UNIVERSITÉ DE SHERBROOKE

TP2

IFT 712 Technique d'apprentissage

Eliott THOMAS — 21 164 874
thoe2303@usherbrooke.ca

Lilian FAVRE GARCIA — 21 153 421
favl2301@usherbrooke.ca

Tsiory Razafindramisa — 21 145 627
raza3902@usherbrooke.ca

Travail présenté à
Martin Vallières

Université de Sherbrooke
Département d'informatique
Date de remise : 10 mars 2022

Table des matières

1 Introduction

On s'intéresse dans ce second TP à des démonstrations ainsi qu'à la mise en place d'un programme informatique permettant une classification linéaire sur des données. Nous avons le choix entre trois algorithmes : un modèle génératif, le modèle du Perceptron codé par nous-même et enfin un modèle d'un Perceptron mais codé à l'aide de Sklearn.

2 Question 1

$$
\begin{aligned}
E(\vec{w}) &= \sum_{n=1}^{N} \left(t_n - \vec{w}^T \vec{\phi}(\vec{x}_n)\right)^2 + \lambda \vec{w}^T \vec{w} \\
&= \sum_{n=1}^{N} \left(\vec{w}^T \vec{\phi}(\vec{x}_n) - t_n\right)^2 + \lambda \vec{w}^T \vec{w} \\
&\Leftrightarrow 0 = \nabla_{\vec{w}} E_D(\vec{w}) \\
&\Leftrightarrow 0 = \sum 2\phi(x_n) * \left(\phi^T(x_n)w - t_n^T\right) + 2\lambda w \\
&\Leftrightarrow 0 = \sum \phi(x_n) * \left(\phi^T(x_n)w - t_n^T\right) + \lambda w \\
&\Leftrightarrow 0 = \left(\sum \phi(x_n)\phi^T(x_n)\right) w - \left(\sum \phi(x_n) t_n^T\right) + (\lambda I) w \\
&\Leftrightarrow (\phi^T \phi + \lambda I)\vec{w} = \sum \phi(x_n) t_n^T \\
&\Leftrightarrow (\phi^T \phi + \lambda I)\vec{w} = \phi^T t \\
&\Leftrightarrow \vec{w} = (\phi^T \phi + \lambda I)^{-1} \phi^T t
\end{aligned}
$$

3 Question 2

On sait que

$$
\begin{aligned}
&(fog)' = (f'og) * g' \\
&\tfrac{\partial \sigma(X)}{\partial x} = \sigma(X) * (1 - \sigma(X))
\end{aligned}
$$

Ainsi,

$$
\begin{aligned}
\tfrac{\partial \sigma(w^T\phi)}{\partial w} &= \sigma(w^T\phi) * (1 - \sigma(w^T\phi)) * \tfrac{\partial (w^T\phi)}{\partial w} \\
&= \sigma(w^T\phi) * (1 - \sigma(w^T\phi)) * \phi
\end{aligned}
$$

On note également pour simplifier la lecture

$$
\sigma(w^T\phi(x_n)) = y_n
$$

On part donc de la formule de l'entropie croisée

$$\sum_{n=1}^{N} t_n * ln(y_n) + (1 - t_n) * ln(1 - y_n)$$

Que l'on dérive

$$\nabla_{\vec{w}} E_D(\vec{w}) = -\sum_{n=1}^{N} t_n(1 - y_n) * \phi(x_n) + (1 - t_n) * (-y_n) * \phi(\mathrm{x}_n)$$
$$\nabla_{\vec{w}} E_D(\vec{w}) = -\sum_{n=1}^{N} (t_n - t_n * y_n + t_n * y_n - y_n) * \phi(x_n)$$
$$\nabla_{\vec{w}} E_D(\vec{w}) = -\sum_{n=1}^{N} (t_n - y_n) * \phi(x_n)$$
$$\nabla_{\vec{w}} E_D(\vec{w}) = \sum_{n=1}^{N} (y_n - t_n) * \phi(x_n)$$

4 Question 3

On a $\mathscr{L} = -p_1.log_2(p_1) - p_2.log_2(p_2) - p_3.log_2(p_3) - \lambda.(p_1 + p_2 + p_3 - 1) - \mu.(p_1 - 2.p_2)$

On veut dériver par rapport à p1, p2, p3. Et mettre les gradients à 0.

(1) $\frac{\partial \mathscr{L}}{\partial p_1} = -log_2(p1) - log_2(e) - \lambda - \mu = 0$

(2) $\frac{\partial \mathscr{L}}{\partial p_2} = -log_2(p2) - log_2(e) - \lambda + 2.\mu = 0$

(3) $\frac{\partial \mathscr{L}}{\partial p_3} = -log_2(p3) - log_2(e) - \lambda = 0$

Par ailleurs on a les contraintes suivantes :
(4) $p_1 + p_2 + p_3 = 1$

(5) $p_1 = 2 * p_2$

On commence par (1) - (2) :

$$-log_2(2.p_2) - log_2(e) - \lambda - \mu + log_2(p_2) + log_2(e) + \lambda - 2.\mu = 0$$

$$\Leftrightarrow -log_2(2) - log_2(p_2) - 3.\mu + log_2(p_2) = 0$$

$$\Leftrightarrow -1 - 3.\mu = 0$$

$$\Leftrightarrow \mu = -\frac{1}{3}$$

En utilisant (4) et (5) on obtient (6) : $p3 = 1 - 3.p_2$
On rentre (6) dans (3) pour obtenir (3').
Puis on fait (2) - (3') :

$$-log_2(p_2) - log_2(e) - \lambda + 2.\mu + log_2(1 - 3p_2) + log_2(e) + \lambda = 0$$

$\Leftrightarrow -log_2(p_2) + 2.(-\frac{1}{3}) + log_2(1 - 3p_2) = 0$

$\Leftrightarrow \frac{2}{3} = log_2\left(\frac{1-3p_2}{p_2}\right)$

$\Leftrightarrow 2^{\frac{2}{3}} = \frac{1-3p_2}{p_2}$

$\Leftrightarrow 2^{\frac{2}{3}}.p_2 = 1 - 3p_2$

$\Leftrightarrow 2^{\frac{2}{3}}.p_2 + 3.p_2 = 1$

$\Leftrightarrow p_2(2^{\frac{2}{3}} + 3) = 1$

$\Leftrightarrow p_2 = \frac{1}{2^{\frac{2}{3}}+3}$

En utilisant (5) : $p_1 = 2 * p_2 = \frac{2}{2^{\frac{2}{3}}+3}$

En utilisant (6) : $p_3 = 1 - 3.p_2 = 1 - 3.\frac{1}{2^{\frac{2}{3}}+3} = \frac{3+2^{\frac{2}{3}}-3}{2^{\frac{2}{3}}+3} = \frac{2^{\frac{2}{3}}}{2^{\frac{2}{3}}+3}$

Finalement on a bien :

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

5 Question 4

Voir code remis. Le lien Github est ici