

# Theoretical tasks, Assignment 2

Parth Kalkar

## 1. Theoretical question on K-means Clustering

The first variant of clustering was given in the problem itself, and it was as follows:

$D1 = \{-2, -2, \dots, -2\}$ , with the number of elements in this multiset equals to  $m$ , while  $D2 = \{0, 0, 0, \dots, a\}$ , with size equals to  $m + 1$ , therefore

$$J1 = m((a/(m + 1))^2) + (a - (a/(m + 1)))^2.$$

The second variant which is considered the worst way for clustering the given dataset is distributing the data as follows:  $D1 = \{-2, -2, \dots, -2, 0, 0, \dots, 0\}$ , with the number of elements in this multiset equals to  $2m$ , while  $D2 = \{a\}$ , which leads to  $J2 = 2m$

And by considering those 2 variants we get:

$$\begin{aligned} J1 &< J2 \\ m((a/(m + 1))^2) + (a - (a/(m + 1)))^2 &< 2m \\ (m * a^2 + (m * a)^2) / (m + 1)^2 &< 2m \\ a^2 * m(m + 1) / (m + 1)^2 &< 2m \\ a^2 &< 2 * (m + 1) \end{aligned}$$

## 2. Theoretical question on SVM

(I)

A - Yes

B - No, the hyperplane must pass through the zero, and it's not passing

C - No, because the hard margin can not contain the support vectors inside the margin.

(II)

A - No, the hyperplane can't pass through the zero, and it's passing, which is not possible.

B - Yes,

C - No, because the hard margin can not contain the support vectors inside the margin.

(III)

A - Yes,

B - No, It is not passing through the zero, which can't be possible.

C - Yes.