Jérémie Béliveau-Lefebvre – 04 494 470
Sébastien Leblanc – 18 206 273

TP2 - solution

1. Nous devons donc trouver un $\vec{w}$ tel que $E(\vec{w})$est minimum. Pour ce faire, nous allons calculer le gradient $\nabla E(\vec{w})$ et trouver la racine de ce gradient soit :

Soit :

$$E(\vec{w}) = \sum_{n=1}^{N}(t_n - \vec{w}^T\vec{\phi}(\vec{x}_n))^2 + \lambda\vec{w}^T\vec{w}$$

Nous devons trouver :

$$\frac{\partial E(\vec{w})}{\partial\vec{w}} = 0 \text{ soit } \frac{\partial \sum_{n=1}^{N}(t_n - \vec{w}^T\vec{\phi}(\vec{x}_n))^2 + \lambda\vec{w}^T\vec{w}}{\partial\vec{w}} = 0$$

<=>

$$\sum_{n=1}^{N} -2\cdot\vec{\phi}(\vec{x}_n)\cdot(t_n - \vec{w}^T\vec{\phi}(\vec{x}_n)) + 2\cdot\lambda\vec{w} = 0 \qquad \text{par définition}$$

<=>

$$\sum_{n=1}^{N} -\vec{\phi}(\vec{x}_n)\cdot(t_n - \vec{w}^T\vec{\phi}(\vec{x}_n)) + \lambda\vec{w} = 0 \qquad \text{arithmétique}$$

<=>

$$\sum_{n=1}^{N} -\vec{\phi}(\vec{x}_n)\cdot t_n + \vec{\phi}(\vec{x}_n)\cdot\vec{w}^T\vec{\phi}(\vec{x}_n) + \lambda\vec{w} = 0 \qquad \text{par distributivité}$$

<=>

$$\sum_{n=1}^{N}\vec{\phi}(\vec{x}_n)\cdot\vec{w}^T\vec{\phi}(\vec{x}_n) + \lambda\vec{w} = \sum_{n=1}^{N}\vec{\phi}(\vec{x}_n)\cdot t_n \qquad \text{arithmétique}$$

<=>

$$\sum_{n=1}^{N}(\vec{\phi}(\vec{x}_n)^T\cdot\vec{\phi}(\vec{x}_n) + \lambda)^T\cdot\vec{w} = \sum_{n=1}^{N}\vec{\phi}(\vec{x}_n)\cdot t_n \qquad \text{arithmétique}$$

<=>

$$(\Phi^T\Phi + \lambda I)^T\cdot\vec{w} = \Phi t \qquad \text{trans. en matrices}$$

<=>

$$\vec{w} = (\Phi^T\Phi + \lambda I)^{-1}\Phi t \qquad \text{arithmétique}$$

CQFD

2. La fonction de perte lors d'une entropie croisée est donnée par :

$$E_D(\vec{w}) = -\sum_{n=1}^{N} t_n\ln\big(y_{\vec{w}}(\vec{x}_n)\big) + (1-t_n)\ln(1-y_{\vec{w}}(\vec{x}_n))$$

Nous devons trouver le gradient en effectuant :

$$\vec{\nabla}E(\vec{w}) = \frac{\partial E_D(\vec{w})}{\partial\vec{w}} = \frac{\partial\Big(-\sum_{n=1}^{N} t_n\ln\big(y_{\vec{w}}(\vec{x}_n)\big) + (1-t_n)\ln\big(1-y_{\vec{w}}(\vec{x}_n)\big)\Big)}{\partial\vec{w}}$$

Puisque :

$$y_{\vec{w}}(\vec{x}_n) = \sigma\left(\vec{w}^T\vec{\phi}(\vec{x}_n)\right) = \frac{1}{\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)} \qquad \text{par définition}$$

Il s'en suit que :

<=>

$$\frac{\partial -\sum_{n=1}^{N} t_n\ln\left(\frac{1}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}\right) + (1-t_n)\ln(1-\frac{1}{\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)})}{\partial\vec{w}} \qquad \text{par définition}$$

<=>

$$\frac{\partial -\sum_{n=1}^{N} t_n\ln\left(\frac{1}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}\right) + (1-t_n)\ln(\frac{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}} - \frac{1}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}})}{\partial\vec{w}} \qquad \text{arithmétique}$$

<=>

$$\frac{\partial -\sum_{n=1}^{N} t_n \ln\left(\frac{1}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}\right)+(1-t_n)\ln\left(\frac{e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}\right)}{\partial \vec{w}}$$ arithmétique

<=>

$$\frac{\partial -\sum_{n=1}^{N} -t_n \ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)+(1-t_n)\left(\ln\left(e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)-\ln(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)})\right)}{\partial \vec{w}}$$ par déf de ln

