Set 4. Due March 11, 2016

Problem 17 Let $(x_1, y_1), \ldots, (x_n, y_n)$ be data in $\mathbb{R}^d \times \{-1, 1\}$. Suppose that the data are *linearly separable*, that is, there exists a $w \in \mathbb{R}^d$ such that $y_i w^T x_i > 0$ for all $i = 1, \ldots, n$. The *margin* of such a vector is

$$\gamma(w) = \min_{i=1,\ldots,n} \frac{y_i w^T x_i}{\|w\|} .$$

Formulate a *convex optimization problem* whose solution is a vector w^* that classifies the data correctly (i.e., $y_i {w^*}^T x_i > 0$ for all $i = 1, \ldots, n$) and maximizes the margin. Show that the optimal solution w^* lies in the vector space spanned by the examples x_i for which the margin $\frac{y_i {w^*}^T x_i}{\|w^*\|}$ is minimal among all examples.

Problem 18 Suppose that data $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^d \times \{-1, 1\}$ are such that each $x_i \in \{0, 1\}^d$ (i.e., the x_i have binary components) and for each $i = 1, \ldots, n$, $y_i = 1$ if and only if at least one component of x_i is 1. Show that the data are linearly separable. What is largest achievable margin (i.e., the smallest distance of all data points to the separating hyperplane) in the worst case?

Repeat the exercise in the case when for each $i = 1, \ldots, n$, $y_i = 1$ if and only if the sum of the components of x_i is at least $d/2$ (assume here that d is odd).

Problem 19 Let $\mathcal{H}$ be the Hilbert space of all sequences $s = \{s_n\}_{n=0}^{\infty}$ satisfying $\sum_{n=0}^{\infty} s_n^2 < \infty$ with inner product $\langle s, t \rangle = \sum_{n=0}^{\infty} s_n t_n$. Consider the feature map $\Phi : \mathbb{R} \to \mathcal{H}$ that assigns, to each real number x, the sequence $\Phi(x)$ whose n-th element equals

$$(\Phi(x))_n = \frac{1}{\sqrt{n!}} x^n e^{-x^2/2} , \quad n = 0, 1, 2, \ldots .$$

Determine the kernel function $K(x, y) = \langle \Phi(x), \Phi(y) \rangle$ for $x, y \in \mathbb{R}$. (You may use the fact that $\sum_{n=0}^{\infty} x^n/n! = e^x$.)

Can you generalize the kernel so that it is defined on $\mathbb{R}^d \times \mathbb{R}^d$ instead of $\mathbb{R} \times \mathbb{R}$? What is the corresponding feature map?

Problem 20 Let $K_1, K_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be kernels. Prove that $K_1 + K_2$ and $K_1 K_2$ are also kernels.

Problem 21 Let $\mathcal{X}_n = \{0, 1\}^n$ be the set of binary strings of length n. Let $m < n$. We say that $s \in \{0, 1\}^m$ is a *substring* of $x = (x_1, \ldots, x_n) \in \mathcal{X}_n$ if for some $i \in \{1, \ldots, n - m + 1\}$, $s = (x_i, \ldots, x_{i+m-1})$.

Define the function $K : \mathcal{X}_n \times \mathcal{X}_n \to \mathbb{R}$ as the number of common substrings of its aguments, that is,

$$K(x, y) = \sum_{s \in \{0,1\}^m} \mathbb{1}_{\{s \text{ is a substring of both } x \text{ and } y\}} \qquad \text{for } x, y \in \mathcal{X}_n.$$

Prove that K is a kernel function. Determine a feature map Φ defined of $\mathcal{X}_n$, mapping to some Hilbert space $\mathcal{H}$ for which $K(x, y)$ is the inner product of $\Phi(x)$ and $\Phi(y)$. What is the dimension of $\mathcal{H}$?