Assignment 2

Author: Danis Alukaev
Email: d.alukaev@innopolis.university

1. Theoretical question on K-means Clustering

Consider two possible cluster-partitionings:

1. $D_1 = \{-2, -2, ..., -2\}$, $D_2 = \{0, 0, ..., 0, a\}$
2. $D_1 = \{-2, -2, ..., -2, 0, 0, ..., 0\}$, $D_2 = \{a\}$

Note that for these cases the sum-of-squared errors is less for the first partitioning than for the second.

Let's find centers of clusters in both cases:

1. $\mu_1 = \frac{-2 \cdot m}{m} = -2$ and $\mu_2 = \frac{0 \cdot m + a}{m+1} = \frac{a}{m+1}$
2. $\mu_1 = \frac{-2 \cdot m + 0 \cdot m}{2 \cdot m} = -1$ and $\mu_2 = \frac{a}{1} = a$

Further, compute the sum-of-squared errors J for both clusters.

$$J(D_1, \mu_1, \mu_2) = \sum_{i:D_i \neq \phi} \sum_{x \in D_i} (x - \mu_i)^2$$

1. $J(D_1, \mu_1, \mu_2) = m \cdot (-2 - (-2))^2 + m \cdot (0 - \frac{a}{m+1})^2 + 1 \cdot (a - \frac{a}{m+1})^2 = \frac{a^2 m}{m+1}$
2. $J(D_1, \mu_1, \mu_2) = m \cdot (-2 - (-1))^2 + m \cdot (0 - (-1))^2 + 1 \cdot (a - a)^2 = 2m$

Recall that the first sum-of-squared errors should be less than the second:

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

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

Thus, the desired function $f(m)$ is equal to $2(m+1)$.

Answer: $2(m+1)$.

2. Theoretical question on SVM

1. Formulation I
   (a) Yes.
   (b) No. Given plane does not pass through the origin, although by the condition $\theta_0 = 0$.
   (c) No. Two positive points located inside the margin, although the SVM (I) is hard-margin and should satisfy $y_t \theta^T x_t \geq 1$.
2. Formulation II
   (a) No. For $\theta_0 \neq 0$ there is a better hyper-plane with smaller θ_0 and $||\theta||^2$ (greater width of margins).
   (b) Yes.
   (c) No. Two positive points located inside the margin, although the SVM (II) is hard-margin and should satisfy $y_t(\theta^T x_t + \theta_0) \geq 1$.
3. Formulation III
   (a) Yes.
   (b) No. Given plane does not pass through the origin, although by the condition $\theta_0 = 0$.
   (c) Yes.