Big Data Analytics - 17CS82

Assignment 3 - Three Clusters

Name: Amogh R

USN: 1KS17CS005

Phone: 9972875490

E-mail: amoghpavan5363@gmail.com (mailto:amoghpavan5363@gmail.com)

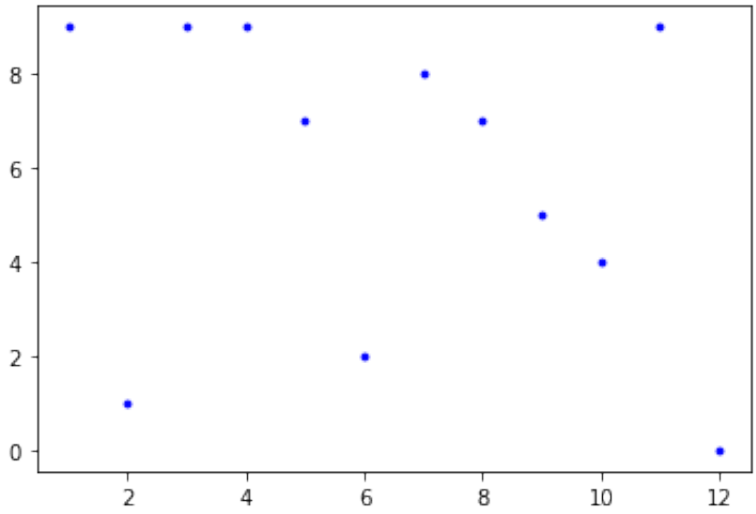
Input

Points	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
X	1	2	3	4	5	6	7	8	9	10	11	12
Y	9	1	9	9	7	2	8	7	5	4	9	0

In [1]:
```
# DRAWING A GRAPH FOR THE ABOVE TABLE
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np

x=[1,2,3,4,5,6,7,8,9,10,11,12]
y=[9,1,9,9,7,2,8,7,5,4,9,0]

plt.scatter(x,y,color='blue',s=8)
data=np.array(list(zip(x,y)))
```



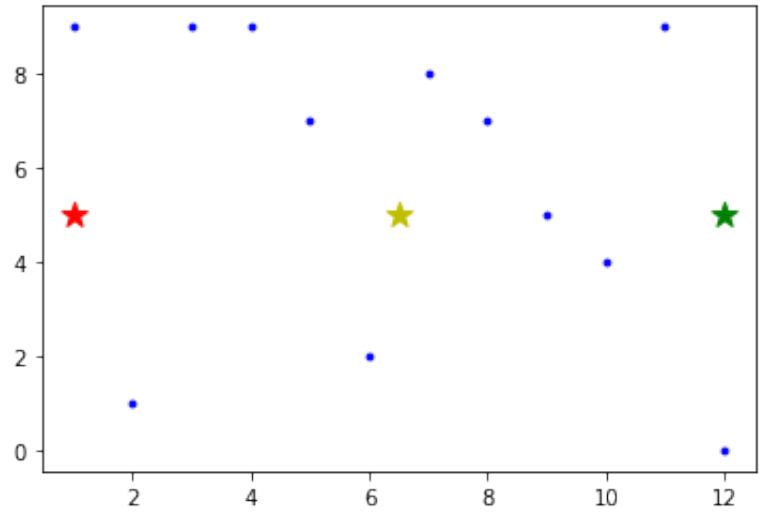
EUCLIDEAN DISTANCE = SQRT [(Xi - Xc) ^ 2 + (Yi - Yc) ^ 2]
Xi = Observed value of X from table
Xc = Centroid value of X
Yi = Observed value of Y from table
Yc = Centroid value of Y

New Centroid Formula = [Sum (Respective Cluster Points (of K1 or K2) + Current Centroid) / Count (Cluster Points + 1)]

In [2]:
```
# FUNCTION FOR CALCULATING EUCLIDEAN DISTANCE
import math
def calculateEuclideanDistance(xi, yi, xc, yc):
    return math.sqrt(math.pow((xi-xc),2) + math.pow((yi-yc),2))
```

