c}C_c + \beta'_c\right)\right] \\
&= -\sum_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,b}\ f'_C\left(\sum_{h=1}^{n_C} W'_{b,h}C_h + \beta'_h\right) C_c\right] \\
&= -\sum_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,b}\right] f'_C\left(\sum_{h=1}^{n_C} W'_{b,h}C_h + \beta'_h\right) C_c
\end{aligned}
$$

En notant $\varepsilon'_b = \begin{bmatrix}\varepsilon_{b,1} & \cdots & \varepsilon_{b,n_C}\end{bmatrix} \in \mathcal{M}_{1,n_C}$, on a donc :

$$
\begin{aligned}
\varepsilon'_b &= -\sum_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,b}\right] f'_C\left(\sum_{h=1}^{n_C} W'_{b,h}C_h + \beta'_h\right)\begin{bmatrix}C_1 & \cdots & C_{n_C}\end{bmatrix} \\
&= -\sum_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,b}\right] f'_C\left(\sum_{h=1}^{n_C} W'_{b,h}C_h + \beta'_h\right) {}^{\mathbf{t}}C
\end{aligned}
$$

Puis en notant $\varepsilon' = \begin{bmatrix}\varepsilon'_{1,1} & \cdots & \varepsilon'_{1,n_C} \\ \vdots & & \vdots \\ \varepsilon'_{n_B,1} & \cdots & \varepsilon'_{n_B,n_C}\end{bmatrix} = \begin{bmatrix}\varepsilon'_1 \\ \vdots \\ \varepsilon'_b\end{bmatrix} \in \mathcal{M}_{n_B,n_C}$, on en déduit que :

$$
\begin{aligned}
\varepsilon' &= \begin{bmatrix}
-\sum\limits_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,1}\right] f'_C\left(\sum\limits_{h=1}^{n_C} W'_{1,h}C_h + \beta'_h\right) {}^{\mathbf{t}}C \\
\vdots \\
-\sum\limits_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,n_B}\right] f'_C\left(\sum\limits_{h=1}^{n_C} W'_{n_B,h}C_h + \beta'_h\right) {}^{\mathbf{t}}C
\end{bmatrix} \\
&= -\begin{bmatrix}
\sum\limits_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,1}\right] f'_C\left(\sum\limits_{h=1}^{n_C} W'_{1,h}C_h + \beta'_h\right) \\
\vdots \\
\sum\limits_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,n_B}\right] f'_C\left(\sum\limits_{h=1}^{n_C} W'_{n_B,h}C_h + \beta'_h\right)
\end{bmatrix} * {}^{\mathbf{t}}C \\
&= -\left(\begin{bmatrix}
\sum\limits_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,1}\right] \\
\vdots \\
\sum\limits_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,n_B}\right]
\end{bmatrix} \times \begin{bmatrix}
f'_C\left(\sum\limits_{h=1}^{n_C} W'_{1,h}C_h + \beta'_h\right) \\
\vdots \\
f'_C\left(\sum\limits_{h=1}^{n_C} W'_{n_B,h}C_h + \beta'_h\right)
\end{bmatrix}\right) * {}^{\mathbf{t}}C \\
&= -\left[\left(\begin{bmatrix}
W_{1,1} & \cdots & W_{n_A,1} \\
\vdots & & \vdots \\
W_{1,n_B} & \cdots & W_{n_A,n_B}
\end{bmatrix} * \begin{bmatrix}
(\alpha_1 - A_1)\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{1,i}B_i + \beta_i\right) \\
\vdots \\
(\alpha_{n_A} - A_{n_A})\, f'_B\left(\sum\limits_{i=1}^{n_B} W_{n_A,i}B_i + \beta_i\right)
\end{bmatrix}\right) \times f'_C\left(\begin{bmatrix}
\sum\limits_{h=1}^{n_C} W'_{1,h}C_h + \beta'_h \\
\vdots \\
\sum\limits_{h=1}^{n_C} W'_{n_B,h}C_h + \beta'_h
\end{bmatrix}\right)\right] * {}^{\mathbf{t}}C \\
&= -\left[\left({}^{\mathbf{t}}W * \left[(\alpha - A) \times f'_B\left(W * B + \beta\right)\right]\right) \times f'_C\left(W' * C + \beta'\right)\right] * {}^{\mathbf{t}}C \\
&= -\left[\left({}^{\mathbf{t}}W * \delta_B\right) \times f'_C\left(W' * C + \beta'\right)\right] * {}^{\mathbf{t}}C
\end{aligned}
$$

On posant $\delta_C = \left[\left({}^{\mathbf{t}}W * \delta_B\right) \times f'_C\left(W' * C + \beta'\right)\right]$, on a $\varepsilon' = -\delta_C * {}^{\mathbf{t}}C$.

