CMSC5724: Quiz 3

Hand-write all your solutions on paper. Take a picture of the paper **together with** your CUHK student ID card. Upload the picture to Blackboard or email it to the instructor at *taoyf@cse.cuhk.edu.hk*. Your must do so within 20 minutes after the quiz has started.

Problem 1 (40%). Consider the kernel function $K(p, q) = (2(p \cdot q) + 1)^2$, where $p = (p[1], p[2])$ and $q = (q[1], q[2])$ are 2D 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.

$$\begin{aligned} K(p, q) &= (2p[1]q[1] + 2p[2]q[2] + 1)^2 \\ &= 4p[1]^2q[1]^2 + 4p[2]^2q[2]^2 + 8p[1]p[2]q[1]q[2] + 4p[1]q[1] + 4p[2]q[2] + 1. \end{aligned}$$

Hence, $\phi(p) = (2p[1]^2, 2p[2]^2, 2\sqrt{2}p[1]p[2], 2p[1], 2p[2], 1)$.

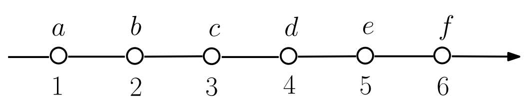
Problem 2 (10%). Consider a 3-class linear classifier in 2D space that is defined by vectors $\boldsymbol{w}_1 = (3, 5)$, $\boldsymbol{w}_2 = (-2, 9)$, and $\boldsymbol{w}_3 = (0, 7)$. Given a point $\boldsymbol{p} = (-5, 1)$, explain what is the label assigned to $\boldsymbol{p}$ and why.

Answer: Computing the dot product between each $\boldsymbol{w_i}$ and $\boldsymbol{p}$ where $i \in [1, 3]$, we have:

- $\boldsymbol{w}_1 \cdot p = -10$;
- $\boldsymbol{w}_2 \cdot p = 19$;
- $\boldsymbol{w}_3 \cdot p = 7$.

Since $\boldsymbol{w}_2 \cdot p$ is largest, the label assigned to $\boldsymbol{p}$ is 2.

Problem 3 (50%). In the lecture, we proved that the k-center algorithm is 2-approximate. In this problem, you will see that the approximation ratio 2 is tight. Consider the k-center problem on the following set P of one-dimensional points (the numbers indicate coordinates):



Answer the following questions for $k = 2$:

1. What is the optimal set C^* of centroids? What is the radius of C^* (namely, $r(C^*)$, using the notations in the lecture notes)?
2. Prove: the k-center algorithm always returns a centroid set whose radius is $2 \cdot r(C^*)$.

Answer: 1. $C^* = \{b, e\}$ and $r(C^*) = 1$.
2: Let $C = \{o_1, o_2\}$ be the set returned by the k-center algorithm. Assume that o_1 (or o_2, resp.) is the first (or the second, resp.) point added into C.

When $o_1 \in \{a, b, c\}$, o_2 must be f. We have $r(C) = 2$.

When $o_1 \in \{d, e, f\}$, o_2 must be a. We also have $r(C) = 2$.

Therefore, the radius of the centroid set returned by the k-center algorithm is always $2 \cdot r(C^*)$.