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

TP4 - solution

1. Le gradient de perte par rapport à a_i est donné en dérivant la formule suivante par rapport à W :

$$E_D(W) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln y_{\vec{w}_i}(\vec{x})$$

Sachant que $y_{\vec{w}_i}(\vec{x}_n) = \frac{e^{a_i}}{\sum_c e^{a_c}}$ et $a_i = \overrightarrow{w_i}^{\mathrm{T}} \phi$

Nous pouvons trouver $\frac{\partial E(w)}{\partial w}$ par la règle de dérivée en chaîne. Ainsi nous pouvons décortiquer le problème tel qu'illustré ici :

$$\frac{\partial E(w)}{\partial w} = \frac{\partial E(w)}{\partial y_{\vec{w}_i}(\vec{x}_n)} \frac{\partial y_{\vec{w}_i}(\vec{x}_n)}{\partial w} = \frac{\partial E(w)}{\partial y_{\vec{w}_i}(\vec{x}_n)} \frac{\partial y_{\vec{w}_i}(\vec{x}_n)}{\partial a_i} \frac{\partial a_i}{\partial w}$$

Et si nous simplifions, nous pouvons effectuer le travail suivant :

$$\frac{\partial E(w)}{\partial A} \cdot \frac{\partial A}{\partial B} \cdot \frac{\partial B}{\partial w}$$

Où : $A = y_{\vec{w}_i}(\vec{x})$

Et $B = a_i$

Ainsi

1) $\frac{\partial E(w)}{\partial A}$

$$\frac{\partial E(w)}{\partial A} = \frac{\partial -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln y_{\vec{w}_i}(\vec{x}_n)}{\partial y_{\vec{w}_i}(\vec{x}_n)}$$

<=>

$$\frac{\partial -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln \mathrm{A}}{\partial A}$$ par définition

<=>

$$-\sum_{n=1}^{N}\sum_{k=1}^{K} \frac{t_{kn}}{\mathrm{A}}$$ par dérivée

2) $\frac{\partial A}{\partial B}$

$$\frac{\partial A}{\partial B} = \frac{\partial \frac{e^{a_i}}{\sum_c e^{a_i}}}{\partial a_i}$$

<=>

$$\frac{\partial \frac{e^{B}}{\sum_c e^{B}}}{\partial B}$$ par définition

<=>

$$\frac{e^B \sum_c e^B - e^B e^B}{\left(\sum_c e^B\right)^2} \qquad \text{selon } \frac{\partial \frac{g(x)}{f(x)}}{\partial x} = \frac{g'(x)\mathrm{f}(x) - \mathrm{g}(x)\mathrm{f}'(x)}{f^2(x)}$$

<=>

$$\frac{e^B(\sum_c e^B - e^B)}{\sum_c e^B} \qquad \text{par associativité}$$

<=>

$$\frac{e^B}{\sum_c e^B}\left(\frac{\sum_c e^B - e^B}{\sum_c e^B}\right) \qquad \text{arythmétique}$$

<=>

$$\frac{e^B}{\sum_c e^B}\left(\frac{\sum_c e^B}{\sum_c e^B} - \frac{e^B}{\sum_c e^B}\right) \qquad \text{arythmétique}$$

<=>

$$\frac{e^B}{\sum_c e^B}\left(I_n - \frac{e^B}{\sum_c e^B}\right) \qquad \text{arythmétique}$$

<=>

$$A(I_n - A) \qquad \text{par définition}$$

3) $\frac{\partial B}{\partial w}$

$$\frac{\partial B}{\partial w} = \frac{\partial \vec{\mathrm{w}_i}^{\mathrm{T}} \vec{\mathrm{x}}}{\partial w}$$

<=>

$$\vec{\mathrm{x}} \qquad \text{par dérivation}$$

Il ne reste qu'à remettre les solutions trouvés dans la définition de dérivée en chaine plus haut :

$$\frac{\partial E(w)}{\partial A} \cdot \frac{\partial A}{\partial B} \cdot \frac{\partial B}{\partial w}$$

<=>

$$-\sum_{n=1}^{N}\sum_{k=1}^{K} \frac{t_{kn}}{\mathrm{A}} \cdot A(I_n - A) \cdot \vec{x}_n \qquad \text{par définition}$$

<=>

$$-\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn}\,(I_n - A) \cdot \vec{x}_n \qquad \text{arythmétique}$$

