CMSC5724: Quiz 3

Name: **Student ID:**

Problem 1 (30%). Given 2D points $p = (p[1], p[2])$ and $q = (q[1], q[2])$, define

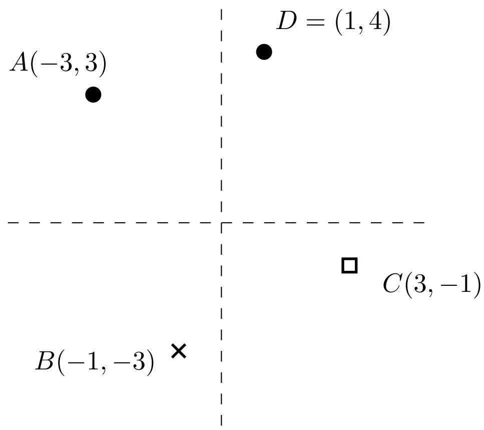
$$K(p, q) = 3 + 4p[2]q[2] + 2(p[1])^2(q[1])^2 + 2(p[1]p[2])^2(q[1]q[2])^2 + 8(p[1]q[1])^3(p[2]q[2])^2.$$

Prove: $K(p, q)$ is a kernel function. Specifically, you need to show a mapping function $\phi : \mathbb{R}^2 \to \mathbb{R}^d$ for some integer d such that $K(p, q)$ equals the dot product of $\phi(p)$ and $\phi(q)$.

Answer:

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

Problem 2 (30%). Consider a training set P including the points below



where the two dots have label 1, the box has label 2, and the cross has label 3. We have a 3-class linear classifier defined by vectors $\boldsymbol{w}_1 = (-1, 3)$, $\boldsymbol{w}_2 = (3, 0)$, and $\boldsymbol{w}_3 = (0, -1)$ (note that this classifier separates P). Calculate the margin of the classifier.

Answer: Let $W = \{w_1, w_2, w_3\}$.

$$\begin{aligned} margin(A \mid W) &= \min\{\frac{\boldsymbol{w_1} \cdot \boldsymbol{A} - \boldsymbol{w_2} \cdot \boldsymbol{A}}{\sqrt{2 \times \sum_{i=1}^{3} |\boldsymbol{w_i}|^2}}, \frac{\boldsymbol{w_1} \cdot \boldsymbol{A} - \boldsymbol{w_3} \cdot \boldsymbol{A}}{\sqrt{2 \times \sum_{i=1}^{3} |\boldsymbol{w_i}|^2}}\} = \min\{\frac{12 - (-9)}{\sqrt{40}}, \frac{12 - (-3)}{\sqrt{40}}\} \\ &= \frac{15}{\sqrt{40}} \end{aligned}$$

Similarly,

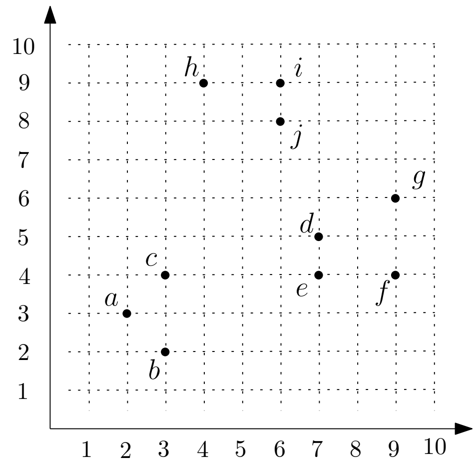
$$margin(B \mid W) = \min\{\frac{3 - (-8)}{\sqrt{40}}, \frac{3 - (-3)}{\sqrt{40}}\} = \frac{6}{\sqrt{40}}$$

$$margin(C \mid W) = \min\{\frac{9 - (-6)}{\sqrt{40}}, \frac{9 - 1}{\sqrt{40}}\} = \frac{8}{\sqrt{40}}$$

$$margin(D \mid W) = \min\{\frac{11 - 3}{\sqrt{40}}, \frac{11 - (-4)}{\sqrt{40}}\} = \frac{8}{\sqrt{40}}$$

Therefore, the classifier's margin equals $\frac{6}{\sqrt{40}}$.

Problem 3 (40%). Consider the set P of points below:



(i) Run the k-center algorithm on P under Euclidean distance. Suppose that $k = 3$ and the first center chosen is a. Explain the second and third centers found by the algorithm.

(ii) Run the k-means algorithm on P with $k = 3$ under Euclidean distance, assuming that the algorithm selects a set $S = \{a, d, g\}$ as the initial centroids. Recall that the algorithm updates S iteratively. Give the content of S after each iteration until the algorithm terminates.

Answer: (i) The second center is g and the third is h.

(ii) *Iteration 1.* Let $o_1 = a, o_2 = d$, and $o_3 = g$. The algorithm divides P into partitions P_1, P_2 and P_3 such that P_i $(1 \leq i \leq 3)$ includes all the points in P with o_i as their closest centroids. Specifically, $P_1 = \{a, b, c\}$, $P_2 = \{d, e, h, i, j\}$, and $P_3 = \{f, g\}$. Then, the algorithm resets o_i to the geometric centroid of P_i: $o_1 = (\frac{8}{3}, 3)$, $o_2 = (6, 7)$, and $o_3 = (9, 5)$.

Iteration 2. The algorithm re-divides P into P_1, P_2 and P_3 based on the current centroids: $P_1 = \{a, b, c\}$, $P_2 = \{h, i, j\}$, and $P_3 = \{d, e, f, g\}$. Accordingly, the centroids are re-computed as $o_1 = (\frac{8}{3}, 3)$, $o_2 = (\frac{16}{3}, \frac{26}{3})$, and $o_3 = (8, \frac{19}{4})$.

Iteration 3. We get $P_1 = \{a, b, c\}$, $P_2 = \{h, i, j\}$, and $P_3 = \{d, e, f, g\}$ again after re-dividing P based on the current centroids. The algorithm terminates.