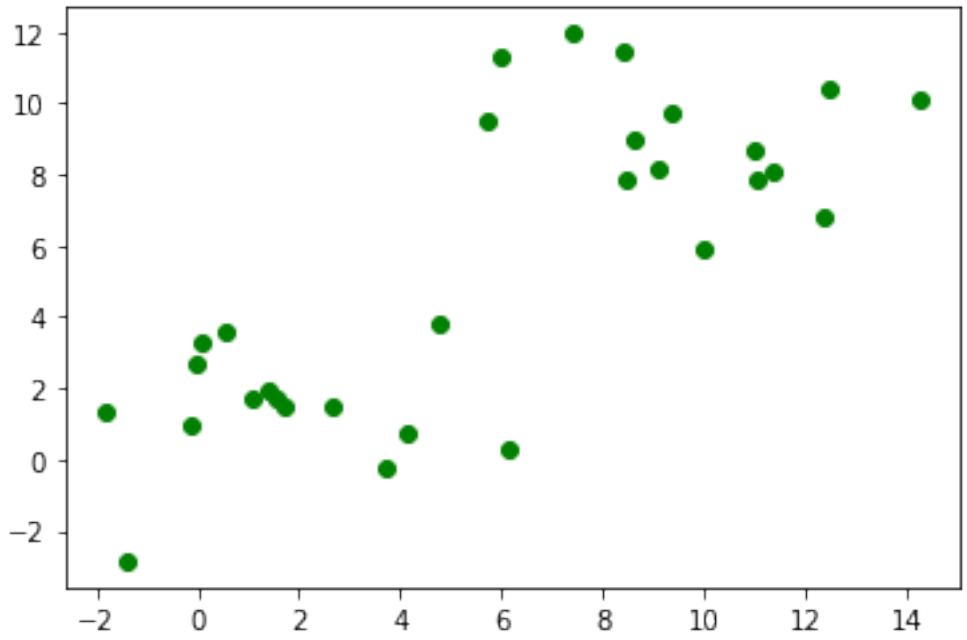
kmeans

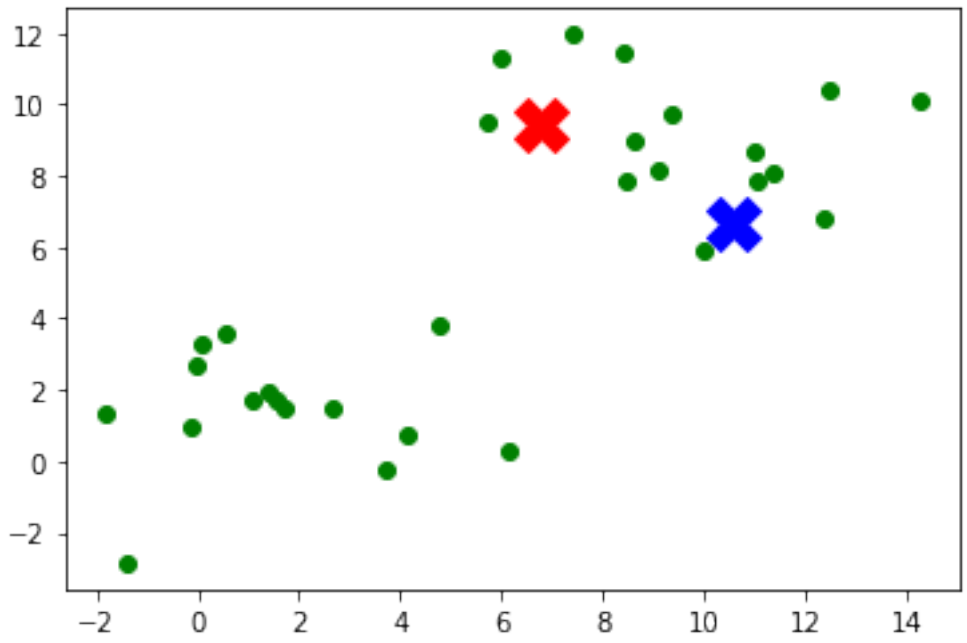
January 13, 2022

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
```

[8]:
```
# generate two clusters of points with normal distributions.
means = [[1,1],[10,10]]
points1 = np.random.normal(means[0],scale=2.5,size=(15,2))
points2 = np.random.normal(means[1],scale=2.5,size=(15,2))
points = np.concatenate([points1,points2],0)
plt.plot(points[:,0],points[:,1], 'o', color='green');
```

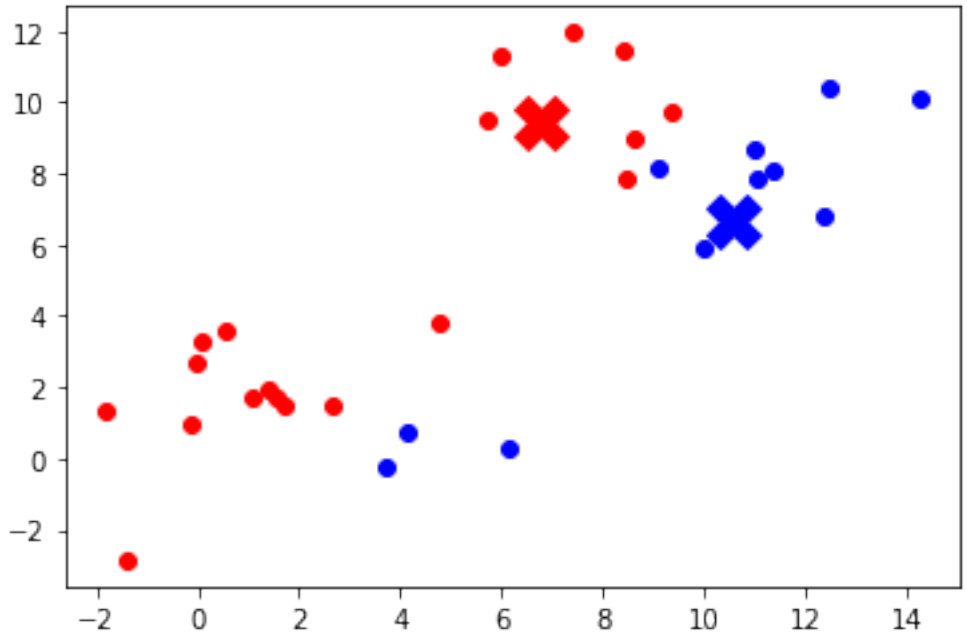


[16]:
```
# Before inner loop of Kmeans. Initilize two random cluster centroids.
centroid1 = np.random.random(size=2)*15
centroid2 = np.random.random(size=2)*15
plt.plot(points[:,0],points[:,1], 'o', color='green');
