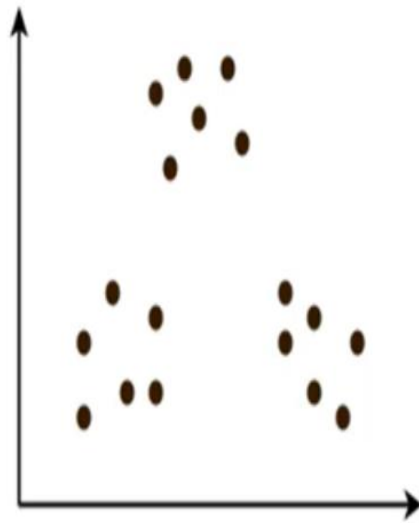
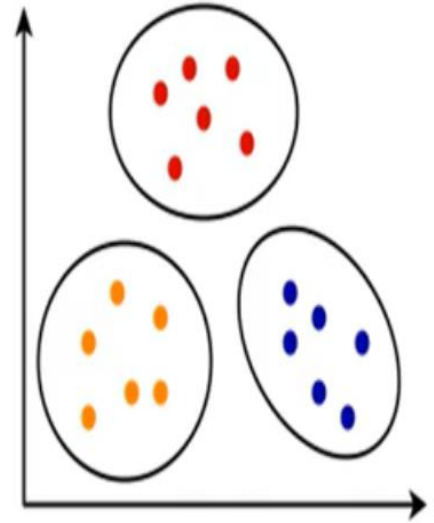


K-Means Clustering



Before K-Means



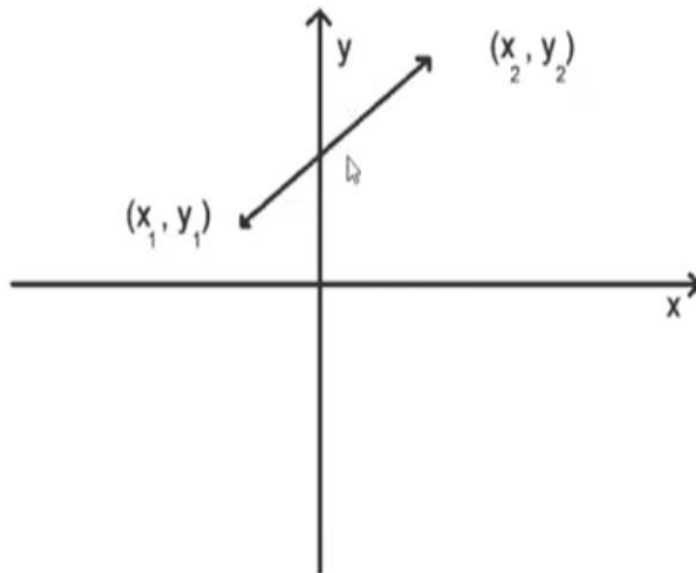
After K-Means

K-Means Clustering

- ***Algorithm (KMeans)***
- ***Metrics***
 - 1. Euclidean Distance***
 - 2. Manhattan Distance***
- ***Elbow method***

K-Means Clustering

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



Euclidean Distance

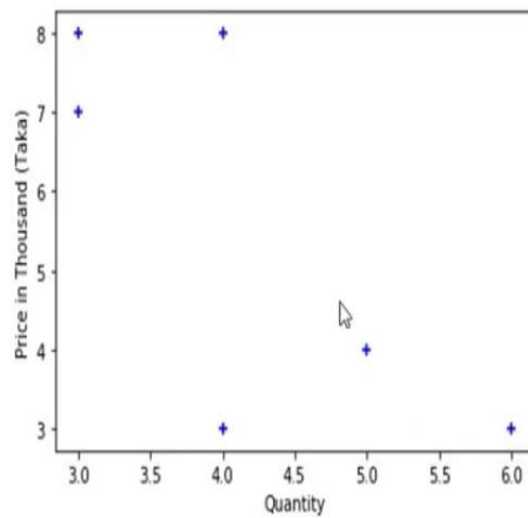
K-Means Clustering



```
In [6]: plt.xlabel('Quantity')
plt.ylabel('Price in Thousand (Taka)')
plt.scatter(dataframe['Quantity'], dataframe['Price(K)'], marker='+', color='blue')
```