Given Initial Centroids are:
C1 = (1,5) which will be represented by RED star
C2 = (6.5,5) which will be represented by YELLOW star
C3 = (12,5) which will be represented by GREEN star

In [3]:
```
# PLOTTING THE DATAPOINTS ALONG WITH CENTROIDS
plt.scatter(x,y,color='blue',s=8)
plt.scatter(1, 5, s=150, c='r', marker='*')
plt.scatter(6.5, 5, s=150, c='y', marker='*');
plt.scatter(12, 5, s=150, c='g', marker='*');
```



ITERATION 1

Centroid 1, C1 = (1,5)
Centroid 2, C2 = (6.5,5)
Centroid 3, C3 = (12,5)

In [4]:
```
# EUCLIDEAN DISTANCE FOR POINT P1=(1, 9), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 1, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 1, 9)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 1, 9)
print(E3)
```

```
4.0
6.800735254367722
11.704699910719626
```

Since **E1 < E2 < E3**, P1(1,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9)}
K2 = {}
K3 = {}

In [6]:
```
# EUCLIDEAN DISTANCE FOR POINT P2=(2, 1), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 2, 1)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 2, 1)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 2, 1)
print(E3)
```

```
4.123105625617661
6.020797289396148
10.770329614269007
```

Since **E1 < E2 < E3**, P2(2,1) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1)}
K2 = {}
K3 = {}

In [7]:
```
# EUCLIDEAN DISTANCE FOR POINT P3=(3, 9), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 3, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 3, 9)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 3, 9)
print(E3)
```

```
4.47213595499958
5.315072906367325
9.848857801796104
```

Since **E1 < E2 < E3**, P3(3,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {}
K3 = {}

In [8]:
```
# EUCLIDEAN DISTANCE FOR POINT P4=(4, 9), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 4, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 4, 9)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 4, 9)
print(E3)
```

```
5.0
4.716990566028302
8.94427190999916
```

Since **E2 < E1 < E3**, P4(4,9) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9)}
K3 = {}

In [9]:
```
# EUCLIDEAN DISTANCE FOR POINT P5=(5, 7), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 5, 7)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 5, 7)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 5, 7)
print(E3)
```

```
4.47213595499958
2.5
7.280109889280518
```

Since **E2 < E1 < E3**, P5(5,7) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7)}
K3 = {}

In [10]:
```
# EUCLIDEAN DISTANCE FOR POINT P6=(6, 2), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 6, 2)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 6, 2)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 6, 2)
print(E3)
```

```
5.830951894845301
3.0413812651491097
6.708203932499369
```

Since **E2 < E1 < E3**, P6(6,2) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2)}
K3 = {}

In [12]:
```
# EUCLIDEAN DISTANCE FOR POINT P7=(7, 8), CENTROIDS=(1,5),(6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 7, 8)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 7, 8)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 7, 8)
print(E3)
```

```
6.708203932499369
3.0413812651491097
5.830951894845301
```

Since **E2 < E3 < E1**, P7(7,8) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2), P7(7,8)}
K3 = {}

In [13]:
```
# EUCLIDEAN DISTANCE FOR POINT P8=(8,7), CENTROIDS=(1,5), (6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 8, 7)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 8, 7)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 8, 7)
print(E3)
```

```
7.280109889280518
2.5
4.47213595499958
```

Since **E2 < E3 < E1**, P8(8,7) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {}

In [15]:
```
# EUCLIDEAN DISTANCE FOR POINT P9=(9,5), CENTROIDS=(1,5), (6.5,5) a
E1 = calculateEuclideanDistance(1, 5, 9, 5)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 9, 5)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 9, 5)
print(E3)
```

```
8.0
2.5
3.0
```

Since **E2 < E3 < E1**, P9(9,5) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2), P7(7,8), P8(8,7), P9(9,5)}
K3 = {}

