

Assignment 2

Georgy Andryushchenko B19-DS-01

1 Theoretical question on K-means Clustering

In the context of the given problem, there are 2 cases of two-cluster partitioning possible:

1. m points with $x = -2$ belong to the 1st cluster
 m points with $x = 0$ and the point with $x = a$ belong to the 2nd cluster
2. m points with $x = -2$ and m points with $x = 0$ belong to the 1st cluster
the point with $x = a$ belong to the 2nd cluster

It is stated in the problem that the first case should minimize J . Therefore, J_1 (J for the first case) should be strictly less than J_2 (J for the second case):

$$J_1 < J_2$$

Calculating the means and obtaining the J s, we have:

$$J_1 = a^2 \frac{m^2 + m}{(m + 1)^2}; J_2 = 2m$$

Let us compose the inequality:

$$a^2 \frac{m^2 + m}{(m + 1)^2} < 2m$$

Transforming the inequality we get:

$$a^2 < 2(m + 1)$$

Therefore, the answer is $f(m) = 2(m + 1)$.

2 Theoretical question on SVM

1. Definition (I)

a. *Illustration (a)*

Yes.

b. *Illustration (b)*

No. The hyperplane does not pass the origin on the current illustration.

c. *Illustration (c)*

No. There are vectors that lie within the margin.

2. **Definition (II)**

a. *Illustration (a)*

No. The margin is not maximized which implies that $\frac{1}{2}||\theta||^2$ is not minimized.

b. *Illustration (b)*

Yes.

c. *Illustration (c)*

No. There are vectors that lie within the margin.

3. **Definition (III)**

a. *Illustration (a)*

Yes.

b. *Illustration (b)*

No. The hyperplane does not pass the origin on the current illustration.

c. *Illustration (c)*

Yes.