<=>

$$\frac{\partial -\sum_{n=1}^{N} -t_n \ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)+(1-t_n)\ln\left(e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)-(1-t_n)\ln(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)})}{\partial \vec{w}}$$

par distributivité : $(1-t_n)$

<=>

$$\frac{\partial -\sum_{n=1}^{N} -t_n \ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)-(1-t_n)\vec{w}^T\vec{\phi}(\vec{x}_n)-(1-t_n)\ln(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)})}{\partial \vec{w}}$$

par définition de $\ln(\exp(f(x)))$

<=>

$$\frac{\partial -\sum_{n=1}^{N}(-t_n-(1-t_n))\ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)-(1-t_n)\vec{w}^T\vec{\phi}(\vec{x}_n)}{\partial \vec{w}}$$

par associativité : $\ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)$

<=>

$$\frac{\partial -\sum_{n=1}^{N} -\ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)-(1-t_n)\vec{w}^T\vec{\phi}(\vec{x}_n)}{\partial \vec{w}}$$ arithmétique

<=>

$$\frac{\partial \sum_{n=1}^{N} \ln\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)+(1-t_n)\vec{w}^T\vec{\phi}(\vec{x}_n)}{\partial \vec{w}}$$ arithmétique

<=>

$$\sum_{n=1}^{N}\frac{\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)'}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}+(1-t_n)\vec{\phi}(\vec{x}_n)$$ par définition de la dérivée

<=>

$$\sum_{n=1}^{N}\frac{-\vec{\phi}(\vec{x}_n)e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}+(1-t_n)\vec{\phi}(\vec{x}_n)$$ par déf de dérivée d'une exp

<=>

$$\sum_{n=1}^{N}\left(\frac{-e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}+1-t_n\right)\vec{\phi}(\vec{x}_n)$$ par associativité $\vec{\phi}(\vec{x}_n)$

<=>

$$\sum_{n=1}^{N}\left(\frac{-e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}+1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}-t_n\right)\vec{\phi}(\vec{x}_n)$$ arithmétique

<=>

$$\sum_{n=1}^{N}\left(\frac{1}{1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}}-t_n\right)\vec{\phi}(\vec{x}_n)$$ arithmétique

<=>

$$\sum_{n=1}^{N}(y_{\vec{w}}(\vec{x}_n)-t_n)\vec{\phi}(\vec{x}_n)$$

par définition : $y_{\vec{w}}(\vec{x}_n)=\frac{1}{\left(1+e^{-\vec{w}^T\vec{\phi}(\vec{x}_n)}\right)}$

CQFD

3. Soit les 3 valeurs possibles {1 2 3} d'une variable aléatoire X, ayant la contrainte $p_1 = 2 \cdot p_2$, nous devons prouver que les propriétés sont les suivantes :

$$p_1 = \frac{2}{2^{2/3}+3}, p_2 = \frac{1}{2^{2/3}+3}, p_3 = \frac{2^{2/3}}{2^{2/3}+3}$$

Soit l'entropie :

$$H[x] = |E \quad [I(x)] = \sum p(x) I(x) = -\sum p(x) \log(p(x))$$

On cherche p1, p2, p3 ou l'entropie est maximale, soit :

$$\text{f (x)=> argmax}(-p_1(x)\log\big(p_1(x)\big) - p_2(x)\log\big(p_2(x)\big) - p_3(x)\log\big(p_3(x)\big))$$

$$g_1(\text{x}) => p_1 + p_2 + p_3 = 1$$

$$g_2(\text{x}) => p_1 + p_2 + p_3 = 1 => 2 \cdot p_2 + p_2 + p_3 = 1 => 3 \cdot p_2 + p_3 = 1$$

$$Lagrange : L = f + \lambda_1 g_1 + \lambda_2 g_2$$

$<=>$

$$L = -p_1(x)\log\big(p_1(x)\big) - p_2(x)\log\big(p_2(x)\big) - p_3(x)\log\big(p_3(x)\big) + \lambda_1(p_1 + p_2 + p_3 - 1) + \lambda_2(3 \cdot p_2 + p_3 - 1)$$

Et nous cherchons

$$\nabla L = 0 \ tel \ que \begin{bmatrix} \frac{\partial L}{\partial p_1} \\ \frac{\partial L}{\partial p_2} \\ \frac{\partial L}{\partial p_3} \\ \frac{\partial L}{\partial \lambda_1} \\ \frac{\partial L}{\partial \lambda_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial L}{\partial p_1} = -\log p_1 - \frac{\cancel{p_1}}{ln2 \cdot \cancel{p_1}} + \lambda_1 = 0 => \log p_1 = -\frac{1}{ln2} - \lambda_1 => p_1 = 2^{-\lambda_1 - \frac{1}{ln2}} \\ \frac{\partial L}{\partial p_2} = -\log p_2 - \frac{\cancel{p_2}}{ln2 \cdot \cancel{p_2}} + \lambda_1 + 3\lambda_2 = 0 => \log p_2 = -\lambda_1 - 3\lambda_2 - \frac{1}{ln2} => p_2 = 2^{-\lambda_1 - 3\lambda_2 - \frac{1}{ln2}} \\ \frac{\partial L}{\partial p_3} = -\log p_3 - \frac{\cancel{p_3}}{ln2 \cdot \cancel{p_3}} + \lambda_1 + \lambda_2 = 0 => \log p_3 = -\lambda_1 - \lambda_2 - \frac{1}{ln2} => p_3 = 2^{-\lambda_1 - \lambda_2 - \frac{1}{ln2}} \\ \frac{\partial L}{\partial \lambda_1} = p_1 + p_2 + p_3 - 1 = 0 => \lambda_1 = p_1 + p_2 + p_3 - 1 \\ \frac{\partial L}{\partial \lambda_2} = 3 \cdot p_2 + p_3 - 1 = 0 => \lambda_2 = 3 \cdot p_2 + p_3 - 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Nous avons donc que :

