

K - Means Clustering Algorithm

	T1	T2	T3	← Input Data
D1	0	0.9	0.4	
D2	0.8	0.3	0.5	
D3	1	0	0	
D4	0	1	0	
D5	0.7	0.4	0.6	

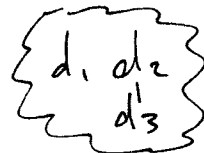
For $K = 2$. Suppose D1 + D2 are the starting (randomly chosen) centroids. What are the clustering results after 1st and 2nd iterations?

Solution

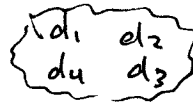
Use the inner ("dot") product between each node and the centroids to see which centroid to place the node into.

Place with the "max" of the two inner products.

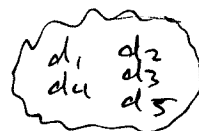
$$d_3 \left(\begin{array}{l} d_{3,1} = (1)(0) + (0)(0.9) + (0)(0.4) = 0 \\ d_{3,2} = (1)(0.8) + (0)(0.3) + (0)(0.5) = 0.8 \end{array} \right. \left[d_3 \text{ goes with } d_2 \right]$$



$$d_4 \left(\begin{array}{l} d_{4,1} = (0)(0) + (1)(0.9) + (0)(0.4) = 0.9 \leftarrow [d_4 \rightarrow d_1] \\ d_{4,2} = (0)(0.8) + (1)(0.3) + (0)(0.5) = 0.3 \end{array} \right.$$



$$d_5 \left(\begin{array}{l} d_{5,1} = (0.7)(0) + 0.4(0.9) + 0.6(0.4) = 0.4 \\ d_{5,2} = (0.7)(0.8) + 0.4(0.3) + 0.6(0.5) = 0.48 \end{array} \right. \left[d_5 \rightarrow d_2 \right]$$



— Iteration 1 complete —

* Now calculate new centroids * [Average the components of each node in the cluster]

$$C_1 = (0+0)/2, (0.9+1)/2, (0.4+0)/2 \\ = 0 \quad 0.95 \quad 0.2$$

$$C_2 = (0.8+1+0.7)/3, (0.3+0+0.4)/3, (0.5+0+0.6)/3 \\ = 0.833 \quad 0.233 \quad 0.367$$

Now calculate the appropriate centroid to place each node (including the initial seed nodes) with.

$$d_{1,c_1} = 0(0) + 0.9(0.95) + 0.4(0.2) = 0.935 \quad \checkmark$$

$$d_{1,c_2} = 0(0.833) + 0.9(0.233) + 0.4(0.367) = 0.36$$

$$d_{2,c_1} = 0.8(0) + 0.3(0.95) + 0.5(0.2) = \dots$$

etc.

$$d_{2,c_2} = 0.8(0.833) + 0.3(0.233) + 0.5(0.367) = \dots$$