In [16]:
```
# EUCLIDEAN DISTANCE FOR POINT P10=(10, 4), CENTROIDS=(1,5), (6.5,5
E1 = calculateEuclideanDistance(1, 5, 10, 4)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 10, 4)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 10, 4)
print(E3)
```

```
9.055385138137417
3.640054944640259
2.23606797749979
```

Since **E3 < E2 < E1**, P10(10,4) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2), P7(7,8), P8(8,7), P9(9,5)}
K3 = {P10(10,4)}

In [17]:
```
# EUCLIDEAN DISTANCE FOR POINT P11=(11, 9), CENTROIDS=(1,5), (6.5,5
E1 = calculateEuclideanDistance(1, 5, 11, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 11, 9)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 11, 9)
print(E3)
```

```
10.770329614269007
6.020797289396148
4.123105625617661
```

Since **E3 < E2 < E1**, P11(11,9) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2), P7(7,8), P8(8,7), P9(9,5)}
K3 = {P10(10,4), P11(11,9)}

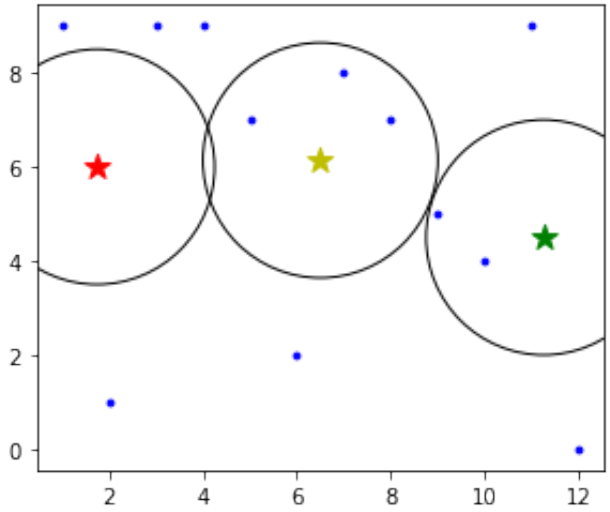
In [18]:
```
# EUCLIDEAN DISTANCE FOR POINT P12=(12, 0), CENTROIDS=(1,5), (6.5,5
E1 = calculateEuclideanDistance(1, 5, 12, 0)
print(E1)
E2 = calculateEuclideanDistance(6.5, 5, 12, 0)
print(E2)
E3 = calculateEuclideanDistance(12, 5, 12, 0)
print(E3)
```

```
12.083045973594572
7.433034373659253
5.0
```

Since **E3 < E2 < E1**, P12(12,0) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {P4(4,9), P5(5,7), P6(6,2), P7(7,8), P8(8,7), P9(9,5)}
K3 = {P10(10,4), P11(11,9), P12(12,0)}

New Centroid Formula = [Sum (Respective Cluster Points (of K1 or K2 or K3) + Current Centroid) / Count (Cluster Points + 1)]
Centroid C1(1,5) becomes ((1+2+3+1)/4,(9+1+9+5)/4) => **new C1 is (1.75,6)**
Centroid C2(6.5,5) becomes ((4+5+6+7+8+9+6.5)/7,(9+7+2+8+7+5+5)/7) => **new C2 is (6.5,6.14)**
Centroid C3(12,5) becomes ((10+11+12+12)/4,(4+9+0+5)/4) => **new C3 is (11.25,4.5)**

In [21]:
```
# PLOTTING THE DATAPOINTS ALONG WITH NEW CENTROIDS
figure, axes = plt.subplots()
circle1 = plt.Circle((1.75, 6), 2.5, fill = False)
circle2 = plt.Circle((6.5, 6.14), 2.5, fill = False)
circle3 = plt.Circle((11.25, 4.5), 2.5, fill = False)
axes.set_aspect(1)
axes.add_artist(circle1)
axes.add_artist(circle2)
axes.add_artist(circle3)
plt.scatter(x,y,color='blue',s=8)
plt.scatter(1.75, 6, s=150, c='r', marker='*')
plt.scatter(6.5, 6.14, s=150, c='y', marker='*')
plt.scatter(11.25, 4.5, s=150, c='g', marker='*')
plt.show()
```



