1- M. que la regression de Ridge ayant pour objectif de minimiser la fct d'energie.

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

a pour solution: $\vec{w} = (\phi^T\phi + \lambda I)^{-1} \phi^T t$

En forçant le gradient de E à 0 on a:

$$\nabla_{\vec{w}} E(\vec{w}) = 0 \Rightarrow \sum_{n=1}^{N} (t_n - \vec{w}^T \vec{\phi}(\vec{x}_n)) \times 2 \vec{\phi}^T(\vec{x}_n) + 2\lambda \vec{w}^T = 0$$

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

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

$$\Rightarrow \vec{w}^T (\lambda I + \phi^T \phi) = \phi^T t$$

$$\Rightarrow \boxed{\vec{w} = (\phi^T\phi + \lambda I)^{-1} \phi^T t}$$

2- Regression logistique: $p(C_1 | \vec{\phi}(\vec{x}_n)) = \sigma(\vec{w}^T \vec{\phi}(\vec{x}_n))$

Si la perte est une cross entropy:

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

La fonction de perte est une entropie croisée

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

$$y_n = y_{\vec{w}}(\phi(\vec{x}_n)) = \sigma(\vec{w} \vec{\phi}(x_n)) = P(C_1 | \vec{\phi}(\vec{x}_n))$$

$$y_n = y_{\vec{w}}(\phi(\vec{x}_n)) = \frac{1}{1 + e^{-\vec{w}^T \vec{\phi}(x_n)}}$$

$$\Rightarrow E(\vec{w}) = -\sum_{n=1}^{N} t_n \ln(y_n) + (1-t_n) \ln(1-y_n)$$

$$\nabla E(\vec{w}) = -\sum_{n=1}^{N} t_n \left(\frac{dy_n}{d\vec{w}}\right)\left(\frac{1}{y_n}\right) + (1-t_n)\left(\frac{dy_n}{d\vec{w}}\right)\left(\frac{-1}{1-y_n}\right)$$

$$= -\sum_{n=1}^{N} \left(\frac{dy_n}{d\vec{w}}\right)\left(\frac{t_n}{y_n} - \frac{1-t_n}{1-y_n}\right)$$

De plus $$\frac{dy_n}{d\vec{w}} = \frac{d\left(\frac{1}{1+e^{-\vec{w}^T \phi(x_n)}}\right)}{d\vec{w}}$$

$$\Rightarrow \frac{dy_n}{d\vec{w}} = -\frac{d}{d\vec{w}}\left(1 + e^{-\vec{w}^T \vec{\phi}(\vec{x}_n)}\right) \times \frac{1}{(1+e^{-\vec{w}^T \vec{\phi}(\vec{x}_n)})^2}$$

$$= \vec{\phi}(\vec{x}_n) e^{-\vec{w}^T \phi(\vec{x}_n)} \times \frac{1}{(1+e^{-\vec{w}^T \vec{\phi}(\vec{x}_n)})^2}$$

$$\frac{dy_n}{d\vec{w}} = \vec{\phi}(\vec{x}_n)\left(\frac{1}{y_n} - 1\right) y_n^2$$ by: $y_n = \frac{1}{1+e^{-\vec{w}^T \vec{\phi}(\vec{x}_n)}}$

$$\Rightarrow \frac{dy_n}{d\vec{w}} = \vec{\phi}(\vec{x}_n)(1-y_n) y_n$$ $\quad \frac{1}{y_n} = 1 + e^{-\vec{w}^T \vec{\phi}(\vec{x}_n)}$

$$\Rightarrow \nabla E(\vec{w}) = -\sum_{n=1}^{N} \vec{\phi}(\vec{x_n})(1-y_n)y_n\left(\frac{t_n}{y_n} - \frac{1-t_n}{1-y_n}\right)$$

$$= -\sum_{n=1}^{N} \vec{\phi}(\vec{x_n})\left((1-y_n)t_n - (1-t_n)y_n\right)$$

$$= -\sum_{n=1}^{N} \vec{\phi}(\vec{x_n})\left(t_n - y_n t_n - 1 + t_n y_n\right)$$

$$= -\sum_{n=1}^{N} \vec{\phi}(\vec{x_n})(t_n - 1)$$

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

$$\hookrightarrow \boxed{\nabla E(\vec{w}) = \sum_{n=1}^{N} (y_n - t_n)\vec{\phi}(\vec{x_n})}$$

3. Soit X une variable aléatoire tq:

$$\begin{cases} P(X=1) = p_1 \\ P(X=2) = p_2 \\ P(X=3) = p_3 \end{cases} \quad \text{tq } p_1 = 2p_2$$

l'entropie associée $H[x] = -\sum_{i=1}^{3} p_i \log_2(p_i)$ tq $\sum_{i=1}^{3} p_i = 1 \Rightarrow p_1 + p_2 + p_3 = 1$

$$= -p_1\log_2(p_1) - p_2\log_2(p_2) - p_3\log_2(p_3)$$

l'entropie la plus élevée satisfaisant la contrainte $p_1 = 2p_2$ s'exprime en lagrangien:

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

on a: $\frac{d}{dx}\log_a(u) = \log_a e \times \frac{1}{u} \times \frac{du}{dx}$

$\nabla \mathcal{L} = 0$

$$\frac{d\mathcal{L}}{p_1} = -\log_2(p_1) - \log_2(e) - \lambda_1 - \lambda_2 = 0 \qquad (a)$$

$$\frac{d\mathcal{L}}{p_2} = -\log_2(p_2) - \log_2(e) + 2\lambda_1 - \lambda_2 = 0 \qquad (b)$$

$$\frac{d\mathcal{L}}{p_3} = -\log_2(p_3) - \log_2(e) - \lambda_2 = 0 \qquad (c)$$

$$\frac{d\mathcal{L}}{d\lambda_1} = -(p_1 - 2p_2) = 0 \qquad (d)$$

$$\frac{d\mathcal{L}}{d\lambda_2} = -(p_1 + p_2 + p_3 - 1) = 0 \qquad (e)$$

$$(c) \Rightarrow -\log_2(p_3) - \log_2(e) = \lambda_2$$

$$\Rightarrow \boxed{p_3 = 2^{-(\log_2(e)+\lambda_2)}}$$

$$(a)-(b) \Rightarrow -\log_2 p_1 + \log_2 p_2 - \lambda_1 - 2\lambda_1 = 0$$

$$\Rightarrow \lambda_1 = \frac{1}{3}\log_2\left(\frac{p_2}{p_1}\right)$$

$$= \frac{1}{3}\log\left(\frac{p_2}{2p_2}\right)$$

$$= -\frac{1}{3}\log(2)$$

$$\boxed{\lambda_1 = -\frac{1}{3}}$$

On a : $\begin{cases} p_1 = 2p_2 \\ p_1 + p_2 + p_3 = 1 \end{cases} \Rightarrow p_3 = 1-3p_2$

$$(b)-(c) \Rightarrow -\log_2 p_2 + \log_2 p_3 + 2\lambda_1 = 0 \quad \hookrightarrow \lambda_1 = \frac{1}{3}$$

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

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

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

$$\Rightarrow \boxed{p_2 = \frac{1}{2^{2/3}+3}}$$

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

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

$$= \frac{2^{2/3}+\cancel{3}-\cancel{3}}{2^{2/3}+3}$$

$$p_3 = \frac{2^{2/3}}{2^{2/3}+3}$$