Machine Learning: Homework Assignment 7

E4525 Spring 2018,

IEOR, Columbia University

Due: May 3rd, 2019

1. **Separating Hyperplane for two points**

   Assume we have two points x_+ and x_- both in $\mathcal{R}^D$. We have that the class of x_+ is $y_+ = +1$ and the class of x_- is $y_- = -1$.

   Give an explicit expression for the parameters w and b of the separating hyper-plane with maximum margin. Set the normalization such that $w^T x_+ + b = 1$.

2. **Dual coefficients and Margin**

   Prove that, for the hard margin case, the optimal margin ρ satisfies

   $$\frac{1}{\rho^2} = \sum_i \alpha_i \tag{1}$$

3. **Kernel Construction and Mercer's Condition**

   Using Mercer's Condition that a kernel $K(x, x')$ generates a scalar product if, and only if, the $N \times N$ matrix,

   $$K_{i,j} = K(x_i, x_j) \tag{2}$$

   is positive semi-definite for all x_i, x_j in $(x_1, \dots x_N) \in \mathcal{X}$.

   And that a matrix is positive semi-definite if and only if

   $$c^T K c \geq 0 \tag{3}$$

   for all vectors $c \in \mathbb{R}^N$.

   Show that

   (a) The sum of two kernels is a kernel

   $$K(x, x') = k_1(x, x') + k_2(x, x') \tag{4}$$

   (b) The product of a kernel k_1 and a scalar $0 \leq \lambda \in R$ is a kernel

   $$K(x, x') = \lambda k_1(x, x') \tag{5}$$

(c) The scaling of a kernel k by a function is a kernel

$$K(x, x') = f(x)k_1(x, x')f(x') \tag{6}$$

4. **Product of two Kernels and Tensor Product**

   Let's assume that we have two kernels k^1 and k^2 that can be written as the scalar product in feature spaces ϕ^1 and ϕ^2 of dimension D_1 and D_2:

$$\begin{aligned} k^1(x, x') &= \sum_i^{D_1} \phi_i^1(x)\phi_i^1(x') \\ k^2(x, x') &= \sum_j^{D_2} \phi_j^2(x)\phi_j^1(x') \end{aligned} \tag{7}$$

   Show that the product kernel

$$K(x, x') = k^1(x, x')k^2(x, x') \tag{8}$$

   can be represented by the *tensor product* feature space ϕ

$$\phi_k(x) = \phi_{i+D_1 j}(x) = \phi_i^1(x)\phi_j^2(x) \tag{9}$$

   for $k = 1, \ldots, D_1 D_2$, and

$$K(x, x') = \sum_{k=1}^{D_1 D_2} \phi_k(x)\phi_k(x') \tag{10}$$

5. **Off Diagonal terms of Gram Matrix**

   (a) Show that the diagonal terms of a kernel $k(x, x)$ must be positive or zero.

   (b) Using the fact that the Gram matrix of two points x and x' must be positive semi-definite show that

$$k(x, x') \leq \sqrt{k(x, x)k(x', x')} \tag{11}$$

   (c) Give an example of a symmetric 2×2 matrix where all entries are positive and yet the matrix is not posite semi-definite.

   (d) Give an example of a symmetric 2×2 matrix where some elements are negative but the matrix is still positive semi-definite.