IFT 712: Devoir n°2

- caiy2401   - CAI Yunfan

1- [1.5 point] Prouvez que la régression de Ridge ayant pour objectif de minimiser la fonction d'énergie suivante (voir les notes pdf sur la régression linéaire):

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

a pour solution l'équation suivante :

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

On veut minimiser la fonction alors:

$$0 = \nabla_{\vec{W}} \vec{E}(\vec{W})$$

$$0 = -\sum_{n=1}^{N} [(t_n - \vec{W}^T \vec{\Phi}(x_n))\vec{\Phi}^T(x_n) + \lambda \vec{W}^T]$$

$$0 = \sum_{n=1}^{N} [\vec{W}^T \vec{\Phi}(x_n)\vec{\Phi}^T(x_n) - t_n \vec{\Phi}(x_n)] + \lambda \vec{W}^T$$

$$\sum_{n=1}^{N} t_n \vec{\Phi}^T(x_n) = \sum_{n=1}^{N} [\vec{W}^T \vec{\Phi}(x_n)\vec{\Phi}^T(x_n)] + \lambda \vec{W}^T$$

$$T^T \Phi = \vec{W}^T \cdot (\lambda I + \Phi^T \Phi)$$

$$\vec{W}^T = T^T \Phi \cdot (\lambda I + \Phi^T \Phi)^{-1}$$

$$\vec{W} = (T^T \Phi \cdot (\lambda I + \Phi^T \Phi)^{-1})^T$$

$$\vec{W} = ((\lambda I + \Phi^T \Phi)^T)^{-1} \cdot (T^T \Phi)^T$$

$$\vec{W} = (\lambda I + \Phi^T \Phi) \cdot \Phi^T T$$

d'où le résultat

2- [2.5 points] Nous avons vu que l'objectif de la régression logistique est de trouver un vecteur de paramètres $\vec{w}$ pouvant reproduire la probabilité conditionnelle suivante

$$p(C_1|\vec{\phi}(\vec{x})) = \sigma(\vec{w}^T\vec{\phi}(\vec{x})).$$

Prouvez que lorsque la fonction de perte est une entropie croisée (*cross-entropy*), le gradient utilisé pour effectuer la descente de gradient est déterminé par la fonction suivante :

$$\vec{\nabla}E(\vec{w}) = \sum_{n=1}^{N}(y_n - t_n)\vec{\phi}(\vec{x}_n).$$

On calcule la dérivé de $y_{\vec{W}}(\Phi(\vec{x}_n))$:

$$\begin{aligned}
\frac{\delta}{\delta W}(\vec{y}_W(\Phi(\vec{x}_n))) &= \frac{\delta}{\delta W}\left(\frac{1}{1+\exp(-W^T\vec{\Phi}(\vec{x}_n))}\right)\\
&= \frac{\exp(-W^T\Phi(\vec{x}_n))}{(1+\exp(-W^T\vec{\Phi}(\vec{x}_n)))^2}\cdot\vec{\Phi}(\vec{x}_n)\\
&= y_W^{\ 2}\cdot\vec{\Phi}(\vec{x}_n)\cdot\left(\frac{1}{y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))} - 1\right)\cdot\vec{\Phi}(\vec{x}_n)
\end{aligned}$$

On a alors :

$$\begin{aligned}
E_D(\vec{W}) &= -\sum_{n=1}^{N} t_n\ln(y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)) + (1-t_n)\ln(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\\
\nabla_W E_D(\vec{W}) &= -\sum_{n=1}^{N} t_n\cdot\frac{\delta}{\delta W}(\vec{y}_W(\vec{\Phi}(\vec{x}_n)))\cdot\frac{1}{\vec{y}_W(\vec{\Phi}(\vec{x}_n))} + (1-t_n)\cdot\frac{-1}{1-\vec{y}_W(\vec{\Phi}(\vec{x}_n))}\cdot\frac{\delta}{\delta W}(\vec{y}_W(\vec{\Phi}(\vec{x}_n)))\\
&= -\sum_{n=1}^{N}\frac{\delta}{\delta W}(\vec{y}_W(\vec{\Phi}(\vec{x}_n)))\cdot\left(\frac{t_n}{\vec{y}_W(\vec{\Phi}(\vec{x}_n))} - \frac{(1-t_n)}{1-\vec{y}_W(\vec{\Phi}(\vec{x}_n))}\right)\\
&= -\sum_{n=1}^{N}(y_{\vec{W}}^{\ 2}(\vec{\Phi}(\vec{x}_n)))\cdot\left(\frac{1}{y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))} - 1\right)\cdot\vec{\Phi}(\vec{x}_n).\left(\frac{t_n}{y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))} - \frac{(1-t_n)}{1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))}\right)\\
&= -\sum_{n=1}^{N}(y_{\vec{W}}^{\ 2}(\vec{\Phi}(\vec{x}_n)))\cdot\left(\frac{1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))}{y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))}\right)\cdot\vec{\Phi}(\vec{x}_n).\left(\frac{t_n.(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)) - (1-t_n).y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))}{y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\,(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)))}\right)\\
&= -\sum_{n=1}^{N}(y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)))\cdot(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)))\cdot\vec{\Phi}(\vec{x}_n).\left(\frac{t_n.(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)) - (1-t_n).y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))}{y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\,(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)))}\right)\\
&= -\sum_{n=1}^{N}\vec{\Phi}(\vec{x}_n).(t_n.(1-y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)) - (1-t_n).y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)))\\
&= -\sum_{n=1}^{N} t_n\cdot\vec{\Phi}(\vec{x}_n) - t_n\cdot y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\cdot\vec{\Phi}(\vec{x}_n) - y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\cdot\vec{\Phi}(\vec{x}_n) + t_n\cdot y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\cdot\vec{\Phi}(\vec{x}_n)\\
&= -\sum_{n=1}^{N} t_n\cdot\vec{\Phi}(\vec{x}_n) - y_{\vec{W}}(\vec{\Phi}(\vec{x}_n))\cdot\vec{\Phi}(\vec{x}_n)\\
&= \sum_{n=1}^{N}(y_{\vec{W}}(\vec{\Phi}(\vec{x}_n)) - t_n)\,\vec{\Phi}(\vec{x}_n)
\end{aligned}$$

