TP3

Thibault Duhamel 18026048, Heng Shi 18171434

March 14, 2019

Question 1

Given a set of vectors $\{x_n\}_{n\in[1...N]}$ and a base function ϕ, we define the matrix Φ whose rows are $\phi(x_n)^T$. Let us define the following loss function, with a positive regularization parameter λ:

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

As we want to minimize this function, it is common to compute its gradient with respect to w:

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

Then, we set it to zero to extract the value at an extrem point, which gives:

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

If we create a vector of variables $a = (a_1, ..., a_N)$, it is possible to write:

$$w = \sum_{n=1}^{N}a_n\phi(x_n) = \Phi^Ta$$

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

Now, let us replace w by Φ^Ta in the loss function:

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

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

$$= \frac{1}{2}\sum_{n=1}^{N}\Big[a^T\Phi\phi(x_n)a^T\Phi\phi(x_n) - 2t_na^T\Phi\phi(x_n) + t_n^2\Big] + \frac{\lambda}{2}a^T\Phi\Phi^Ta$$

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

Using $t = (t_1, ...t_N)$, and matching corresponding dimensions between matrices and vectors:

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

This expression can be simplified by introducing $K = \Phi\Phi^T$, and defining a kernel $k(x_n, x_m) = K_{nm} = \phi(x_n)^T\phi(x_m)$:

$$= \frac{1}{2}a^TKKa - a^TKt + \frac{1}{2}t^Tt + \frac{\lambda}{2}a^TKa$$

Again, let us set compute the gradient of the loss function J, with respect to a this time:

$$\frac{dJ}{da} = KKa - Kt + \lambda Ka$$

Setting this gradient to zero, we get:

$$KKa - Kt + \lambda Ka = 0$$

As $K = \Phi\Phi^T$, this matrix is invertible:

$$Ka - t + \lambda a = 0$$

$$Ka + \lambda a = t$$

$$(K + \lambda I_N)a = t$$

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

Eventually, reporting this value in the model function y:

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

with $k(x)$ is the vector built with the kernel function of x and every other point in the set, such as $k(x) = (k(x_1, x), ..., k(x_N, x))$.

Question 2

Let x be the concatenation of two sub-vectors x_a and x_b, such as $x = (x_a, x_b)$. It is known that both $k_a(x_a, x'_a)$ and $k_b(x_b, x'_b)$ are two valid kernels over respective dimensions. So, there exist $\phi_.(.)$ functions such as:

$$k_a(x_a, x'_a) = \phi_a(x_a)^T \phi_a(x'_a)$$
$$k_b(x_b, x'_b) = \phi_b(x_b)^T \phi_b(x'_b)$$

Then, let us write down the expression of $k(x, x')$:

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

$$= \phi_a(x_a)^T \phi_a(x'_a) + \phi_b(x_b)^T \phi_b(x'_b)$$

$$= (\phi_a(x_a), \phi_b(x_b))^T (\phi_a(x'_a), \phi_b(x'_b))$$

where $(\phi_a(.), \phi_b(.))$ is the concatenation of $\phi_a(.)$ and $\phi_b(.)$.

Hence, as the first term only depends on x and the second one only depends on x', it is possible to write:

$$k(x, x') = \phi(x)^T \phi(x')$$

where $\phi(x) = (\phi_a(x_a), \phi_b(x_b))$

$k(x, x')$ thus is a valid kernel.