Info retrieval sheet 9

Exercise 1

To show: $d(\overline{A}_i,\ \mu_2) < d(\overline{A}_i,\ \mu_2) \Leftrightarrow A_{i1} * \mu_{11} + ... + A_{in} * \mu_{1n} > A_{i1} * \mu_{21} + ... + A_{in} * \mu_{2n}$

$\sqrt{(\overline{A}_{i1} - \mu_{11})^2 + ... + (\overline{A}_{in} - \mu_{1n})^2} < \sqrt{(\overline{A}_{i1} - \mu_{21})^2 + ... + (\overline{A}_{in} - \mu_{2n})^2}$
$\Leftrightarrow -((\overline{A}_{i1} * \mu_{11}) + ... + (\overline{A}_{in} * \mu_{1n})) < -((\overline{A}_{i1} * \mu_{21}) + ... + (\overline{A}_{in} * \mu_{2n}))$
$\Leftrightarrow (\overline{A}_{i1} * \mu_{11}) + ... + (\overline{A}_{in} * \mu_{1n}) > (\overline{A}_{i1} * \mu_{21}) + ... + (\overline{A}_{in} * \mu_{2n})$

We know for each A_i there is a normalizaton factor k so that $k * \overline{A}_i = A_i$

$\Leftrightarrow k * ((\overline{A}_{i1} * \mu_{11}) + ... + (\overline{A}_{in} * \mu_{1n})) > k * ((\overline{A}_{i1} * \mu_{21}) + ... + (\overline{A}_{in} * \mu_{2n}))$
$\Leftrightarrow (A_{i1} * \mu_{11}) + ... + (A_{in} * \mu_{1n}) > (A_{i1} * \mu_{21}) + ... + (A_{in} * \mu_{2n})$
q.e.d.

Exercise 2

$M^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$
$M^T * A = \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}$
1.$Iteration$:
$Clusters$:
c_1 : $\{d_1, d_2\}$
c_2 : $\{d_3, d_4, d_5\}$

$c_1 - vector : \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ $normalized : \begin{pmatrix} \sqrt{2/3} \\ \sqrt{1/3} \\ 0 \\ 0 \end{pmatrix}$

$c_2 - vector : \begin{pmatrix} 0 \\ 1 \\ 2 \\ 1 \end{pmatrix}$ $normalized : \begin{pmatrix} 0 \\ 1/2 \\ 1/\sqrt{2} \\ 1/2 \end{pmatrix}$

2.$Iteration$:
$M^T * A = \begin{pmatrix} (\sqrt{2}+1)/\sqrt{3} & \sqrt{2/3} & \sqrt{1/3} & 0 & 0 \\ 1/2 & 0 & 1/2 & 1/\sqrt{2} & 1/2 \end{pmatrix}$
$Clusters$:
c_1 : $\{d_1, d_2, d_3\}$
c_2 : $\{d_4, d_5\}$

$$c_1 - vector: \begin{pmatrix} 2 \\ 2 \\ 0 \\ 0 \end{pmatrix} \quad normalized: \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \\ 0 \\ 0 \end{pmatrix}$$

$$c_2 - vector: \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1 \end{pmatrix} \quad normalized: \begin{pmatrix} 0 \\ 0 \\ \sqrt{2/3} \\ \sqrt{1/3} \end{pmatrix}$$

3.$Iteration$:

$$M^T * A = \begin{pmatrix} \sqrt{2} & \sqrt{1/2} & \sqrt{1/2} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{2/3} & (\sqrt{2}+1)/\sqrt{3} \end{pmatrix}$$

$Clusters$:

c_1: $\{d_1, d_2, d_3\}$

c_2: $\{d_4, d_5\}$

$$c_1 - vector: \begin{pmatrix} 2 \\ 2 \\ 0 \\ 0 \end{pmatrix} \quad normalized: \begin{pmatrix} \sqrt{1/2} \\ \sqrt{1/2} \\ 0 \\ 0 \end{pmatrix}$$

$$c_2 - vector: \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1 \end{pmatrix} \quad normalized: \begin{pmatrix} 0 \\ 0 \\ \sqrt{2/3} \\ \sqrt{1/3} \end{pmatrix}$$

$\Rightarrow$ The clusters don't change anymore so the algorithm terminates and we are finished