ITERATION 2

Centroid 1, C1 = (1.75,6)
Centroid 2, C2 = (6.5,6.14)
Centroid 3, C3 = (11.25,4.5)

In [35]:
```
# EUCLIDEAN DISTANCE FOR POINT P1=(1, 9), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 1, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 1, 9)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 1, 9)
print(E3)
```

```
3.092329219213245
6.199161233586363
11.194306588619055
```

Since **E1 < E2 < E3**, P1(1,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9)}
K2 = {}
K3 = {}

In [36]:
```
# EUCLIDEAN DISTANCE FOR POINT P2=(2, 1), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 2, 1)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 2, 1)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 2, 1)
print(E3)
```

```
5.006246098625197
6.83151520528206
9.89002022242624
```

Since **E1 < E2 < E3**, P2(2,1) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1)}
K2 = {}
K3 = {}

In [38]:
```
# EUCLIDEAN DISTANCE FOR POINT P3=(3, 9), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 3, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 3, 9)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 3, 9)
print(E3)
```

```
3.25
4.519911503558449
9.39747306460625
```

Since **E1 < E2 < E3**, P3(3,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {}
K3 = {}

In [39]:
```
# EUCLIDEAN DISTANCE FOR POINT P4=(4, 9), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 4, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 4, 9)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 4, 9)
print(E3)
```

```
3.75
3.7986313324669982
8.533024082938006
```

Since **E1 < E2 < E3**, P4(4,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {}
K3 = {}

In [40]:
```
# EUCLIDEAN DISTANCE FOR POINT P5=(5, 7), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 5, 7)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 5, 7)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 5, 7)
print(E3)
```

```
3.400367627183861
1.7290459797240791
6.73145600891813
```

Since **E2 < E1 < E3**, P5(5,7) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7)}
K3 = {}

In [41]:
```
# EUCLIDEAN DISTANCE FOR POINT P6=(6, 2), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 6, 2)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 6, 2)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 6, 2)
print(E3)
```

```
5.836308764964376
4.170083932009042
5.814851674806504
```

Since **E2 < E1 < E3**, P6(6,2) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2)}
K3 = {}

In [42]:
```
# EUCLIDEAN DISTANCE FOR POINT P7=(7, 8), CENTROIDS=(1.75,6), (6.5,
E1 = calculateEuclideanDistance(1.75, 6, 7, 8)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 7, 8)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 7, 8)
print(E3)
```

```
5.618051263561058
1.926032190800559
5.50567888638631
```

Since **E2 < E3 < E1**, P7(7,8) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8)}
K3 = {}

In [43]:
```
# EUCLIDEAN DISTANCE FOR POINT P8=(8,7), CENTROIDS=(1.75,6), (6.5,6
E1 = calculateEuclideanDistance(1.75, 6, 8, 7)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 8, 7)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 8, 7)
print(E3)
```

```
6.329494450586082
1.7290459797240791
4.100304866714182
```

Since **E2 < E3 < E1**, P8(8,7) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {}

In [44]:
```
# EUCLIDEAN DISTANCE FOR POINT P9=(9,5), CENTROIDS=(1.75,6), (6.5,6
E1 = calculateEuclideanDistance(1.75, 6, 9, 5)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 9, 5)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 9, 5)
print(E3)
```

```
7.318640584152224
2.7476535443901944
2.3048861143232218
```

Since **E3 < E2 < E1**, P9(9,5) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5)}

In [45]:
```
# EUCLIDEAN DISTANCE FOR POINT P10=(10, 4), CENTROIDS=(1.75,6), (6.
E1 = calculateEuclideanDistance(1.75, 6, 10, 4)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 10, 4)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 10, 4)
print(E3)
```

```
8.488963423174823
4.102389547568587
1.346291201783626
```

