Jérémie Béliveau-Lefebvre – 04 494 470
Sébastien Leblanc – 18 206 273

TP2 - solution

1. La régression linéaire est donné par la formule suivante :

$$J(w) = \frac{1}{2}\sum_{n=1}^{N}(w^T\phi(x_n) - t_n)^2 + \frac{\lambda}{2}w^Tw$$

Nous voulons prouver avec l'aide de la représentation duale, que la solution de type maximum à postériori de cette équation est la suivante :

$$y(x) = w^T\phi(x) = a^T\Phi\phi(x) = k(x)^T(K + \lambda \mathrm{I}_N)^{-1}t$$

Nous devons d'abord trouver le gardient de $J(\vec{w})$ et le rendre à 0 soit :

$$\frac{\partial J(w)}{\partial w} = 0 \text{ soit } \frac{\partial \frac{1}{2}\sum_{n=1}^{N}(w^T\phi(x_n)-t_n)^2+\frac{\lambda}{2}w^Tw}{\partial w} = 0$$

<=>

$$\frac{1}{2}\sum_{n=1}^{N}(w^T\phi(x_n) - t_n)2\phi(x_n) + \lambda w = 0 \qquad \text{par dérivation}$$

<=>

$$\sum_{n=1}^{N}(w^T\phi(x_n) - t_n)\phi(x_n) + \lambda w = 0 \qquad \text{arithmétique}$$

<=>

$$\lambda w = -\sum_{n=1}^{N}(w^T\phi(x_n) - t_n)\phi(x_n) \qquad \text{arithmétique}$$

<=>

$$w = -\frac{1}{\lambda}\sum_{n=1}^{N}(w^T\phi(x_n) - t_n)\phi(x_n) \qquad \text{arithmétique}$$

Nous pouvons combiner des termes de cet expressions comme suit :

$$w = -\frac{1}{\lambda}\sum_{n=1}^{N}(w^T\phi(x_n) - t_n)\phi(x_n) = \sum_{n=1}^{N} a_n\phi(x_n) = \Phi^Ta$$

Où

$$a = -\frac{1}{\lambda}\{w^T\phi(x_n) - t_n\}$$

Au lieu de travailler avec le paramètre w, nous pouvons travailler avec le paramètre $w = \Phi^Ta$, et le susbtituer dans J(w), ce qui donnerait $J(\Phi^Ta)$ et que nous nommeront J(a) et se décrit comme ceci :

$$J(a) = \frac{1}{2}\sum_{n=1}^{N}((\Phi^Ta)^T\phi(x_n) - t_n)^2 + \frac{\lambda}{2}(\Phi^Ta)^T(\Phi^Ta) \qquad \text{par définition}$$

<=>

$$J(a) = \frac{1}{2}((\Phi^Ta)^T\Phi - t)^2 + \frac{\lambda}{2}(\Phi^Ta)^T\Phi^Ta$$

par changement matriciel

<=>

$$J(a) = \frac{1}{2}(a^T\Phi\Phi - t)^2 + \frac{\lambda}{2}a^T\Phi\Phi^Ta \qquad \text{arythmétique}$$

<=>

$$J(a) = \frac{1}{2}(a^T\Phi\Phi^T\Phi\Phi^Ta - 2a^T\Phi\Phi^Tt + t^Tt) + \frac{\lambda}{2}a^T\Phi\Phi^Ta \qquad \text{arythmétique}$$

<=>

$$J(a) = \frac{1}{2}a^T\Phi\Phi^T\Phi\Phi^Ta - a^T\Phi\Phi^Tt + \frac{1}{2}t^Tt + \frac{\lambda}{2}a^T\Phi\Phi^Ta \qquad \text{distributivité}$$

Où $t = \{t_1, t_2, \dots, t_n\}^T$
Si nous définissons la matrice de Gram comme suit : $K = \Phi\Phi^T$, laquelle est une matrice NxN avec comme éléments :

$$K_{nm} = \phi(x_n)^T \phi(x_m)$$

La formule suivante devient :

$$J(a) = \frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a$$

Pour trouver les maximum et minimum, nous devons mettre le gradient J(a) à 0, ce qui donne :

$$\frac{\partial J(a)}{\partial a} = \frac{\partial\left(\frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a\right)}{\partial a} = 0$$
définition

<=>

$$a^T K K - K t + \lambda a^T K = 0$$
par dérivation

<=>

$$a^T K - t + \lambda a^T = 0$$
arythmétique

<=>

$$a^T (K + \lambda I_N) = t$$
arythmétique

<=>

$$a = (K + \lambda I_N)^{-1} t$$
arythmétique

Si nous subtituons cette expression dans le modèle de régression linéaire, nous trouvons que :

$$y(x) = w^T \phi(x) = \Phi^T a \phi(x)$$
puisque $w = \Phi^T a$

<=>

$$a^T \Phi \phi(x)$$
arythmétique

<=>

$$a^T \Phi \phi(x)$$
arythmétique

<=>

$$((K + \lambda I_N)^{-1} t\,)^T \Phi \phi(x)$$
par définition de a

<=>

$$\left(\Phi \phi(x)\right)^T (K + \lambda I_N)^{-1} t$$
arythmétique

Puisque $K_{nm} = \phi(x_n)^T \phi(x_m)$, nous pourrions définir un vecteur tel que : k(x) avec n élément tel que $k_n(x) = k(x_n, x)$. Nous aurons donc l'expression suivante et finale :

$$y(x) = k(x)(K + \lambda I_N)^{-1} t$$

CQFD

2. Nous devons prouver que l'expression suivante est un noyau valide, sachant que $k_a\left(x_a, x'_a\right)$ et $k_b(x_b, x'_b)$ sont des noyaux valides :

$$k_a\left(x_a, x'_a\right) + k_b\left(x_b, x'_b\right) = k\ (x, x')$$

$$k_a\left(x_a, x'_a\right) + k_b\left(x_b, x'_b\right)$$

<=>

$$\phi(x_a)^T \phi\left(x'_a\right) + \phi(x_b)^T \phi\left(x'_b\right) \qquad \text{par définition}$$

<=>

$$[\phi(x_a) \quad \phi(x_b)] \begin{bmatrix} \phi\left(x'_a\right) \\ \phi\left(x'_b\right) \end{bmatrix} \qquad \text{par déduction matricielle}$$

<=>

$$\phi(\{x_a, x_b\})^T \phi\left(\{x'_a, x'_b\}\right) \qquad \text{par définition}$$

<=>

$$k\ \left(\{x_a, x_b\}, \{x'_a, x'_b\}\right) \qquad \text{par définition}$$

<=>

$$k\ (x, x') \qquad \text{puisque } x = \{x_a, x_b\}$$

Ainsi, $k\ (x, x')$ est un noyau valide

CQFD