$$p_1 = 2^{-\lambda_1 - \frac{1}{ln2}}, p_2 = 2^{-\lambda_1 - 3\lambda_2 - \frac{1}{ln2}}, p_3 = 2^{-\lambda_1 - \lambda_2 - \frac{1}{ln2}}, \lambda_1 = p_1 + p_2 + p_3 - 1 \text{ et } \lambda_2 = 3 \cdot p_2 + p_3 - 1$$

Et puisque

$$p_1 + p_2 + p_3 - 1 = 0 \text{ et } 3 \cdot p_2 + p_3 - 1 = 0$$

Si on suppose une $p_2{}^* = 1$, donc d'après la 1ere contrainte où $p_1 = 2p_2$,
ça voudrait dire qu'une $p_1{}^* = 2$.

D'après les lois des probabilités, nous pouvons établis les relations suivantes pour $p_1{}^*$, $p_2{}^* et\ p_3{}^*$:

$$p_1 + p_2 + p_3 = 1$$

$<=>$

$$\frac{p_1{}^*}{p_1{}^* + p_2{}^* + p_3{}^*} + \frac{p_2{}^*}{p_1{}^* + p_2{}^* + p_3{}^*} + \frac{p_3{}^*}{p_1{}^* + p_2{}^* + p_3{}^*} = 1$$

Il suffit de trouver les $\lambda_1{}^*$ et $\lambda_2{}^*$ selon les λ_1 et λ_2 que nous avons trouvés précédemment.

$$p_1{}^* = 2^{-\lambda_1{}^* - \frac{1}{ln2}} = 2^1$$ par définition

<=>

$$-\lambda_1{}^* - \frac{1}{ln2} = 1$$ par équivalence des exposants

<=>

$$\lambda_1{}^* = -1 - \frac{1}{ln2}$$ arithmétique

Que nous remplaçons dans :

$$p_2{}^* = 1 = 2^{-\lambda_1{}^* - 3\lambda_2{}^* - \frac{1}{ln2}} = 2^0$$ par définition

<=>

$$-\lambda_1{}^* - 3\lambda_2{}^* - \frac{1}{ln2} = 0$$ par équivalence des exposants

<=>

$$-\left(-1 - \frac{1}{ln2}\right) - 3\lambda_2{}^* - \frac{1}{ln2} = 0$$ par remplacement : $\lambda_1{}^* = -1 - \frac{1}{ln2}$

<=>

$$(1) - 3\lambda_2{}^* = 0$$ arithmétique

<=>

$$\lambda_2{}^* = \frac{1}{3}$$ arithmétique

Que nous remplaçons avec $\lambda_1{}^*$ dans :

$$p_3{}^* = 2^{-\lambda_1{}^* - \lambda_2{}^* - \frac{1}{ln2}}$$ par définition

<=>

$$p_3{}^* = 2^{-\left(-1-\frac{1}{ln2}\right) - \frac{1}{3} - \frac{1}{ln2}}$$

par remplacement : $\lambda_1{}^* = -1 - \frac{1}{ln2}$ et $\lambda_2{}^* = \frac{1}{3}$

<=>

$$p_3{}^* = 2^{1-\frac{1}{3}}$$ arithmétique

<=>

$$p_3{}^* = 2^{2/3}$$ arithmétique

D'après la relation que nous avons établi précédemment :

$$p_1 + p_2 + p_3 = 1$$ tel qu'établi

<=>

$$\frac{p_1{}^*}{p_1{}^* + p_2{}^* + p_3{}^*} + \frac{p_2{}^*}{p_1{}^* + p_2{}^* + p_3{}^*} + \frac{p_3{}^*}{p_1{}^* + p_2{}^* + p_3{}^*} = 1$$ tel qu'établi

Nous retrouvons que :

$$p_1 = \frac{2}{2^{2/3}+3}, p_2 = \frac{1}{2^{2/3}+3}, p_3 = \frac{2^{2/3}}{2^{2/3}+3}$$

par remplacement et arithmétique

CQFD