Since **E3 < E2 < E1**, P10(10,4) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5), P10(10,4)}

In [46]:
```
# EUCLIDEAN DISTANCE FOR POINT P11=(11, 9), CENTROIDS=(1.75,6), (6.
E1 = calculateEuclideanDistance(1.75, 6, 11, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 11, 9)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 11, 9)
print(E3)
```

```
9.724325169388361
5.331941485050263
4.5069390943299865
```

Since **E3 < E2 < E1**, P11(11,9) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5), P10(10,4), P11(11,9)}

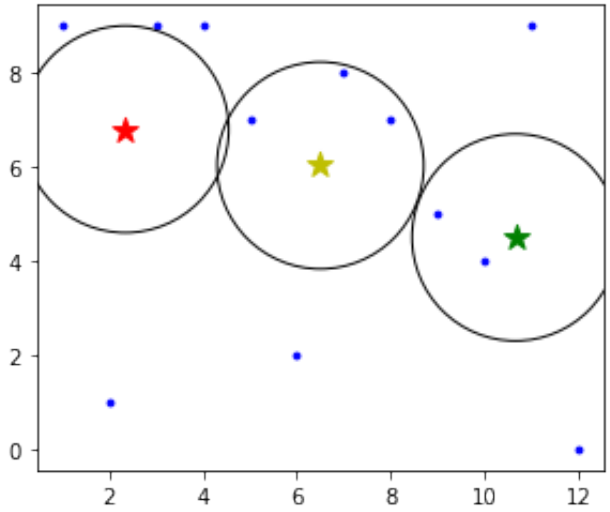
In [48]:
```
# EUCLIDEAN DISTANCE FOR POINT P12=(12, 0), CENTROIDS=(1.75,6), (6.
E1 = calculateEuclideanDistance(1.75, 6, 12, 0)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.14, 12, 0)
print(E2)
E3 = calculateEuclideanDistance(11.25, 4.5, 12, 0)
print(E3)
```

```
11.876973520219703
8.243154735900571
4.562071897723665
```

Since **E3 < E2 < E1**, P12(12,0) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5), P10(10,4), P11(11,9), P12(12,0)}

New Centroid Formula = [Sum (Respective Cluster Points (of K1 or K2 or K3) + Current Centroid) / Count (Cluster Points + 1)]
Centroid C1(1,5) becomes ((1+2+3+4+1.75)/5,(9+1+9+9+6)/5) => **new C1 is (2.35,6.8)**
Centroid C2(6.5,5) becomes ((5+6+7+8+6.5)/5,(7+2+8+7+6.14)/5) => **new C2 is (6.5,6.03)**
Centroid C3(12,5) becomes ((9+10+11+12+11.25)/5,(5+4+9+0+4.5)/5) => **new C3 is (10.65,4.5)**

In [50]:
```
# PLOTTING THE DATAPOINTS ALONG WITH NEW CENTROIDS
figure, axes = plt.subplots()
circle1 = plt.Circle((2.35,6.8), 2.2, fill = False)
circle2 = plt.Circle((6.5,6.03), 2.2, fill = False)
circle3 = plt.Circle((10.65,4.5), 2.2, fill = False)
axes.set_aspect(1)
axes.add_artist(circle1)
axes.add_artist(circle2)
axes.add_artist(circle3)
plt.scatter(x,y,color='blue',s=8)
plt.scatter(2.35, 6.8, s=150, c='r', marker='*')
plt.scatter(6.5, 6.03, s=150, c='y', marker='*')
plt.scatter(10.65, 4.5, s=150, c='g', marker='*')
plt.show()
```



ITERATION 3

Centroid 1, C1 = (2.35,6.8)
Centroid 2, C2 = (6.5,6.03)
Centroid 3, C3 = (10.65,4.5)

In [51]:
```
# EUCLIDEAN DISTANCE FOR POINT P1=(1, 9), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 1, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 1, 9)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 1, 9)
print(E3)
```

```
2.581181899828062
6.2506719638771635
10.647652323399745
```

