1. Consider a dataset with 3 points in $\mathbb{R}$,

<table>
<thead>
<tr>
<th>Class $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$+$ 0</td>
</tr>
<tr>
<td>$-$ +1</td>
</tr>
<tr>
<td>$-$ −1</td>
</tr>
</tbody>
</table>

(a) Are the classes $\{+, −\}$ linearly separable?

(b) Consider mapping each point to $\mathbb{R}^3$ using new feature vectors $\phi(x) = [1, 2x, x^2]^T$. Are the classes now linearly separable? If so, find a separating hyperplane.

(c) Define a class variable $y \in \{+1, −1\}$ which denotes the class of $x$ and let $w = [w_1, w_2, w_3]^T$.

The maximum margin SVM classifier solves the following problem,

$$
\begin{align*}
\min_{w,b} & \quad \frac{1}{2}||w||^2 \\
\text{s.t.} & \quad y_i(w^T \phi(x_i) + b) \geq 1, \quad i = 1, 2, 3.
\end{align*}
$$

Using the method of Lagrange multipliers find the solution is $w, b$ and calculate the maximum margin.

(d) Show that the solution remains the same if the constraints are changed to $y_i(w^T \phi(x_i) + b) \geq \rho$, $i = 1, 2, 3$, for any $\rho \geq 1$.

2. Expressiveness of kernels

(a) Construct a support vector machine that computes the $XOR$ function. Use values of $+1$ and $−1$ (instead of 1 and 0) for both inputs and outputs, so that an example looks like $\left([-1, 1], 1\right)$ or $\left([-1, -1], -1\right)$. Map the input $[x_1, x_2]$ into a space consisting of $x_1$ and $x_1x_2$. Draw the four input points in this space, and the maximal margin separator. What is the margin? Now draw the separating line back in the original Euclidean input space.

(b) Recall that the equation of the circle in $\mathbb{R}^2$ is $(x_1 - a)^2 + (x_2 - b)^2 = r^2$. Expand out the formula and show that every circular region is linearly separable from the rest of the plane in the feature space $(x_1, x_2, x_1^2, x_1x_2)$.

(c) Recall that the equation of an ellipse in $\mathbb{R}^2$ is \( \frac{(x_1 - a)^2}{r_1^2} + \frac{(x_2 - b)^2}{r_2^2} = 1 \). Show that an SVM using the polynomial kernel of degree 2, $k(x, y) = (1 + x^T y)^2$, is equivalent to a linear SVM in the feature space $(1, x_1, x_2, x_1^2, x_1x_2)$ and hence that SVMs with this kernel can separate any elliptic region from the rest of the plane.

3. Let $k_1, k_2$ be kernels over $\mathbb{R}^n \times \mathbb{R}^n$, and $k_3$ be a kernel over $\mathbb{R}^d \times \mathbb{R}^d$. Let $a \in \mathbb{R}_+$ be a positive real number, $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function, $\phi : \mathbb{R}^n \to \mathbb{R}^d$ be a function mapping from $\mathbb{R}^n$ to $\mathbb{R}^d$, and let $p$ be a polynomial over $\mathbb{R}_+$ with positive coefficients. For each of the functions $k$ below, state whether it is necessarily a kernel. If you think it is, prove it; if you think it isn’t, give a counter-example.

(a) $k(x, y) = k_1(x, y) + k_2(x, y)$

(b) $k(x, y) = k_1(x, y) − k_2(x, y)$

(c) $k(x, y) = a k_1(x, y)$

(d) $k(x, y) = − a k_1(x, y)$

(e) $k(x, y) = k_1(x, y) k_2(x, y)$

(f) $k(x, y) = f(x) f(y)$
4. Recall the Representer theorem stated in class. Let $k : X \times X \to \mathbb{R}$ be a PDS kernel and $\mathcal{H}$ its corresponding RKHS. Then, for any non-decreasing function $G : \mathbb{R} \to \mathbb{R}$ and any loss function $L : \mathbb{R}^m \to \mathbb{R} \cup \{\pm \infty\}$, the optimization problem

$$
\arg \min_{h \in \mathcal{H}} G(\|h\|_{\mathcal{H}}) + L(h(x_1), \ldots, h(x_m))
$$

admits a solution of the form $h^* = \sum_{i=1}^m \alpha_i k(x_i, \cdot)$.

(a) Now rewrite equation (1) by using $G(x) = x^2$ and $L(h(x_1), \ldots, h(x_m)) = \sum_{i=1}^m (y_i - h(x_i))^2$, where $y_i \in \mathbb{R}$ for all $i = 1, \ldots, m$ and solve for $h^*$.

(b) Express your solution in the form $h^* = \sum_{i=1}^m \alpha_i k(x_i, \cdot)$, or equivalently find $\alpha_i, i = 1, \ldots, m$.

(c) Is this solution unique? Give your reasons.

5. **Explicit polynomial kernel mapping.** Let $K$ be a polynomial kernel of degree $d$, i.e., $K : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$, $K(x, x') = (x^T x' + c)^d$, with $c > 0$. Show that the dimension of the feature space associated to $K$ is

$$\left(\begin{array}{c}
N + d \\
N - d
\end{array}\right).$$

6. Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a PDS kernel. For any $x \in \mathcal{X}$, define $\Phi(x) : \mathcal{X} \to \mathbb{R}$ for all $x' \in \mathcal{X}$ as follows:

$$\Phi(x)(x') = K(x, x').$$

We define $\mathbb{H}_0$ as the set of finite linear combinations of such functions $\Phi(x)$:

$$\mathbb{H}_0 = \left\{ \sum_{i \in I} a_i \Phi(x_i) : a_i \in \mathbb{R}, x_i \in \mathcal{X}, |I| < \infty \right\}.$$

Now, we introduce the operation $\langle , \rangle$ on $\mathbb{H}_0 \times \mathbb{H}_0$ defined for all $f, g \in \mathbb{H}_0$ with $f = \sum_{i \in I} a_i \Phi(x_i)$ and $g = \sum_{j \in J} b_j \Phi(x_j)$ by

$$\langle f, g \rangle = \sum_{i \in I, j \in J} a_i b_j K(x_i, x_j).$$

Show that for any $f \in \mathbb{H}_0$ and any $x \in \mathcal{X}$,

$$\langle f, \Phi(x) \rangle^2 \leq \langle f, f \rangle \langle \Phi(x), \Phi(x) \rangle.$$

7. **Normalization of kernels.** To any kernel $K$ we associate the normalized kernel $K'$ by

$$
\forall x, x' \in \mathcal{X}, K'(x, x') = \begin{cases} 
0, & \text{if } K(x, x) = 0 \land (K(x', x') = 0) \\
\frac{K(x, x')}{\sqrt{K(x, x)K(x', x')}}, & \text{otherwise.}
\end{cases}
$$

(a) Show that if $K$ is a PDS kernel, then the normalized kernel $K'$ associated to $K$ is also PDS.

(b) Consider the kernel, $K : \langle x, x' \rangle \mapsto \exp \left( \frac{x^T x'}{2\sigma^2} \right)$. Find the normalized kernel associated with $K$.

What is this kernel popularly known as?

8. **Closure properties of kernels.** Prove that PDS kernels are closed under sum, product, tensor product, pointwise limit, and composition with a power series $\sum_{n=0}^{\infty} a_n x^n$ with $a_n \geq 0$ for all $n \in \mathbb{N}$. 

\[ \text{E2 336 Homework 2, Page 2 of 2 Aug 27, 2018} \]