CMSC5724: Exercise List 6

Problem 1. Prove the theorem on Slide 6 of the lecture notes on the kernel method without the interleaving assumption.

Answer: Sort the input P and divide it into *maximal* subsets such that the points in each subset are consecutive and share the same label. Denote the subsets as $S_1, S_2, ..., S_l$ in ascending order (for some $l \geq 1$). For example, suppose P has points $p_1, p_2, ..., p_{10}$ where p_2, p_3, and p_4 have label 1, and the other points label -1. Then, $l = 3$; and $S_1 = \{p_1\}, S_2 = \{p_2, p_3, p_4\}$, and $S_3 = \{p_5, p_6, ..., p_{10}\}$.

W.o.l.g., we will assume that the points in S_1 have label -1 and that l is an odd number. Find a point q_i for each $i \in [1, l-1]$ such that q_i is larger than the points in S_i but smaller than those in S_{i+1}. Construct a function:

$$f(x) = -(x - q_1)(x - q_2)...(x - q_{l-1}). \tag{1}$$

For an odd i, $f(p) < 0$ for all $p \in S_i$. For an even i, $f(p) > 0$ for all $p \in S_i$. The rest of the proof proceeds as discussed in the lecture.

Problem 2. Consider the kernel function $K(p, q) = (\boldsymbol{p} \cdot \boldsymbol{q} + 1)^3$, where $\boldsymbol{p} = (p[1], p[2])$ and $\boldsymbol{q} = (q[1], q[2])$ are 2-dimensional vectors. Recall that there is a mapping function ϕ from $\mathbb{R}^2$ to $\mathbb{R}^d$ for some integer d such that $K(p, q)$ equals the dot product of $\phi(p)$ and $\phi(q)$. Give the details of ϕ.

Answer: Rewrite K as dot product form.

$$\begin{aligned}
K(p,q) &= (p[1]q[1] + p[2]q[2] + 1)^3 \\
&= p[1]^3q[1]^3 + p[2]^3q[2]^3 + 1 + 3p[1]q[1]p[2]^2q[2]^2 \\
&\quad + 3p[1]^2q[1]^2p[2]q[2] + 3p[1]q[1] + 3p[1]^2q[1]^2 + 3p[2]q[2] + 3p[2]^2q[2]^2 + 6p[1]q[1]p[2]q[2] \\
&= (p[1]^3, p[2]^3, 1, \sqrt{3}p[1]p[2]^2, \sqrt{3}p[1]^2p[2], \sqrt{3}p[1], \sqrt{3}p[2], \sqrt{3}p[1]^2, \sqrt{3}p[2]^2, \sqrt{6}p[1]p[2]) \\
&\quad \cdot (q[1]^3, q[2]^3, 1, \sqrt{3}q[1]q[2]^2, \sqrt{3}q[1]^2q[2], \sqrt{3}q[1], \sqrt{3}q[2], \sqrt{3}q[1]^2, \sqrt{3}q[2]^2, \sqrt{6}q[1]q[2])
\end{aligned}$$

Therefore, $\phi(x) = (x[1]^3, x[2]^3, 1, \sqrt{3}x[1]x[2]^2, \sqrt{3}x[1]^2x[2], \sqrt{3}x[1], \sqrt{3}x[2], \sqrt{3}x[1]^2, \sqrt{3}x[2]^2, \sqrt{6}x[1]x[2])$.

Problem 3. Consider a set P of 2D points each labeled either -1 or 1. It is known that the points of the two labels can be linearly separated after applying the Kernel function $K(p, q) = (\boldsymbol{p} \cdot \boldsymbol{q} + 1)^2$. Prove that they can also be linearly separated by applying the kernel function $K'(p, q) = (2\boldsymbol{p} \cdot \boldsymbol{q} + 3)^2$.

