Fonction de perte quadratique et dérivées de matrices

m : exemples, d : dimension

On représente les données par une matrice à m lignes et d colonnes (les exemples toujours en ligne) : $X \in \mathbb{R}^{m \times d}$

Fonction de perte : $J(\theta) = \frac{1}{2m} \sum_{j=1}^{m} (\theta^t x^j - y^j)^2$

On peut écrire cette somme sous forme d'un produit de matrices :

$$J(\theta) = \frac{1}{2m} (\underbrace{X}_{m \times d} \underbrace{\vec{\theta}}_{d \times 1} - \underbrace{\vec{y}}_{m \times 1})^t (X\vec{\theta} - \vec{y})$$

$X\theta$ est de dimension $m \times 1$
et $(X\theta - y)^t (X\theta - y)$ est bien un scalaire.

Développons J en laissant tomber la constante $\frac{1}{2m}$:

$$J(\theta) = (X\theta - y)^t (X\theta - y) = ((X\theta)^t - y^t)(X\theta - y)$$
$$= (X\theta)^t X\theta - (X\theta)^t y - y^t (X\theta) + y^t y$$

on a $(X\theta)^t y = y^t (X\theta)$ car c'est un scalaire.

d'où $J(\theta) = \theta^t X^t X \theta - 2(X\theta)^t y + y^t y$

On dérive par rapport à θ :

$$\frac{dJ(\theta)}{d\theta} = 2X^t X \theta - 2X^t y + 0 = 0$$
$$\Rightarrow \cancel{2} X^t X \theta = \cancel{2} X^t y$$
$$\Rightarrow \boxed{\theta = (X^t X)^{-1} X^t y}$$

À vous de vous convaincre que $\frac{d}{d\theta}(\theta^t X^t X \theta) = 2X^t X \theta$

et $\frac{d}{d\theta}(X\theta)^t = \frac{d}{d\theta}(\theta^t X^t) = X^t$ ...