Diverses fonctions utiles

Pascal Vincent

November 18, 2015

1 Exponentielle et Logarithme

Exponentielle:
$\exp(x) = e^x$
$\mathbb{R} \to \mathbb{R}^{+*}$
Logarithme en base naturelle (base e)
$\log(x) = \log_e(x) = \ln(x)$
$\mathbb{R}^{+*} \to \mathbb{R}$

Ce sont des inverses l'une de l'autre:

$$\log(\exp(x)) = x$$

$$\exp(\log(x)) = x$$

Utilité:

- Souvent rencontrées en apprentissage automatique, quand on manipule des probabilités ou des densités de probabilité.
- SOUVENT, on va manipuler des log de (densité de) probabilité plutôt que des (densité de) probabilités
- POURQUOI? Pour des raisons de stabilité et précision numérique. Car souvent, surtout en haute dimension, les probabilités sont très très petites.
- EX: on va souvent manipuler la log vraisemblance (somme de log prob) plutôt que la vraisemblance (produit de prob)
- AUTRE RAISON: l'expression du gradient de la log vraisemblance est plus facile (la dérivée d'une somme est la somme des dérivées) que de la vraisemblance, et admet les mêmes minima.

Propriétés utiles:

- $\exp(0) = 1$
- $\exp(a+b) = \exp(a)\exp(b)$
- $\exp(ab) =$ pas de simplification
- $\exp(-a) = \frac{1}{\exp(a)}$
- $\exp(a-b) = \frac{\exp(a)}{\exp(b)}$
- $\log(1) = 0$
- $\log(ab) = \log(a) + \log(b)$
- Ex: $\log\left(\prod_k a_k\right) = \sum_k \log(a_k)$
- $\log(a+b) =$ pas de simplification
- $\log(\frac{1}{a}) = -\log a$
- $\log \frac{a}{b} = \log(a) - \log(b)$

2 Logadd

Comme on conserve souvent des probabilités sous forme de leur log, et que parfois on est appelé à sommer des probabilité, comment faire pour obtenir le résultat sous forme de log.

Soit $l_a = \log(p_a)$ et $l_b = \log(p_b)$

Produit des probabilités: facile: somme des log prob.

$\log(p_a p_b) = l_a + l_b$

Somme des probabilités: **logadd**

$\log(p_a + p_b) = \text{logadd}(l_a, l_b) = \log\left(\exp(l_a) + \exp(l_b)\right)$

Attention pour la stabilité numérique:

Pour calculer $\text{logadd}(l_a, l_b)$ de manière numériquement stable:

- $m = \max(l_a, l_b)$
-

$$
\begin{aligned}
\text{logadd}(l_a, l_b) &= \log\left(\exp(l_a) + \exp(l_b)\right) \\
&= \log\left(\frac{\exp(m)}{\exp(m)}\left(\exp(l_a) + \exp(l_b)\right)\right) \\
&= \log\left(\exp(m)\left(\exp(l_a - m) + \exp(l_b - m)\right)\right) \\
&= \log\left(\exp(m)\right) + \log\left(\exp(l_a - m) + \exp(l_b - m)\right) \\
&= m + \log\left(\exp(l_a - m) + \exp(l_b - m)\right)
\end{aligned}
$$

Cette façon de procéder se généralise pour une série:

- $m = \max(l_1, \ldots, l_n)$
- $\text{logadd}(l_1, \ldots, l_n) = m + \log\left(\exp(l_1 - m) + \ldots + \exp(l_n - m)\right)$

3 Sigmoid logistique

$\mathbb{R} \to (0, 1)$

$$\text{sigmoid(a)} = \frac{1}{1 + \exp(-\text{a})}$$

4 Tangente hyperbolique

$\mathbb{R} \to (-1, 1)$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$

5 Max $\mathbb{R}^d \to \mathbb{R}$ et ArgMax $\mathbb{R}^d \to \mathbb{N}^+$

- soit $a = (3.5, 25, -3, 2)$
- $\max(a) = 25$
- $\arg\max(a) = 2$ (attention, en python/numpy ce serait 1)

6 OneHot

$\mathbb{N}^+ \to \{0, 1\}^d$

- $\text{onehot}_5(2) = (0, 1, 0, 0, 0)$
- Propriété: $\max(a) = \langle a, \text{onehot}(\arg\max(a))\rangle$

7 Softmax

$\mathbb{R}^d \to \mathbb{R}^{+d}$

$$\begin{aligned}
\mathrm{softmax}((a_1,\ldots,a_d)) &= \left(\frac{\exp(a_1)}{\exp(a_1)+\ldots+\exp(a_d)},\ldots,\frac{\exp(a_d)}{\exp(a_1)+\ldots+\exp(a_d)}\right) \\
&= \left(\frac{\exp(a_1)}{\sum_{i=1}^d \exp(a_i)},\ldots,\frac{\exp(a_d)}{\sum_{i=1}^d \exp(a_i)}\right) \\
&= \frac{1}{\sum_{i=1}^d \exp(a_i)}\left(\exp(a_1),\,\ldots,\,\exp(a_d)\right)
\end{aligned}$$

Proriétés intéressantes:

Soit $s = \mathrm{softmax}(a)$ alors $\forall i,\ 0 < s_i < 1$

Et aussi: $\sum_{i=1}^d s_i = 1$

Les valeurs résultantes d'un softmax peuvent donc s'interpréter comme des probabilités de classe (ou catégories mutuellement exclusives)

Autre propriété intéressante:

$$\lim_{\alpha\to+\infty} \mathrm{softmax}(\alpha a) = \mathrm{onehot}(\arg\max(a))$$

7.1 Calcul numériquement stable du softmax

Pour éviter des problèmes avec des exp de nombres trop grands:

$$\begin{aligned}
\mathrm{softmax}((a_1,\ldots,a_d)) &= \frac{1}{\sum_{i=1}^d \exp(a_i)}\left(\exp(a_1),\,\ldots,\,\exp(a_d)\right) \\
&= \frac{1}{\sum_{i=1}^d \exp(a_i)/\exp(a_{max})}\left(\exp(a_1)/\exp(a_{max}),\,\ldots,\,\exp(a_d)/\exp(a_{max})\right) \\
&= \frac{1}{\sum_{i=1}^d \exp(a_i - a_{max})}\left(\exp(a_1 - a_{max}),\,\ldots,\,\exp(a_d - a_{max})\right)
\end{aligned}$$

où $a_{max} = \max(a)$

8 Rectifieur

ou "rectified linear unit" (RELU):

$$\mathrm{rect}(x) = \max(0, x)$$

9 Softplus

$$\mathrm{softplus}(x) = \log(1 + \exp(x))$$