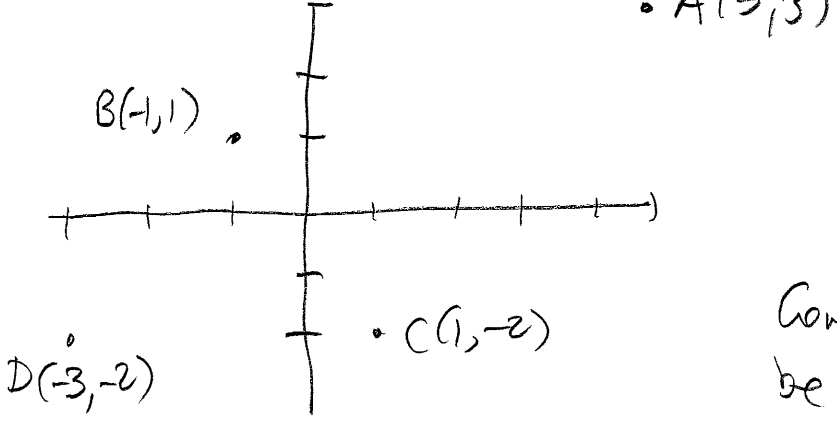
Detailed explanation of manual calculations for a simple example. (k-means)

We have measurement of two variables X_1 & X_2 on four items A, B, C, D.

Item	Observat X_1	X_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

Objective: Divide these items into $k=2$ clusters such that items within a cluster are closer to one another than items in different clusters

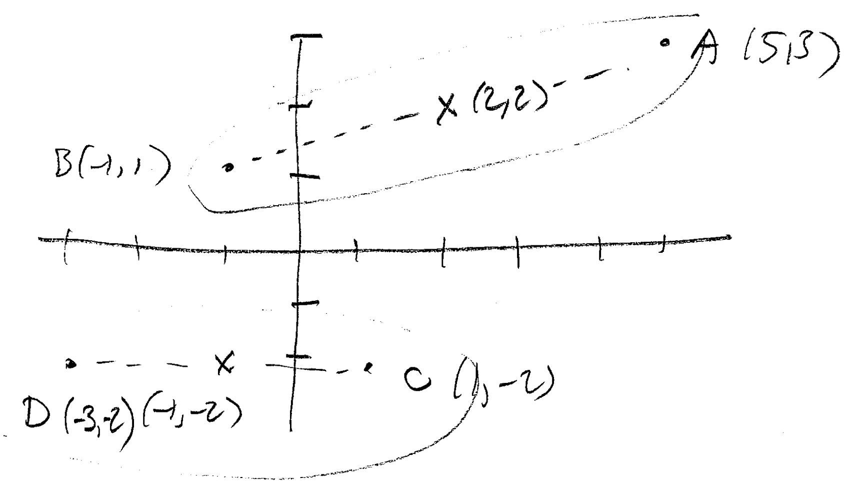


Could BCD be a cluster?

Step 1. Arbitrary partition AB, CD

Centroids (mean)

Cluster	$\bar{x}_1$	$\bar{x}_2$
AB	$\frac{1}{2}(5-1)=2$	$\frac{1}{2}(3+1)=2$
CD	$\frac{1}{2}(1+(-3))=-1$	$\frac{1}{2}(-2+(-2))=-2$



Step 2: Compute distance of each item from all centroids and reassign to nearest group (we use squared distances w.l.o.g)

$$d^2(A, AB) = (5-2)^2 + (3-2)^2 = 10$$
$$d^2(A, CD) = (5+1)^2 + (3+2)^2 = 61$$

Keep A in AB

And, $d^2(B, AB) = \quad = 10$

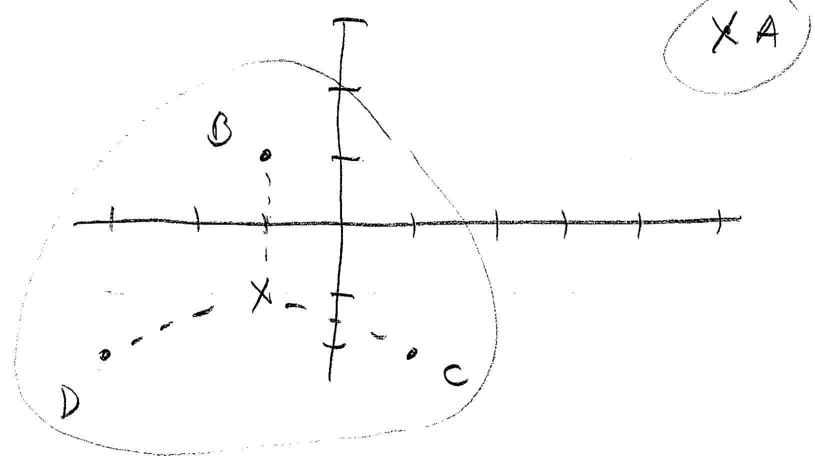
$d^2(B, CD) = \quad = 9$ ← B goes to CD

Of course, C & D should still stay in CD, since $d^2(C,AB) = 17$, $d^2(C,CD) = 2$, and $d^2(D,AB) = 41$, $d^2(D,CD) = 2$.

Step 3

Two clusters now →

	$\bar{x}_1$	$\bar{x}_2$
A	5	3
BCD	−1	−1



Check (squared) distances

	A	B	C	D
A	0 ✓	40	41	89
BCD	52	4 ✓	5 ✓	5 ✓

A is still alone, B, C, D should stay in BCD

(Rcmdr will give the same result!)