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 $a_i = \vec{w_i}^{\mathrm{T}}\phi$

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 :

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

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

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

Où :

$$A = y_{\vec{w}_i}(\vec{x})$$

$$\text{Et B} = a_i$$

Ainsi

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

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

<=>

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

<=>

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

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

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

<=>

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

<=>

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

$<=>$

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

$<=>$

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

$<=>$

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

$<=>$

$$\frac{e^B}{\sum_c\ e^B}\left(1 - \frac{e^B}{\sum_c\ e^B}\right)$$ arythmétique

$<=>$

$$A(1-A)$$ par définition

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

$$\frac{\partial B}{\partial w} = \frac{\partial \overrightarrow{\mathrm{w}_i}^{\mathrm{T}} \phi}{\partial w}$$

$<=>$

$$\phi$$ par dérivation

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

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

$<=>$

$$-\sum_{n=1}^{N} \sum_{k=1}^{K} \frac{t_{kn}}{\mathrm{A}} \cdot A(1-A) \cdot \phi$$ par définition

$<=>$

$$-\sum_{n=1}^{N} \sum_{k=1}^{K} t_{kn}\,(1-A)\phi$$ arythmétique

$<=>$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} t_{kn}\,(A-1)\phi$$ arythmétique

$<=>$

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

$<=>$

$$\frac{1}{N}\sum_{n=1}^{N} \vec{x}(y_{\vec{w}}(\vec{x}) - t_{kn})$$ par une twist quelconque

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. 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

a : accélération
v : vitesse
d distance
i : initiale

Les formules de position, vitesses et accélération en fonction du temps sont définis de la manière suivante :

$$a = \frac{(v_f - v_i)}{\Delta t}$$
$$v = \frac{(p_f - p_i)}{\Delta t}$$

$$a = \frac{(v_i - v_{i-\delta})}{\Delta t} \text{ vs } a = \frac{(v_i - v_{i-\delta})}{d}$$
$$v = \frac{(d_i - d_{i-\delta})}{\Delta t} \text{ vs } v = \frac{(d_i - d_{i-\delta})}{d}$$

Où

a : accélération
v : vitesse
p position
i : initiale
f : finale
Δt: période de temps

De même, lorsque nous voulons calculer la nouvelle position d'un objet en mouvement, nous pouvons isoler p_f de la formule de vitesse :

$$p_f = p_i + v\Delta t$$
par rapport à
$$w_{t+1} = w_t - \eta v_{t+1}$$

Laquelle est similaire que pour trouver w_{t+1}. La variable η sert à contrôler la vitesse de déplacement dans le déplacement suivant notre gradient.

Lorsque nous voulons calculer la nouvelle vitesse, nous devons isoler v_f de la formule d'accélération :

$$v_f = v_i + a\Delta t$$

par rapport à

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

Laquelle est similaire que pour trouver v_{t+1}.

CQFD