plt.plot(centroid1[0],centroid1[1],marker='X',markersize=20, color='red');
plt.plot(centroid2[0],centroid2[1],marker='X',markersize=20, color='blue');
```

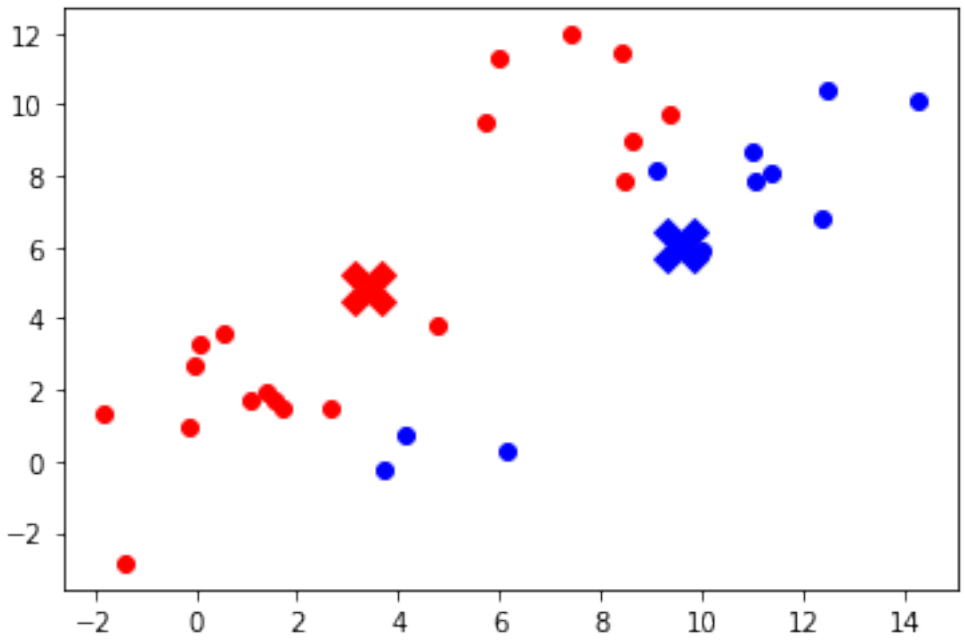


[23]:
```
# innner loop of kmeans. cluster assignment step
def assignment(points,cluster1,cluster2, plot=True):
    cluster1 = []
    cluster2 = []
    for point in points:
        dist1 = np.linalg.norm(point - centroid1)
        dist2 = np.linalg.norm(point - centroid2)
        if dist1 < dist2:
            cluster1.append(point)
        else:
            cluster2.append(point)
    cluster1 = np.stack(cluster1,0)
    cluster2 = np.stack(cluster2,0)
    if plot:
        plt.plot(cluster1[:,0],cluster1[:,1], 'o', color='red');
        plt.plot(centroid1[0],centroid1[1],marker='X',markersize=20,␣
↪color='red');

