

TP2 - solution

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

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

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

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

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

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

\Leftrightarrow

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

\Leftrightarrow

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

\Leftrightarrow

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

\Leftrightarrow

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

Nous pouvons combiner des termes de cet expressions comme suit :

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

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 $w = \Phi^T a$, et le substituer dans $J(w)$, ce qui donnerait $J(\Phi^T a)$ et que nous nommeront $J(a)$ et se décrit comme ceci :

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

\Leftrightarrow

$$J(a) = \frac{1}{2} ((\Phi^T a)^T \Phi - t)^2 + \frac{\lambda}{2} (\Phi^T a)^T \Phi^T a \quad \text{par changement matriciel}$$

\Leftrightarrow

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

\Leftrightarrow

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

\Leftrightarrow

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

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 $N \times N$ 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 :

$$\begin{aligned} \frac{\partial J(a)}{\partial a} &= \frac{\partial \left(\frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a \right)}{\partial a} = 0 && \text{définition} \\ \Leftrightarrow & a^T K K - K t + \lambda a^T K = 0 && \text{par dérivation} \\ \Leftrightarrow & a^T K - t + \lambda a^T = 0 && \text{arithmétique} \\ \Leftrightarrow & a^T (K + \lambda I_N) = t && \text{arithmétique} \\ \Leftrightarrow & a = (K + \lambda I_N)^{-1} t && \text{arithmétique} \end{aligned}$$

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

$$\begin{aligned} y(x) &= w^T \Phi(x) = \Phi^T a \Phi(x) && \text{puisque } w = \Phi^T a \\ \Leftrightarrow & a^T \Phi \Phi(x) && \text{arithmétique} \\ \Leftrightarrow & a^T \Phi \Phi(x) && \text{arithmétique} \\ \Leftrightarrow & ((K + \lambda I_N)^{-1} t)^T \Phi \Phi(x) && \text{par définition de } a \\ \Leftrightarrow & (\Phi \Phi(x))^T (K + \lambda I_N)^{-1} t && \text{arithmétique} \end{aligned}$$

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 $k_a(x_a, x'_a)$ et $k_b(x_b, x'_b)$ sont des noyaux valides :

$$\begin{aligned} & k_a(x_a, x'_a) + k_b(x_b, x'_b) = k(x, x') \\ \Leftrightarrow & k_a(x_a, x'_a) + k_b(x_b, x'_b) \end{aligned}$$

$a(k(x, x')) + b(k(x, x'))$ par concaténation,
les noyaux k_a et k_b sont des portions du noyau k

\Leftrightarrow

$(a + b)(k(x, x'))$ par associativité

\Leftrightarrow

$k(x, x')$ puisque $k(x, x') = ck(x, x')$

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

CQFD