Answer: Using the method explained in Problem 1, we can find the mapping functions ϕ and ϕ' for K and K', respectively:

$$\begin{aligned}
\phi(p) &= (p[1]^2, p[2]^2, 1, \sqrt{2}p[1], \sqrt{2}p[2], \sqrt{2}p[1]p[2]) \\
\phi'(p) &= (2p[1]^2, 2p[2]^2, 3, 2\sqrt{3}p[1], 2\sqrt{3}p[2], 2\sqrt{2}p[1]p[2]).
\end{aligned}$$

Let π be the plane that separates the points under ϕ. If $\boldsymbol{w} \cdot \phi(x) = 0$ is the equation for π, then (i) for every point p of label 1, $\boldsymbol{w} \cdot \phi(p) > 0$, and (ii) for every point p of label -1, $\boldsymbol{w} \cdot \phi(p) < 0$.

Set $\boldsymbol{w}' = (\frac{\boldsymbol{w}[1]}{2}, \frac{\boldsymbol{w}[2]}{2}, \frac{\boldsymbol{w}[3]}{3}, \frac{\boldsymbol{w}[4]}{\sqrt{6}}, \frac{\boldsymbol{w}[5]}{\sqrt{6}}, \frac{\boldsymbol{w}[6]}{\sqrt{6}})$. Let π' be the plane given by the equation $\boldsymbol{w}' \cdot \phi'(x) = 0$.

We claim that π' also separates the points. Indeed, for every point p of label 1, we have:

$$\begin{aligned}
&\boldsymbol{w}' \cdot \phi'(p) \\
&= \frac{\boldsymbol{w}[1]}{2} \cdot 2p[1]^2 + \frac{\boldsymbol{w}[2]}{2} \cdot 2p[2]^2 + \frac{\boldsymbol{w}[3]}{3} \cdot 3 + \frac{\boldsymbol{w}[4]}{\sqrt{6}} \cdot 2\sqrt{3}p[1] + \frac{\boldsymbol{w}[5]}{\sqrt{6}} \cdot 2\sqrt{3}p[2] + \frac{\boldsymbol{w}[6]}{\sqrt{6}} \cdot 2\sqrt{3}p[1]p[2] \\
&= \boldsymbol{w}[1] \cdot p[1]^2 + \boldsymbol{w}[2] \cdot p[2]^2 + \boldsymbol{w}[3] + \sqrt{2}\boldsymbol{w}[4] \cdot p[1] + \sqrt{2}\boldsymbol{w}[5] \cdot p[2] + \sqrt{2}\boldsymbol{w}[6] \cdot p[1]p[2] \\
&= \boldsymbol{w} \cdot \phi(p) > 0.
\end{aligned}$$

Likewise, we can prove that, for every point p of label -1, it holds that $\boldsymbol{w}' \cdot \phi'(p) = \boldsymbol{w} \cdot \phi(p) < 0$.

Problem 4. Consider a set P of 2D points that has three label-1 points $p_1(-2,-2), p_2(1,1), p_3(3,3)$, and two label-$(-1)$ points $q_1(-2,2), q_2(2,-2)$. Answer the following questions:

- Use Perceptron to find a separation plane π using the Kernel function $K(x,y) = (x \cdot y + 1)^2$.
- According to π, what is the label of point $(2,2)$?

Answer: Initially, let $\boldsymbol{w}_0 = 0$. Perceptron runs as follows:

Iteration 1. Since $\boldsymbol{w}_0 \cdot \phi(p_1) = 0$, we set $\boldsymbol{w}_1 = \boldsymbol{w}_0 + \phi(p_1) = \phi(p_1)$.

Iteration 2. Since $\boldsymbol{w}_1 \cdot \phi(q_1) = K(p_1, q_1) = 1 > 0$, we set $\boldsymbol{w}_2 = \boldsymbol{w}_1 - \phi(q_1) = \phi(p_1) - \phi(q_1)$.