        plt.plot(cluster2[:,0],cluster2[:,1], 'o', color='blue');
        plt.plot(centroid2[0],centroid2[1],marker='X',markersize=20,␣
↪color='blue');

    return cluster1, cluster2
```

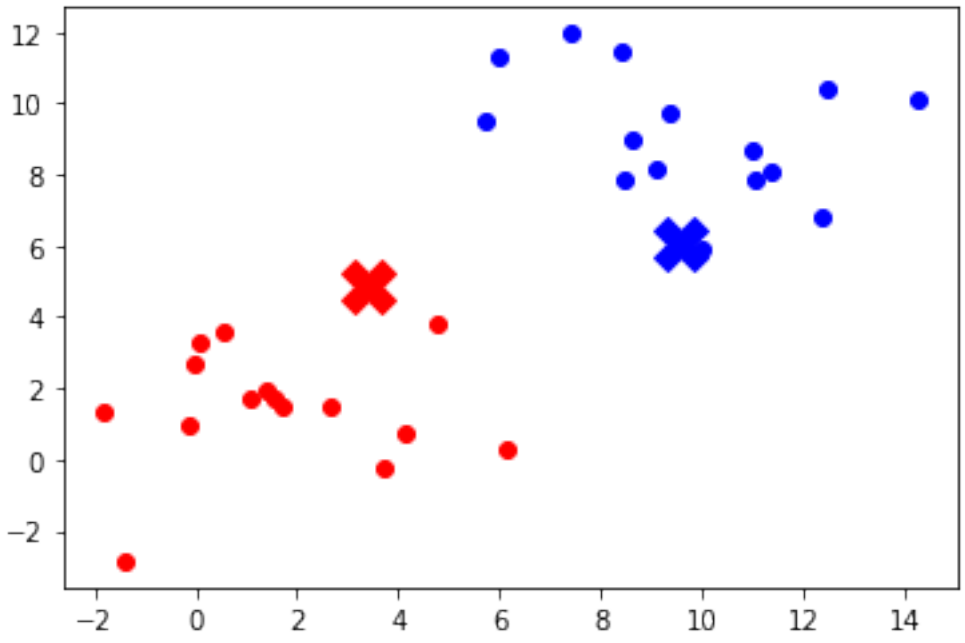
```
cluster1, cluster2 = assignment(points,cluster1,cluster2, plot=True)
```



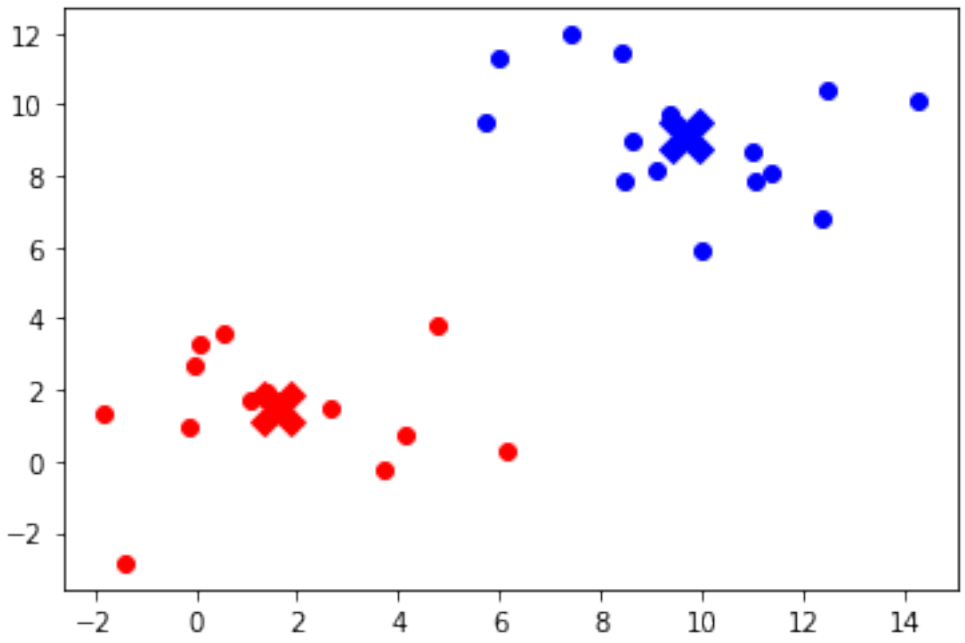
[26]:
```
# innner loop of kmeans, update centroid step
def move_centroid(cluster1,cluster2, plot=True):
    centroid1 = np.mean(cluster1,0)
    centroid2 = np.mean(cluster2,0)