Since **E1 < E2 < E3**, P1(1,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9)}
K2 = {}
K3 = {}

In [52]:
```
# EUCLIDEAN DISTANCE FOR POINT P2=(2, 1), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 2, 1)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 2, 1)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 2, 1)
print(E3)
```

```
5.810550748423079
6.749140686042928
9.331264651696468
```

Since **E1 < E2 < E3**, P2(2,1) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1)}
K2 = {}
K3 = {}

In [53]:
```
# EUCLIDEAN DISTANCE FOR POINT P3=(3, 9), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 3, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 3, 9)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 3, 9)
print(E3)
```

```
2.2940139493908926
4.590305000759753
8.87538731549221
```

Since **E1 < E2 < E3**, P3(3,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9)}
K2 = {}
K3 = {}

In [54]:
```
# EUCLIDEAN DISTANCE FOR POINT P4=(4, 9), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 4, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 4, 9)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 4, 9)
print(E3)
```

```
2.75
3.88212570636243
8.029476944359452
```

Since **E1 < E2 < E3**, P4(4,9) will become part of cluster K1 representing C1 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {}
K3 = {}

In [55]:
```
# EUCLIDEAN DISTANCE FOR POINT P5=(5, 7), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 5, 7)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 5, 7)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 5, 7)
print(E3)
```

```
2.6575364531836625
1.7863090438107285
6.178389757857625
```

Since **E2 < E1 < E3**, P5(5,7) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7)}
K3 = {}

In [56]:
```
# EUCLIDEAN DISTANCE FOR POINT P6=(6, 2), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 6, 2)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 6, 2)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 6, 2)
print(E3)
```

```
6.030132668523969
4.060898915264945
5.279441258315126
```

Since **E2 < E1 < E3**, P6(6,2) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2)}
K3 = {}

In [57]:
```
# EUCLIDEAN DISTANCE FOR POINT P7=(7, 8), CENTROIDS=(2.35,6.8), (6.
E1 = calculateEuclideanDistance(2.35, 6.8, 7, 8)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 7, 8)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 7, 8)
print(E3)
```

```
4.802343178074637
2.0324615617521524
5.056925943693462
```

Since **E2 < E3 < E1**, P7(7,8) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8)}
K3 = {}

In [58]:
```
# EUCLIDEAN DISTANCE FOR POINT P8=(8,7), CENTROIDS=(2.35,6.8), (6.5
E1 = calculateEuclideanDistance(2.35, 6.8, 8, 7)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 8, 7)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 8, 7)
print(E3)
```

```
5.653538714822779
1.7863090438107285
3.643144246389374
```

Since **E2 < E3 < E1**, P8(8,7) will become part of cluster K2 representing C2 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {}

In [59]:
```
# EUCLIDEAN DISTANCE FOR POINT P9=(9,5), CENTROIDS=(2.35,6.8), (6.5
E1 = calculateEuclideanDistance(2.35, 6.8, 9, 5)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 9, 5)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 9, 5)
print(E3)
```

```
6.889303302947258
2.7038676003088615
1.7240939649566671
```

Since **E3 < E2 < E1**, P9(9,5) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5)}

In [60]:
```
# EUCLIDEAN DISTANCE FOR POINT P10=(10, 4), CENTROIDS=(2.35,6.8), (
E1 = calculateEuclideanDistance(2.35, 6.8, 10, 4)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 10, 4)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 10, 4)
print(E3)
```

```
8.146318186763883
4.046096884653159
0.8200609733428366
```

Since **E3 < E2 < E1**, P10(10,4) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5), P10(10,4)}

In [61]:
```
# EUCLIDEAN DISTANCE FOR POINT P11=(11, 9), CENTROIDS=(2.35,6.8), (
E1 = calculateEuclideanDistance(2.35, 6.8, 11, 9)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 11, 9)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 11, 9)
print(E3)
```

```
8.925385145751415
5.391743688270057
4.513590588434002
```

