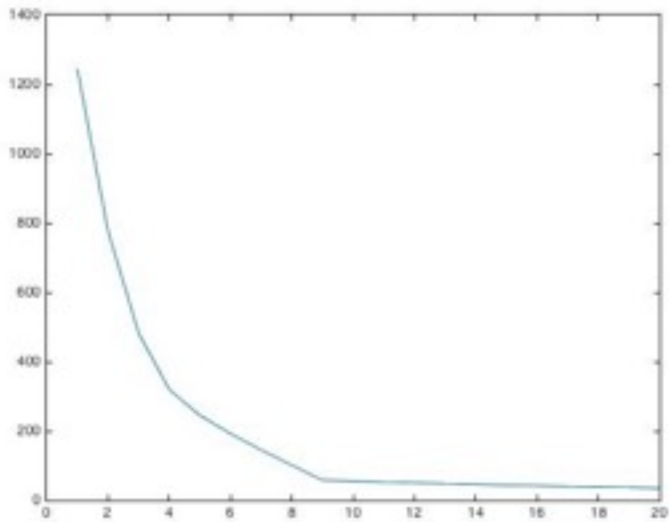
10601 Machine Learning HW5
Hsueh Lin Huang
hsuehlih

Problem1: k-means Clustering

(a) There are total k^n partitionings. Compute the optimal center of each sets and their corresponding objective values. The algorithm's running time is exponential in the number of data points.

Experiment 1: The effect of k on Lloyd's method

(f)



(g) Choose k = 9. The results agrees with the intuitive clusters of 2D dataset.

Experiment 2: The effect of initialization on Lloyd's method

(h)
obj_kmeans++ = 77.1705
obj_random = 88.6968

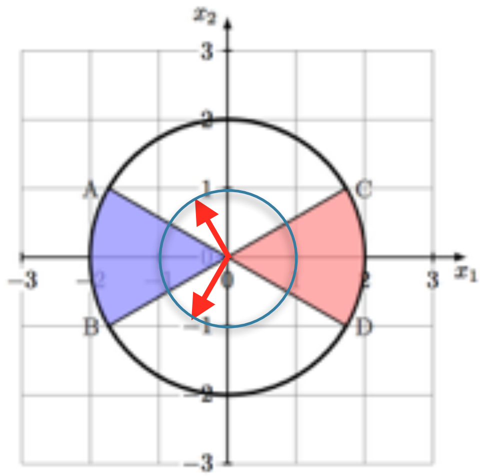
(i)
k-means picks the initial centers randomly. Since they're based on pure luck, they can be selected really badly. The K-means++ algorithm tries to solve this problem, by spreading the initial centers evenly.

Problem2: Disagreement-based Active Learning

(a)
True, because labeling function h* will always be in our hypothesis class H.

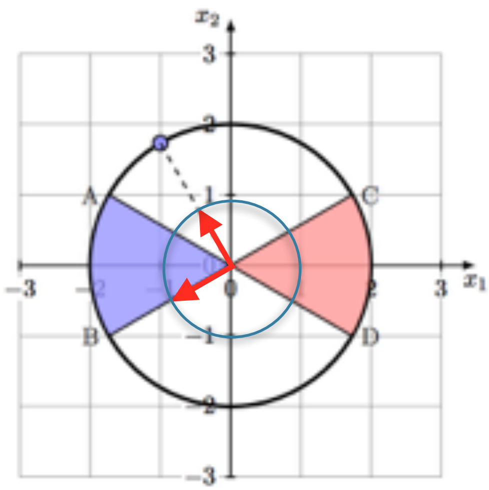
(b)
False, lH_tl >= 1. $\hat{H} \subset H'$

(c)
True, H_hat is always included in H'.
(d)



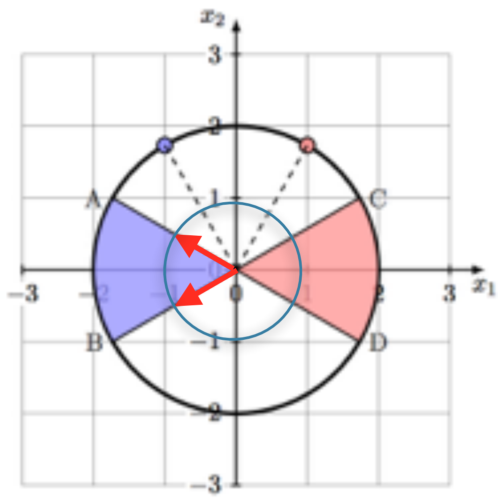
The version space H1 is between the red arrows (120°, 240°), |w| = 1.

(e)



The version space H1 is between the red arrows (120°, 210°), |w| = 1.

(f)



The new version space H3 is between the red arrows (150°, 210°), |w| = 1.

Problem3: Parity Functions

(a)

To prove that VCdim(H_parity) <= n, we need to prove that n+1 points can not be shattered. For n+1 points, we need 2^(n+1) combinations of x to shatter; however, we have only 2^n combinations. Therefore, the VC dimension of H_parity <= n.

(b)

x = {1,0}

pts1 = (1,0)

pts2 = (0,1)

for case: '1,1': S = {1,2}

    h_s(pts1) = 1
    h_s(pts2) = 1

for case: '1,0': S = {1}

    h_s(pts1) = 1
    h_s(pts2) = 0

for case: '0,1': S = {2}

    h_s(pts1) = 0
    h_s(pts2) = 1

for case: '0,0': S = { }

    h_s(pts1) = 0
    h_s(pts2) = 0

VCdim(H_parity) = n