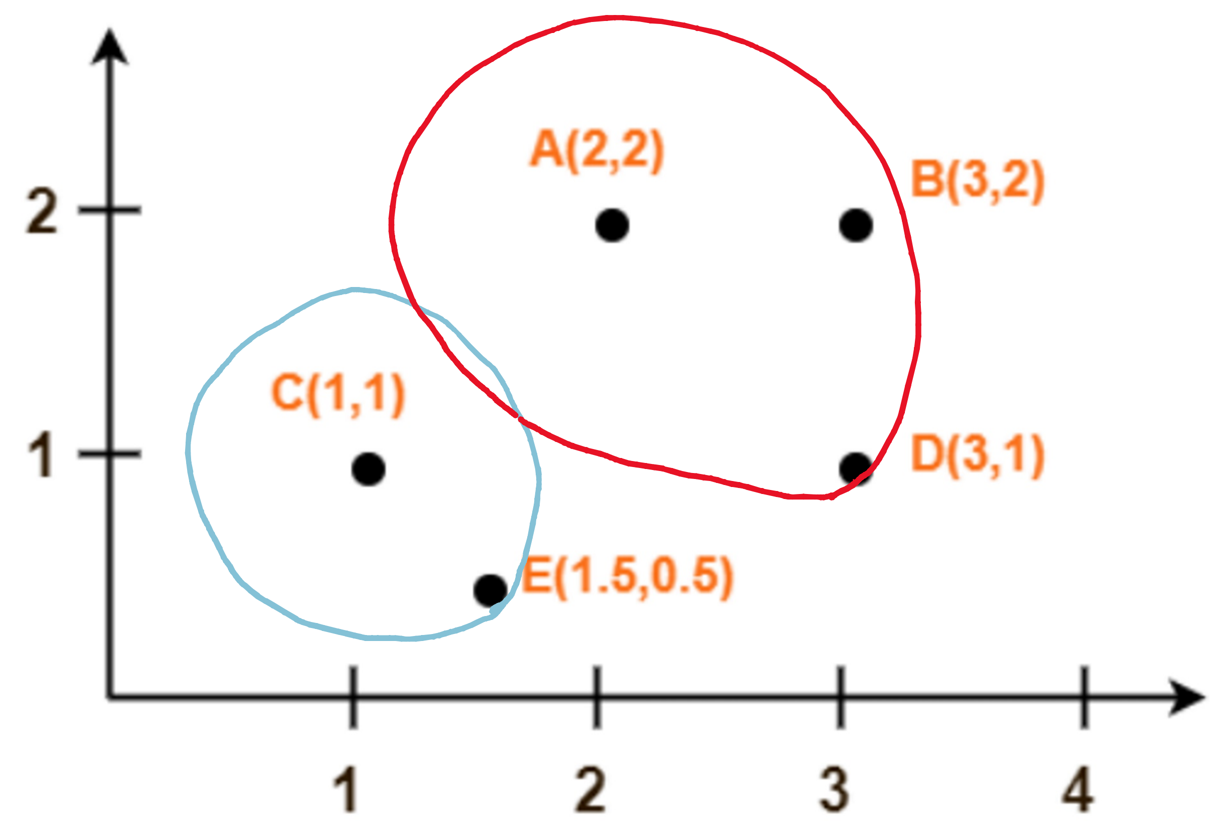
HW6 - Matthew Joel

Problem 1

Use the K-Means Algorithm to create two clusters for the following data:



Assume $A(2,2)$ and $C(1,1)$ are the initial centers of the two clusters. For the first iteration, cluster the points, and update the centers.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

point $B(3,2)$

from $A(2,2)$: $\sqrt{(3-2)^2 + (2-2)^2} = 1$

from $C(1,1)$: $\sqrt{(3-1)^2 + (2-1)^2} = \sqrt{5} \approx 2.2$

$\therefore$ B clusters w/A

point $D(3,1)$

from $A(2,2)$: $\sqrt{(3-2)^2 + (1-2)^2} = \sqrt{2} \approx 1.4$

from $C(1,1)$: $\sqrt{(3-1)^2 + (1-1)^2} = 2$

$\therefore$ D clusters w/A

point $E(1.5, 0.5)$

from $A(2,2)$: $\sqrt{(1.5-2)^2 + (.5-2)^2} = \sqrt{4.5} \approx 2.1$

from $A(2,2)$: $\sqrt{(1.5-2)^2 + (.5-2)^2} = \sqrt{4.5} \approx 2.1$

from $C(1,1)$: $\sqrt{(1.5-1)^2 + (0.5-1)^2} = \sqrt{.5} \approx 0.71$

$\therefore$ E clusters w/ C

Cluster 1:

$$\left(\frac{A_x + B_x + D_x}{3}, \frac{A_y + B_y + D_y}{3}\right) = \text{center}$$

$$= \left(\frac{2+3+3}{3}, \frac{2+2+1}{3}\right) = \left(\frac{8}{3}, \frac{5}{3}\right) \approx (2.7, 1.7)$$

Cluster 2:

$$\left(\frac{C_x + E_x}{2}, \frac{C_y + E_y}{2}\right) = \left(\frac{1.5+1}{2}, \frac{1+.5}{2}\right)$$

$$= (1.25, 0.75)$$

Problem 2 (MATH 5027 ONLY)

Show that $\boldsymbol{\mu}_k$ can be updated by Eqn. (4) once r_{ik}'s are fixed, i.e., if r_{ik}'s are fixed, Eqn. (1) can be minimized by choosing $\boldsymbol{\mu}_k$'s as shown in Eqn. (4).

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} ||\boldsymbol{x}_i - \boldsymbol{\mu}_k||_2^2 \tag{1}$$

$$\boldsymbol{\mu}_k = \frac{\sum_i r_{ik} \boldsymbol{x}_i}{\sum_i r_{ik}} \tag{4}$$

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} ||X_i - \mu_k||_2^2$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} \left((X_i - \mu_k)^T (X_i - \mu_k)\right) r_{ik}$$

$$\sum_{i=1} \sum_{k=1}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} (x_i^T x_i - 2x_i^T \mu_k + \mu_k^T \mu_k)$$

since we want to minimize J w/respect to μ_k:

$$J_k = \sum_{i=1}^{N} r_{ik}(x_i^T x_i - 2x_i^T \mu_k + \mu_k^T \mu_k)$$

$$\frac{\partial J_k}{\partial \mu_k} = \frac{\partial}{\partial \mu_k}\left(\sum_{i=1}^{N} r_{ik}(x_i^T x_i - 2x_i^T \mu_k + \mu_k^T \mu_k)\right)$$

$$= \sum_{i=1}^{N} r_{ik}(-2x_i + 2\mu_k)$$

$$= -2\sum_{i=1}^{N} r_{ik} x_i + 2\sum_{i=1}^{N} r_{ik} \mu_k$$

set derivative $\frac{\partial J_k}{\partial \mu_k}$ equal to 0:

$$0 = -2\sum_{i=1}^{N} r_{ik} x_i + 2\sum_{i=1}^{N} r_{ik} \mu_k$$

$$\sum_{i=1}^{N} r_{ik} x_i = \sum_{i=1}^{N} r_{ik} \mu_k$$

$$\frac{\sum_{i=1}^{N} r_{ik} x_i}{\sum_{i=1}^{N} r_{ik}} = \mu_k \qquad (4)$$

$\therefore$ choosing μ_k's by (4) does minimize (1) when r_{ik}'s are fixed