Out[6]: <matplotlib.collections.PathCollection at 0x2a9e524a048>

Scatter Plot



K-Means Clustering

	A	B	C	
1	Products	Quantity	Price(K)	
2	FaceWash	3	7	
3	Cream	5	4	
4	Shoes	4	3	
5	Bags	4	8	
6	Jacket	6	3	
7	Shirt	3	8	
8				
9				

$c_1 = (3, 7)$ and $c_2 = (5, 4)$

* For First data point $(3, 7)$ • Facewash:

Distance from $c_1 = 0$ * (1)

Distance from $c_2 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{4+9}$

$= 4.24$

* For second data point $(5, 4)$ • Cream:

Distance from $c_1 = \sqrt{(5-3)^2 + (4-7)^2}$

$= \sqrt{4+9}$

$= \sqrt{13} = 3.60$

Distance from $c_2 = 0$ * (2)

* For third data point $(4, 3)$ shoes:

Distance from $c_1 = \sqrt{(4-3)^2 + (3-7)^2}$

$= \sqrt{1+16}$

$= 4.123$

Distance from $c_2 = \sqrt{(4-5)^2 + (3-4)^2}$

$= \sqrt{1+1}$

$= 1.41$ * (2)

so new centroid $= \left(\frac{5+4}{2}, \frac{4+3}{2} \right)$

$c_2 = (4.5, 3.5)$

K-Means Clustering

	A	B	C	
1	Products	Quantity	Price(K)	
2	FaceWash	3	7	
3	Cream	5	4	
4	Shoes	4	3	
5	Bags	4	8	
6	Jacket	6	3	
7	Shirt	3	8	
8				
9				

$c_1 = (3, 7)$ and $c_2 = (4.5, 3.5)$

For 4th data point (4, 8) bags:

Distance from $c_1 = \sqrt{(4-3)^2 + (8-7)^2}$
 $= \sqrt{1+1}$
 $= 1.41$ (*) (c_1)

Distance from $c_2 = \sqrt{(4-4.5)^2 + (8-3.5)^2}$
 $= 0.25 + 20.25$
 $= 20.50$

\therefore New centroid $= \left(\frac{3+4}{2}, \frac{7+8}{2} \right)$
 $c_1 = (3.5, 7.5)$

For 5th data point (6, 3) Jacket:

Distance from $c_1 = \sqrt{(6-3.5)^2 + (3-7.5)^2}$
 $= 6.25 + 20.25$
 $= 26.5$

Distance from $c_2 = \sqrt{(4.5-6)^2 + (3-3.5)^2}$
 $= 2.25 + 0.25$
 $= 2.50$ (*) (c_2)

New centroid $= \left(\frac{5+4+6}{3}, \frac{4+3+3}{3} \right)$
 $c_2 = (5, 3.33)$

K-Means Clustering

	A	B	C	
1	Products	Quantity	Price(K)	
2	FaceWash	3	7	
3	Cream	5	4	
4	Shoes	4	3	
5	Bags	4	8	
6	Jacket	6	3	
7	Shirt	3	8	
8				
9				

$c_1 = (3.5, 7.5)$ and $c_2 = (5, 3.3)$

For 6th data point (3, 8) shirt :

Distance from $c_1 = \sqrt{(3-3.5)^2 + (8-7.5)^2}$

$$= \sqrt{.25 + .25}$$

$$= 0.70 \text{ * } (c_1)$$

Distance from $c_2 = \sqrt{(3-5)^2 + (8-3.33)^2}$

$$= \sqrt{4 + 2.16}$$

$$= 2.48$$

New centroid = $\left(\frac{3+4+3}{3}, \frac{7+8+8}{3} \right)$

$$c_1 = (3.33, 7.67)$$

$$c_2 = (5, 3.33)$$

K-Means Clustering

The Elbow Method

