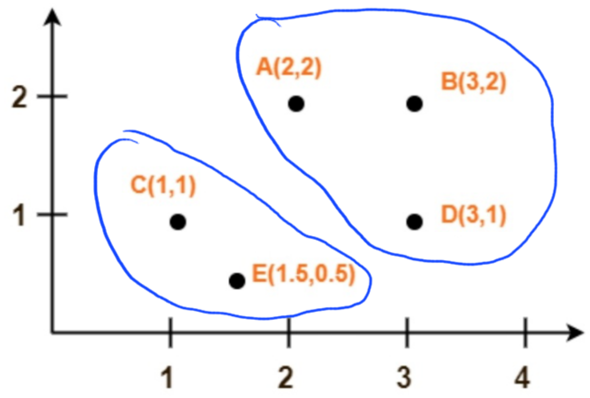
Problem 1

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.

	Center Pt A $(2,2)$	Center Pt C $(1,1)$
$B(3,2)$	$(3-2)^2+(2-2)^2$ $=1$ ✓	$(3-1)^2+(2-1)^2$ $=4+1=5$
$D(3,1)$	$(3-2)^2+(1-2)^2$ $=2$ ✓	$(3-1)^2+(1-1)^2$ $=4$
$E(1.5,0.5)$	$(1.5-2)^2+(0.5-2)^2$ $=0.25+2.25$ $=2.5$	$(1.5-1)^2+(0.5-1)^2$ $0.25+0.25$ $=0.5$ ✓

μ_1 Center for point C&E : $\dfrac{(1+1.5),(1+0.5)}{2} = (1.25, 0.75)$ //

μ_2 Center for point A, B, D : $\dfrac{(2+3+3),(2+2+1)}{3} = \left(\dfrac{8}{3}, \dfrac{5}{3}\right)$ //

Problem 2

Show that if r_{ik}'s are fixed

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

can be minimized by choosing

μ_k as shown in eqn(4) : $\mu_k = \dfrac{\sum_i r_{ik} X_i}{\sum_i r_{ik}}$

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

if r_{ik} is fixed, only one of the r_{ik} is not zero, so this can be reduced to this

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

write term in matrix form

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

expand

$$\frac{dJ}{d\mu_k} = \sum_{i=1}^{N} r_{ik} (-2X_i + 2\mu_k) = 0$$

set the partial derivative to zero

$$\Rightarrow \sum_{i=1}^{N} r_{ik} X_i - \sum_{i=1}^{N} r_{ik} \mu_k = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{i=1}^{N} r_{ik} X_i}{\sum_{i=1}^{N} r_{ik}} \quad //$$

Eqn (4)

$J'' = 2 \sum_{i=1}^{N} r_{ik} > 0$, so Eqn(1) is minimized by choosing μ_k's shown in Eqn(4). //