Iteration 3. There are no more violations for $\boldsymbol{w}_2$. So we have found a separation plane $\boldsymbol{w}_2 \cdot \phi(x) = 0$ such that (i) $\boldsymbol{w}_2 \cdot \phi(x) > 0$ for every label-1 point p, and (ii) $\boldsymbol{w}_2 \cdot \phi(x) < 0$ for every label-(-1) point p.

Now consider the point $r = (2,2)$. As $\boldsymbol{w}_2 \cdot \phi(r) = K(p_1, r) - K(q_1, r) = 48 > 0$, we classify r as label 1.

Problem 5. Same settings as in Problem 3. Calculate the distance from $\phi(p_1)$ to the separation plane you find in the feature space.

Answer: We know from the solution of Problem 3 that the weight vector of the separation plane (in the feature space) is $\boldsymbol{w} = \phi(p_1) - \phi(q_1)$.

The distance from $\phi(p_1)$ to this plane equals

$$\begin{aligned}
\frac{\boldsymbol{w} \cdot \phi(p_1)}{|\boldsymbol{w}|} &= \frac{\boldsymbol{w} \cdot \phi(p_1)}{\sqrt{\boldsymbol{w} \cdot \boldsymbol{w}}} \\
&= \frac{(\phi(p_1) - \phi(q_1)) \cdot \phi(p_1)}{\sqrt{(\phi(p_1) - \phi(q_1)) \cdot (\phi(p_1) - \phi(q_1))}} \\
&= \frac{\phi(p_1) \cdot \phi(p_1) - \phi(p_1) \cdot \phi(q_1)}{\sqrt{\phi(p_1) \cdot \phi(p_1) - 2\phi(p_1) \cdot \phi(q_1) + \phi(q_1) \cdot \phi(q_1)}} \\
&= \frac{K(p_1, p_1) - K(p_1, q_1)}{\sqrt{K(p_1, p_1) - 2K(p_1, q_1) + K(q_1, q_1)}} \\
&= \frac{81 - 1}{\sqrt{81 - 2 \times 1 + 81}} \\
&= 80/\sqrt{160}.
\end{aligned}$$

Problem 6. Let P be a set of points in $\mathbb{R}^d$. Prove: the Gaussian kernel produces a kernel space where every point $p \in P$ is mapped to a point $\phi(p)$ satisfying $|\phi(p)| = 1$, namely, $\phi(p)$ is on the surface of an infinite-dimensional sphere.

Answer: A Gaussian kernel has the form $K(p, q) = \exp(-\frac{dist(p,q)^2}{2\sigma^2})$ where p and q are points in $\mathbb{R}^d$. in the kernel space, The distance of $\phi(p)$ to the origin is $\sqrt{\phi(p) \cdot \phi(p)}$, which equals

$$\sqrt{K(p, p)} = \sqrt{\exp(-\frac{dist(p, p)^2}{2\sigma^2})} = \sqrt{\exp(0)} = 1.$$

Problem 7. For any a d-dimensional sphere centered at the origin of $\mathbb{R}^d$, we know that any set of $d + 1$ points on the sphere's surface can be shattered by the set of linear classifiers. Use this fact to prove that any finite set P of points in $\mathbb{R}^d$ can be linearly separated in the kernel space produced by the Gaussian kernel. (Hint: use the conclusion of Problem 6 and use the fact that the Gaussian kernel produces a kernel space of infinite dimensionality.)

Answer: By the given fact that any $d + 1$ points on a sphere's surface can be shattered, we know:

> **Fact 1:** For any d-dimensional sphere centered at the origin of $\mathbb{R}^d$ and any set S of n points on the sphere such that $d \geq n - 1$, S can be shattered by the set of d-dimensional linear classifiers.

By the conclusion of Problem 6, every point $p \in P$ is mapped into a point $\phi(p)$ on the surface of an infinite-dimensional sphere centering at the origin. The claim in Problem 7 then follows directly from Fact 1 and Problem 6.