3- [1.0 point] Soit une variable aléatoire X pouvant prendre 3 valeurs possibles {1,2,3} avec probabilités $P(X=1) = p_1$, $P(X=2) = p_2$ et $P(X=3) = p_3$. Démontrez que la loi de probabilité ayant l'entropie la plus élevé et satisfaisant la contrainte $p_1 = 2p_2$ a les probabilités suivantes :

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

Pour ce faire, utilisez des multiplicateurs de Lagrange afin de tenir compte de la contrainte de sommation à 1 et de la contrainte $p_1 = 2p_2$. Pour le calcul de l'entropie, prenez un logarithme en base 2.

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

On a :

$$E(I[X]) = \sum_{i=1}^{3} p_i \log_2(p_i)$$

avec $p_1 + p_2 + p_3 = 1$et $p_1 = 2p_2$

1. on obtient: $g_1(p_1, p_2, p_3) = p_1 + p_2 + p_3 - 1 = 0$

   $g_2(p_1, p_2, p_3) = p_1 - 2p_2 = 0$

alors :

$$E(I[X]) = -p_1\log_2 p_1 - p_2\log_2 p_2 - p_3\log_2 p_3$$

et :

$$\mathcal{L} = -p_1\log_2 p_1 - p_2\log_2 p_2 - p_3\log_2 p_3 + \lambda_1(p_1 - 2p_2) + \lambda_2(p_1 + p_2 + p_3 - 1)$$

$$\frac{\delta\mathcal{L}}{\delta p_1} = -\frac{1}{\ln 2}\cdot p_1 \cdot \frac{1}{p_1} - \log_2 p_1 + \lambda_1 + \lambda_2 = 0$$

$$\Rightarrow \log_2 p_1 = -\frac{1}{\ln 2} + \lambda_1 + \lambda_2$$

$$\Rightarrow p_1 = 2^{(\lambda_1+\lambda_2-\frac{1}{\ln 2})}$$

$$\frac{\delta\mathcal{L}}{\delta p_2} = -\frac{1}{\ln 2}\cdot p_2 \cdot \frac{1}{p_2} - \log_2 p_2 - 2\lambda_1 + \lambda_2 = 0$$

$$\Rightarrow \log_2 p_2 = -\frac{1}{\ln 2} - 2\lambda_1 + \lambda_2$$

$$\Rightarrow p_2 = 2^{(-2\lambda_1+\lambda_2-\frac{1}{\ln 2})}$$

$$\frac{\delta\mathcal{L}}{\delta p_3} = -\frac{1}{\ln 2}\cdot p_3 \cdot \frac{1}{p_3} - \log_2 p_3 + \lambda_2 = 0$$

$$\Rightarrow \log_2 p_3 = -\frac{1}{\ln 2} + \lambda_2$$

$$\Rightarrow p_3 = 2^{(\lambda_2-\frac{1}{\ln 2})}$$

on en déduit que $p_1 = 2^{\lambda_1} p_3 \qquad p_2 = 2^{-2\lambda_1} p_3$

donc en remplaçant dans g_2:

$$g_2(p_1, p_2, p_3) = p_1 - 2p_2 = 0$$

$$\Leftrightarrow 2^{\lambda_1} p_3 - 2\cdot 2^{-2\lambda_1} p_3 = 0$$

$$\Leftrightarrow p_3(2^{\lambda_1} - 2^{-2\lambda_1+1}) = 0$$

$$\Leftrightarrow 2^{\lambda_1} = 2^{-2\lambda_1+1}$$

$$\Leftrightarrow \lambda_1 = -2\lambda_1 + 1$$

$$\Leftrightarrow \lambda_1 = 1/3$$

de plus on a $g_1(p_1, p_2, p_3) = p_1 + p_2 + p_3 - 1 = 3p_2 + p_3 - 1 = 0$

$$\Leftrightarrow p_3 + 3 \cdot 2^{-2/3} p_3 = 1$$

$$\Leftrightarrow p_3 = \frac{1}{1+3\cdot 2^{-\frac{2}{3}}} = \frac{2^{2/3}}{2^{2/3}+3}$$

et $\Rightarrow p_2 = 2^{-\frac{2}{3}} p_3 = \frac{1}{2^{2/3}+3}$

et $\Rightarrow p_1 = 2p_2 = \frac{2}{2^{2/3}+3}$