Introduction aux réseaux de neurones
Notes additionnelles

Les batch seront stockés dans des matrices avec 1 exemple par ligne. Par exemple pour l'entrée on a une matrice $\mathbf{X}$ de taille $N \times n_x$ pour N exemples chacun de dimension n_x. Chaque exemple $\mathbf{x}$ sera donc un vecteur ligne (dimension $1 \times n_x$). On fera de même pour les résultats intermédiaires. Ainsi, la matrice de poids $\mathbf{W}^h$ sera par exemple de taille $n_h \times n_x$. Le vecteur de biais $\mathbf{b}^h$ sera un vecteur ligne (dimension $1 \times n_h$).

$\odot$ designe le produit terme-à-terme $\mathrm{sum}_{\mathsf{line}}(\mathbf{X})$ avec X de taille $N \times n_x$ fait une somme par ligne et retourne un vecteur colonne de taille N. $\mathrm{repmat}_{N\ \mathsf{lines}}(\mathbf{b})$ répète le vecteur-ligne $\mathbf{b}$ (dimension $1 \times p$) N fois en ligne pour produire une matrice de taille $N \times p$.

Forward

Elementwise

$$
\begin{cases}
\tilde{h}_i = \sum_{j=1}^{n_x} W_{i,j}^h\ x_j + b_i^h \\[2mm]
h_i = \tanh(\tilde{h}_i) \\[2mm]
\tilde{y}_i = \sum_{j=1}^{n_h} W_{i,j}^y\ h_j + b_i^y \\[2mm]
\hat{y}_i = \mathrm{SoftMax}(\tilde{y}_i) = \dfrac{e^{\tilde{y}_i}}{\sum_{j=1}^{n_y} e^{\tilde{y}_j}}
\end{cases}
$$

Vectoriel

$$
\begin{cases}
\tilde{\mathbf{h}} = \mathbf{x}\mathbf{W}^{h\top} + \mathbf{b}^h \\[2mm]
\mathbf{h} = \tanh(\tilde{\mathbf{h}}) \\[2mm]
\tilde{\mathbf{y}} = \mathbf{h}\mathbf{W}^{y\top} + \mathbf{b}^y \\[2mm]
\hat{\mathbf{y}} = \mathrm{SoftMax}(\tilde{\mathbf{y}})
\end{cases}
$$

Vectoriel par batch

$$
\begin{cases}
\tilde{\mathbf{H}} = \mathbf{X}\mathbf{W}^{h\top} + \mathrm{repmat}_{N\ \mathsf{lines}}(\mathbf{b}_h) \\[2mm]
\mathbf{H} = \tanh(\tilde{\mathbf{H}}) \\[2mm]
\tilde{\mathbf{Y}} = \mathbf{H}\mathbf{W}^{y\top} + \mathrm{repmat}_{N\ \mathsf{lines}}(\mathbf{b}_y) \\[2mm]
\hat{\mathbf{Y}} = \mathrm{SoftMax}_{\mathsf{line}}(\tilde{\mathbf{Y}})
\end{cases}
$$

Loss

$$
\begin{cases}
\ell(\mathbf{y}, \tilde{\mathbf{y}}) = -\sum_{i=1}^{n_y} y_i \log \hat{y}_i = -\sum_{i=1}^{n_y} y_i \tilde{y}_i + \log \sum_{j=1}^{n_y} e^{\tilde{y}_j} \\[3mm]
\mathcal{L}(\mathbf{Y}, \hat{\mathbf{Y}}) = -\dfrac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{n_y} Y_{k,i} \log \hat{Y}_{k,i} = -\mathrm{mean}_{\mathsf{col}}(\mathrm{sum}_{\mathsf{line}}(\mathbf{Y} \odot \log \hat{\mathbf{Y}}))
\end{cases}
$$

Backward

Elementwise

$$
\begin{cases}
\delta_i^y = \dfrac{\partial \ell}{\partial \tilde{y}_i} = \hat{y}_i - y_i \\[3mm]
\dfrac{\partial \ell}{\partial W_{i,j}^y} = \delta_i^y\ h_j \\[3mm]
\dfrac{\partial \ell}{\partial b_i^y} = \delta_i^y \\[3mm]
\delta_i^h = \dfrac{\partial \ell}{\partial \tilde{h}_i} = (1 - h_i^2) \sum_{j=1}^{n_y} \delta_j^y W_{j,i}^y \\[3mm]
\dfrac{\partial \ell}{\partial W_{i,j}^h} = \delta_i^h\ x_j \\[3mm]
\dfrac{\partial \ell}{\partial b_i^h} = \delta_i^h
\end{cases}
$$

Vectoriel

$$
\begin{cases}
\nabla_{\tilde{\mathbf{y}}} = \hat{\mathbf{y}} - \mathbf{y} \\[2mm]
\nabla_{\mathbf{W}^y} = \nabla_{\tilde{\mathbf{y}}}{}^\top \mathbf{h} \\[2mm]
\nabla_{\mathbf{b}^y} = \nabla_{\tilde{\mathbf{y}}}^\top \\[2mm]
\nabla_{\tilde{\mathbf{h}}} = (\nabla_{\tilde{\mathbf{y}}}\mathbf{W}^y) \odot (1 - \mathbf{h}^2) \\[2mm]
\nabla_{\mathbf{W}^h} = \nabla_{\tilde{\mathbf{h}}}{}^\top \mathbf{x} \\[2mm]
\nabla_{\mathbf{b}^h} = \nabla_{\tilde{\mathbf{h}}}^\top
\end{cases}
$$

Vectoriel par batch

$$
\begin{cases}
\nabla_{\tilde{\mathbf{Y}}} = \hat{\mathbf{Y}} - \mathbf{Y} \\[2mm]
\nabla_{\mathbf{W}^y} = \nabla_{\tilde{\mathbf{Y}}}{}^\top \mathbf{H} \\[2mm]
\nabla_{\mathbf{b}^y} = \mathrm{sum}_{\mathsf{ligne}}(\nabla_{\tilde{\mathbf{Y}}})^\top \\[2mm]
\nabla_{\tilde{\mathbf{H}}} = (\nabla_{\tilde{\mathbf{Y}}}\mathbf{W}^y) \odot (1 - \mathbf{H}^2) \\[2mm]
\nabla_{\mathbf{W}^h} = \nabla_{\tilde{\mathbf{H}}}{}^\top \mathbf{X} \\[2mm]
\nabla_{\mathbf{b}^h} = \mathrm{sum}_{\mathsf{ligne}}(\nabla_{\tilde{\mathbf{H}}})^\top
\end{cases}
$$