

exo_retroprop

July 24, 2019

1 exo_retroprop.ipynb

Solutions de l'exercice 10.134 du livre **Python. Apprentissage Actif** (JP. Roy, Editions Ellipses, 2019). Le pdf de ce *notebook Jupyter* se trouve dans `exo_retroprop.ipynb`. Pour le rendu mathématique, j'utilise L^AT_EX qui est installé sur ma machine et reconnu par Jupyter.

1.1 Les équations de la rétro-propagation

```
[1]: def sig(x) : return 1/(1+ exp(-x))      # la fonction sigmoïde
```

Question b_1 . La dérivée de la fonction `sig` se calcule facilement à la main. Mais avec `sympy`, on peut vérifier symboliquement l'identité demandée $\sigma'(x) = \sigma(x)(1 - \sigma(x))$

```
[2]: from sympy import *
var('x')
if simplify(diff(sig(x),x)-sig(x)*(1-sig(x))) == 0:
    print('Egalité correcte')
del x
```

Egalité correcte

On rappelle les notations du réseau de neurones utilisées dans le livre.

$$\tilde{z}_k = \sum_j w_{j,k} z_j \text{ et } z_k = \sigma(\tilde{z}_k).$$

Question b_2 . Montrons que $\partial z_k / \partial \tilde{z}_k = z_k (1 - z_k)$ et que $\partial \tilde{z}_k / \partial z_j = w_{jk}$.

Puisque z_k est fonction par σ de \tilde{z}_k , il vient :

$$\partial z_k / \partial \tilde{z}_k = \sigma'(\tilde{z}_k) = \sigma(\tilde{z}_k)(1 - \sigma(\tilde{z}_k)). \text{ D'où :}$$

$$\partial z_k / \partial \tilde{z}_k = z_k(1 - z_k)$$

Et dans la somme $\tilde{z}_k = \sum_j w_{j,k} z_j$, la variable z_j n'apparaît qu'une fois, avec une contribution de w_{jk} . D'où :

$$\partial \tilde{z}_k / \partial z_j = w_{jk}$$

Question b_3 . Montrons que $\partial E / \partial z_k = y_k - z_k$.

On utilise l'erreur quadratique $E = \frac{1}{2} \sum_k (y_k - z_k)^2$. En ne conservant que la contribution de z_k dans cette somme, il vient immédiatement :

$$\partial E / \partial z_k = y_k - z_k$$

Question b_4 . Nous avons dit que nous nous déplaçons dans l'espace des poids dans le sens inverse du gradient, ce qui revient à utiliser une formule de modification des poids du type $\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}}$. Décomposons-le en $\Delta w_{jk} = -\eta \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial \tilde{z}_k} \frac{\partial \tilde{z}_k}{\partial w_{jk}}$.

Mais $\frac{\partial z_k}{\partial \bar{z}_k} = z_k(1 - z_k)$ et $\frac{\partial \bar{z}_k}{\partial w_{jk}} = z_j$ par la question b_2), d'où :

$$\Delta w_{jk} = -\eta z_j z_k (1 - z_k) \frac{\partial E}{\partial z_k}.$$