Since **E3 < E2 < E1**, P11(11,9) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5), P10(10,4), P11(11,9)}

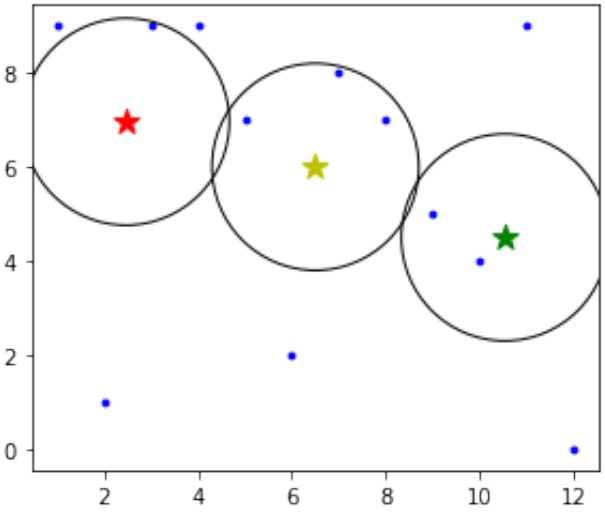
In [62]:
```
# EUCLIDEAN DISTANCE FOR POINT P12=(12, 0), CENTROIDS=(2.35,6.8), (
E1 = calculateEuclideanDistance(2.35, 6.8, 12, 0)
print(E1)
E2 = calculateEuclideanDistance(6.5, 6.03, 12, 0)
print(E2)
E3 = calculateEuclideanDistance(10.65, 4.5, 12, 0)
print(E3)
```

```
11.80518953680965
8.161550097867439
4.698137929009747
```

Since **E3 < E2 < E1**, P12(12,0) will become part of cluster K3 representing C3 as centroid
Therefore,
K1 = {P1(1,9), P2(2,1), P3(3,9), P4(4,9)}
K2 = {P5(5,7), P6(6,2), P7(7,8), P8(8,7)}
K3 = {P9(9,5), P10(10,4), P11(11,9), P12(12,0)}

New Centroid Formula = [Sum (Respective Cluster Points (of K1 or K2 or K3) + Current Centroid) / Count (Cluster Points + 1)]
Centroid C1(1,5) becomes ((1+2+3+4+2.35)/5,(9+1+9+9+6.8)/5) => **new C1 is (2.47,6.96)**
Centroid C2(6.5,5) becomes ((5+6+7+8+6.5)/5,(7+2+8+7+6.03)/5) => **new C2 is (6.5,6)**
Centroid C3(12,5) becomes ((9+10+11+12+10.65)/5,(5+4+9+0+4.5)/5) => **new C3 is (10.53,4.5)**

In [63]:
```
# PLOTTING THE DATAPOINTS ALONG WITH NEW CENTROIDS
figure, axes = plt.subplots()
circle1 = plt.Circle((2.47,6.96), 2.2, fill = False)
circle2 = plt.Circle((6.5,6), 2.2, fill = False)
circle3 = plt.Circle((10.53,4.5), 2.2, fill = False)
axes.set_aspect(1)
axes.add_artist(circle1)
axes.add_artist(circle2)
axes.add_artist(circle3)
plt.scatter(x,y,color='blue',s=8)
plt.scatter(2.47, 6.96, s=150, c='r', marker='*')
plt.scatter(6.5, 6, s=150, c='y', marker='*')
plt.scatter(10.53, 4.5, s=150, c='g', marker='*')
plt.show()
```



Initial Centroids

C1 = (1,5)
C2 = (6.5,5)
C3 = (12,5)

Centroids after Iteration 1

C1 = (2.35,6.8)
C2 = (6.5,6.03)
C3 = (10.65,4.5)

Centroids after Iteration 2

C1 = (2.47,6.96)
C2 = (6.5,6)
C3 = (10.53,4.5)

We see that the centroids are stabilising at this point, hence the iterations have been stopped here.