<=>

$$\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn}\,(A - I_n) \cdot \vec{x}_n \qquad \text{arythmétique}$$

<=>

$$\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn}\left(y_{\vec{w}_i}(\vec{x}_n) - I_n\right) \cdot \vec{x}_n \qquad \text{par définition}$$

<=>

$$\sum_{n=1}^{N}\sum_{k=1}^{K} \left(t_{kn} y_{\vec{w}_i}(\vec{x}_n) - t_{kn}\right) \cdot \vec{x}_n \qquad \text{par distributivité}$$

<=>

$$\sum_{n=1}^{N}\left(y_{\vec{w}_i}(\vec{x}_n) - t_{ni}\right) \cdot \vec{x}_n \qquad \text{par définition d'un « one-hot »}$$

CQFD

2.

a. Une distribution de vraisemblance est la distribution de probabilité conditionnelle d'observer un paramètre dans une classe donnée, ex : $p(x|t)$ soit la probabilité d'observer x étant donné t. Pour la calculer, nous pourrions supposer que chaque classe de véhicule suit une distribution gaussienne et ainsi calculer une gaussienne par classe. Ainsi, dans chaque classe t, nous aurons une formule gaussienne qui nous permet d'en connaitre la répartition x.
b. La distribution à priori est la distribution de probabilité d'observer une catégorie divisée par le nbr de véhicule total, ex : $p(t)$. Pour chaque catégorie, nous pourrions calculer le nombre de véhicule divisé par le nombre total de véhicule.
c. Oui, puisque les voitures sport sont de même catégorie et ont plus de chance de consommer tous beaucoup d'essence et peu de chance qu'une de ces voitures consomme très peu d'essence.

3. Si nous partons des principes de forces qui s'appliquent sur un corps en mouvement, nous avons les paramètres suivants : la masse d'un objet, la gravité qui agit sur celui-ci, la pression du sol sur ce corps, la pente du sol sur cette masse ainsi que la friction entre la masse et le sol. Ainsi, l'équilibre complexe ainsi formé peut être résumé dans les fonctions suivantes :

$$a = \frac{(v_t - v_{t-\delta})}{\delta}$$
$$v = \frac{(d_t - d_{t-\delta})}{\delta}$$

où

$$\delta = \frac{d_t - d_{t-\delta}}{v_t - v_{t-\delta}}$$

et où

a : accélération
v : vitesse
d : position
δ: variation de pente de vitesse

Le formule de descente de gradient de type momentum se définissent comme suit :

$$w_{t+1} = w_t - \eta \nabla E_{\vec{x}_n}(w_t)$$
$$v_{t+1} = \rho v_t - \nabla E_{\vec{x}_n}(w_t)$$
$$w_{t+1} = w_t - \eta v_{t+1}$$

Où

w : un certain poid
v : vitesse
η : constante de contrôle de vitesse initiale
ρ constante de contrôle de variation de vitesse
t : temps donné
$\nabla E_{\vec{x}_n}$: gradient

Pour mettre en valeur les liens entre la descente de gradient de type momentum et la position, vitesse et accélération, il suffit de prendre les équations de vitesse et d'accélération, d'isoler v_t et d_t et d'étudier les similarités. Nous obtiendrons les équations suivantes :

$$d_t = d_{t-\delta} + v\delta$$
$$v_t = v_{\mathrm{t}-\delta} + a\delta$$

Par rapport aux formules du momentum suivantes :

$$w_{t+1} = w_t - \eta v_{t+1}$$
$$v_{t+1} = \rho v_t - \nabla E_{\vec{x}_n}(w_t)$$

- Les w_t tout comme d_t peuvent d'abord être interprétés comme la hauteur de la masse dans une
  pente en un temps t.
- η peut être vue comme étant la variation de l'itération à laquelle on regarde la vitesse.
- ρ peut être vu comme la friction qui ralenti la masse
- $\nabla E_{\vec{x}_n}(w_t)$ et l'accélération (a) peuvent être vu comme la force qui fait varier la position de la masse dans la pente (la gravité).
- La différences des signes + et – s'explique par la direction d'application des variations de vitesses et de gradient sur la masse. Dans le gradient, + signifie que le poid remonte alors que dans la vitesse, un δ + signifie que la masse accélère vers le bas et prend plus de distance.

C'était les similitudes entre la descente de gradient de type momentum et la position, vitesse et accélération.