Task17

Nawar Saeed

Artificial Intelligence for the Web, VT21
DT506A

16 april 2021

Solution. Lest A be the term-document matrix and u_1 be the $Mx1$ vector, then the Euclidean distance between them:

$$|A - u_1|^2 = (A - u_1) \cdot (A - u_1)$$

$$|A|^2 - |u_1|^2 - 2A \cdot u_1 \Longrightarrow 2 - 2A \cdot u_1 \qquad (1)$$

Now let's execute k-means on the term-document matrix A. Let also M_0 be the initial centroids:

$$M_0 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let transpose M_0 and then compute dot products between A and M_0:

$$M_0^T \cdot A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Using equation (1), it obtains that document 1,2 belongs to the first centroid and document 3,4,5 belongs to the seconds centroid. M_1 becomes:

$$M_1 = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}$$

$$M_1^T \cdot A = \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & 1 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix}$$

Using equation (1), it obtains that document 1,2,3 belongs to the first centroid and document 4,5 belongs to the seconds centroid. M_2 becomes:

$$M_2 = \begin{bmatrix} \frac{2}{3} & 0 \\ \frac{2}{3} & 0 \\ 0 & 1 \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$M_2^T \cdot A = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{3}{2} \end{bmatrix}$$

Using equation (1), it obtains that document 1,2,3 belongs to the first centroid and document 4,5 belongs to the seconds centroid same as previous round. This means that the converges is obtained and M_2 represents as the final centroid

$\square$