1.5. Correction des biais de l'avant-dernière couche

Pour la correction des biais, c'est exactement la même démarche, sauf que cette fois on dérive par rapport à β'_c et non plus par rapport à $W'_{b,c}$.
En effectuant, les mêmes calculs, on trouve que :

$$\epsilon_c = \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\beta'_c} = -\sum_{\ell=1}^{n_A}\left[(\alpha_\ell - A_\ell)\, f'_B\left(\sum_{i=1}^{n_B} W_{\ell,i}B_i + \beta_i\right) W_{\ell,b}\right] f'_C\left(\sum_{h=1}^{n_C} W'_{b,h}C_h + \beta'_h\right)$$

Puis, en posant $\epsilon' = \begin{bmatrix}\epsilon'_1\\ \vdots \\ \epsilon'_{n_C}\end{bmatrix}$, on trouve que $\epsilon' = -\left[\left({}^{\mathbf{t}}W * \delta_B\right) \times f'_C\left(W' * C + \beta'\right)\right] = -\delta_C$

1.6. Correction des poids et des biais des précédentes couches

Pour une couche D ayant n_D neurones, la fonction d'activation f_D et les poids $W''_{c,d}$ $((c,d) \in [\![1, n_C]\!] \times [\![1, n_D]\!])$ entre les couches C et D, on peut montrer que :

$$\varepsilon'' = -\left[\left({}^{\mathbf{t}}W' * \delta_C\right) \times f'_D\left(W'' * D + \beta''\right)\right] * {}^{\mathbf{t}}D$$

Et que :

$$\epsilon'' = -\left[\left({}^{\mathbf{t}}W' * \delta_C\right) \times f'_D\left(W'' * D + \beta''\right)\right]$$

On posant $\delta_D = \left[\left({}^{\mathbf{t}}W' * \delta_C\right) \times f'_D\left(W'' * D + \beta''\right)\right]$, on a $\varepsilon'' = -\delta_D * {}^{\mathbf{t}}D$ et $\epsilon'' = -\delta_D$.
De même, pour une couche E ayant n_E neurones, la fonction d'activation f_E et les poids $W'''_{d,e}$ $((d,e) \in [\![1, n_D]\!] \times [\![1, n_E]\!])$ entre les couches D et E, on peut montrer que :

$$\varepsilon''' = -\left[\left({}^{\mathbf{t}}W'' * \delta_D\right) \times f'_E\left(W''' * E + \beta'''\right)\right] * {}^{\mathbf{t}}E$$

Et que :

$$\epsilon''' = -\left[\left({}^{\mathbf{t}}W'' * \delta_D\right) \times f'_E\left(W''' * E + \beta'''\right)\right]$$

On posant $\delta_E = \left[\left({}^{\mathbf{t}}W'' * \delta_D\right) \times f'_E\left(W''' * E + \beta'''\right)\right]$, on a $\varepsilon''' = -\delta_E * {}^{\mathbf{t}}E$ et $\epsilon''' = -\delta_E$.

1.7. Modification des poids et des biais

Une fois qu'on a calculé les variations à appliquer à la matrice des poids synaptiques et à la matrice des biais, il ne reste plus qu'à effectuer les calculs suivants :

$$W = W - \mu\varepsilon \text{ et } \beta = \beta - \mu\epsilon$$

où μ est le coefficient d'apprentissage, W la matrice des poids synaptiques que l'on souhaite corriger, ε la matrice associée à W et contenant les variations à effectuer sur chaque poids synaptique de W, β la matrice des biais que l'on souhaite corriger et ϵ la matrice associée à β et contenant les variations à effectuer sur chaque biais de β.