

VeriMadenciliği – Data Mining

HW – 5 - Solution

1. $x_a = \{6,5,5\}$ $x_b = \{5,4,2\}$ $x_c = \{5,3,4\}$
 x_a noktası x_b ve x_c noktalarından hangisine daha yakın, manhattan mesafeye göre ve öklit mesafeye göre ayrı ayrı bul

(eng) Find whether x_a is closer to x_b or x_c using Manhattan and Euclidean distances separately.

Manhattan için:

$$D(A,B) = |6-5| + |5-4| + |5-2| = 5 \quad D(A,C) = |6-5| + |5-3| + |5-4| = 4$$

Euclidean için:

$$D(A,B) = \sqrt{11} \quad D(A,C) = \sqrt{6}$$

2. Asagıda verilen 8 nokta ile $k=3$ için k-means demetleme yapmak istiyoruz. Öklit (Euclidean) mesafeye kullanılıyor. Basta seçtiğimiz rastgele noktalar $X(0,0)$ $Y(3,3)$ ve $Z(5,5)$ olarak secilmiş. İlk rounddan sonra (algoritmayı devam ettirmeyin) demet merkez noktaları ne olur ve 3 demet hangi elemanlardan oluşur.

(Eng) We want to cluster the given 8 points using k-means algorithm, where $k=3$. Use Euclidean distance function. The initial random points are chosen as $X(0,0)$ $Y(3,3)$ and $Z(5,5)$. Find the centroids and intermediate clusters after the first round. Do not iterate further.

A1 (2,10), A2(2,5), A3(8,4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), A8(4,9)

	A1	A2	A3	A4	A5	A6	A7	A8
X	12	7	12	13	12	10	3	13
Y	8	3	6	7	6	4	3	7
Z	8	3	4	3	2	2	7	5

$$C_X = (1,2)$$

$$C_Y = (2,7.5)$$

$$C_Z = (30/5, 30/5) = (6,6)$$