İstanbul Bilgi University

Department of Computer Engineering

SPRING, 2022
Campus: Santral

CMPE 407
MACHINE LEARNING

HW on K-means

(Duration: minutes)

Name :

Student ID :

Make sure that you explain in detail all your steps - thoughts. You may get extra points for an appropriate observation, you may lose some marks due to an obscure solution.

1. {100 points} K-means algorithm is one of the most successful technique for unsupervised classification. Using K-means the given training set is partitioned into K clusters. Let us consider the case with K=2, n=2, the given training set X ={$x^1, x^2, x^3, x^4, x^5, x^6, x^7$}={(1,1),(1,2), (2,2), (-1,2), (1,7), (-2,1),(-1,8)}, and let us fix the initial centroids of the two clusters to $\mu_1 = (0, 2)$ and $\mu_2 = (0, 5)$. Following the pseudo-code and using the square of the L2 norm as distance [1], show the most significant steps of the K-means algorithm:

K-means algorithm

Randomly initialize K cluster centroids $\mu_1, \mu_2, \ldots, \mu_K \in \mathbb{R}^n$
Repeat {
  for i = 1 to m
    $c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$
  for k = 1 to K
    μ_k := average (mean) of points assigned to cluster k

}

Figure 1: Pseudo-code of the K means algorithm

[1] $||x^i - \mu_j||_2^2 = \sum_t (x_t^i - \mu_{j,t})^2, \forall t = 1, 2$