Polytech - Deep Learning - TP 3-4

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 $n_x \times n_h$. Le vecteur de bias $\mathbf{b}^h$ 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

$$\begin{cases} \tilde{h}_i = \sum_{j=1}^{n_x} W^h_{j,i}\ x_j + b^h_i \\ h_i = \tanh(\tilde{h}_i) \\ \tilde{y}_i = \sum_{j=1}^{n_h} W^y_{j,i}\ h_j + b^y_i \\ \hat{y}_i = \mathrm{SoftMax}(\tilde{y}_i) = \dfrac{e^{\tilde{y}_i}}{\sum_j e^{\tilde{y}_j}} \end{cases}$$

Vectoriel

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

Vectoriel par batch

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

Loss

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

Backward

Elementwise

$$\begin{cases} \delta^y_i = \dfrac{\partial \ell}{\partial \tilde{y}_i} = \hat{y}_i - y_i \\ \dfrac{\partial \ell}{\partial W^y_{i,j}} = \delta_{y,i}\ h_j \\ \dfrac{\partial \ell}{\partial b^y_i} = \delta^y_i \\ \delta^h_i = \dfrac{\partial \ell}{\partial \tilde{h}_i} = (1 - h_i^2)\sum_{j=1}^{n_y} \delta^y_j W^y_{i,j} \\ \dfrac{\partial \ell}{\partial W_{h,ij}} = \delta^h_i\ x_j \\ \dfrac{\partial \ell}{\partial b_{h,i}} = \delta^h_i \end{cases}$$

Vectoriel

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

Vectoriel par batch

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