

# RDFIA TME 4-5

## 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  $X$  de taille  $N \times n_x$  pour  $N$  exemples chacun de dimension  $n_x$ . On fera de même pour les résultats intermédiaires. Ainsi, la matrice de poids  $W_h$  sera par exemple de taille  $n_h \times n_x$ .

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

### Forward

Elementwise

Vectoriel

Vectoriel par batch

$$\left\{ \begin{array}{l} \tilde{h}_i = \sum_j W_{h,ij} x_j + b_{h,i} \\ h_i = \tanh(\tilde{h}_i) \\ \tilde{y}_i = \sum_j W_{y,ij} h_j + b_{y,i} \\ \hat{y}_i = \text{SoftMax}(\tilde{y}_i) = \frac{e^{\tilde{y}_i}}{\sum_j e^{\tilde{y}_j}} \end{array} \right. \quad \left\{ \begin{array}{l} \tilde{h} = W_h x + b_h \\ h = \tanh(\tilde{h}) \\ \tilde{y} = W_y h + b_y \\ \hat{y} = \text{SoftMax}(\tilde{y}) \end{array} \right. \quad \left\{ \begin{array}{l} \tilde{H} = X W_h^\top + \text{repmat}_{N \text{ lines}}(b_h^\top) \\ H = \tanh(\tilde{H}) \\ \tilde{Y} = H W_y^\top + \text{repmat}_{N \text{ lines}}(b_y^\top) \\ \hat{Y} = \text{SoftMax}_{\text{line}}(\tilde{Y}) \end{array} \right.$$

### Loss

$$\left\{ \begin{array}{l} \ell(y, \tilde{y}) = - \sum_i y_i \log \hat{y}_i = - \sum_i y_i \tilde{y}_i + \log \sum_j e^{\tilde{y}_j} \\ \mathcal{L}(Y, \hat{Y}) = - \frac{1}{N} \sum_k \sum_i Y_{k,i} \log \hat{Y}_{k,i} = - \text{mean}_{\text{col}}(\text{sum}_{\text{line}}(Y \odot \log \hat{Y})) \end{array} \right.$$

### Backward

Elementwise

Vectoriel

Vectoriel par batch

$$\left\{ \begin{array}{l} \delta_{y,i} = \frac{\partial \ell}{\partial \tilde{y}_i} = \hat{y}_i - y_i \\ \frac{\partial \ell}{\partial W_{y,ij}} = \delta_{y,i} h_j \\ \frac{\partial \ell}{\partial b_{y,i}} = \delta_{y,i} \\ \delta_{h,i} = \frac{\partial \ell}{\partial \tilde{h}_i} = (1 - h_i^2) \sum_j \delta_{y,j} W_{y,ji} \\ \frac{\partial \ell}{\partial W_{h,ij}} = \delta_{h,i} x_j \\ \frac{\partial \ell}{\partial b_{h,i}} = \delta_i^h \end{array} \right. \quad \left\{ \begin{array}{l} \nabla_{\tilde{y}} = \hat{y} - y \\ \nabla_{W_y} = \nabla_{\tilde{y}} h^\top \\ \nabla_{b_y} = \nabla_{\tilde{y}} \\ \nabla_{\tilde{h}} = W_y^\top \nabla_{\tilde{y}} \odot (1 - h^2) \\ \nabla_{W_h} = \nabla_{\tilde{h}} x^\top \\ \nabla_{b_h} = \nabla_{\tilde{h}} \end{array} \right. \quad \left\{ \begin{array}{l} \nabla_{\tilde{Y}} = \hat{Y} - Y \\ \nabla_{W_y} = \nabla_{\tilde{Y}}^\top H \\ \nabla_{b_y} = \text{sum}_{\text{col}}(\nabla_{\tilde{Y}})^\top \\ \nabla_{\tilde{H}} = \nabla_{\tilde{Y}} W_y \odot (1 - H^2) \\ \nabla_{W_h} = \nabla_{\tilde{H}}^\top X \\ \nabla_{b_h} = \text{sum}_{\text{col}}(\nabla_{\tilde{H}})^\top \end{array} \right.$$