    if plot:
        plt.plot(cluster1[:,0],cluster1[:,1], 'o', color='red');
        plt.plot(centroid1[0],centroid1[1],marker='X',markersize=20,␣
↪color='red');

        plt.plot(cluster2[:,0],cluster2[:,1], 'o', color='blue');
        plt.plot(centroid2[0],centroid2[1],marker='X',markersize=20,␣
↪color='blue');

    return centroid1, centroid2
centroid1, centroid2 = move_centroid(cluster1,cluster2)
```



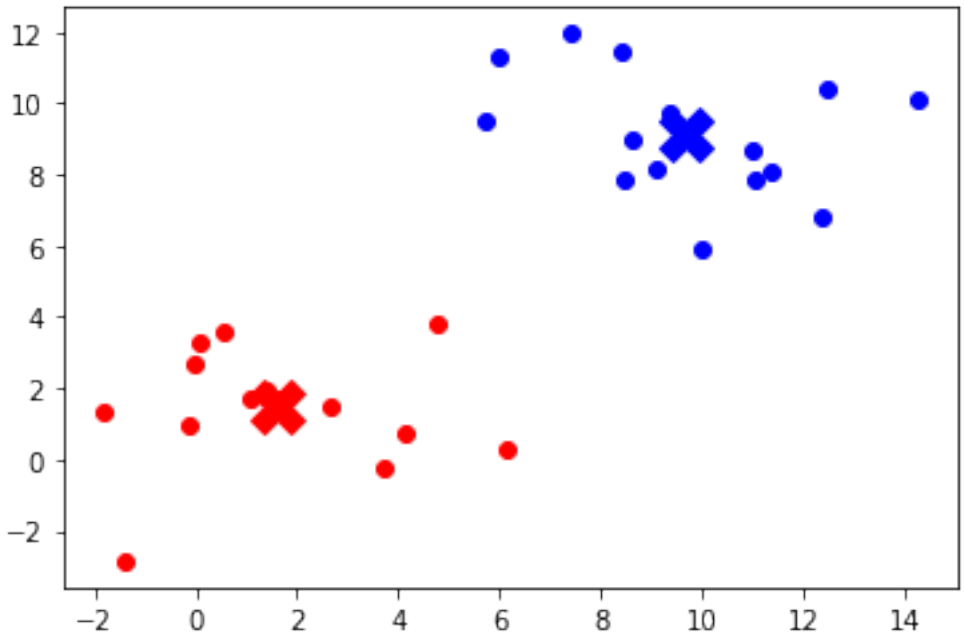
[27]:
```
cluster1, cluster2 = assignment(points,cluster1,cluster2, plot=True)
```



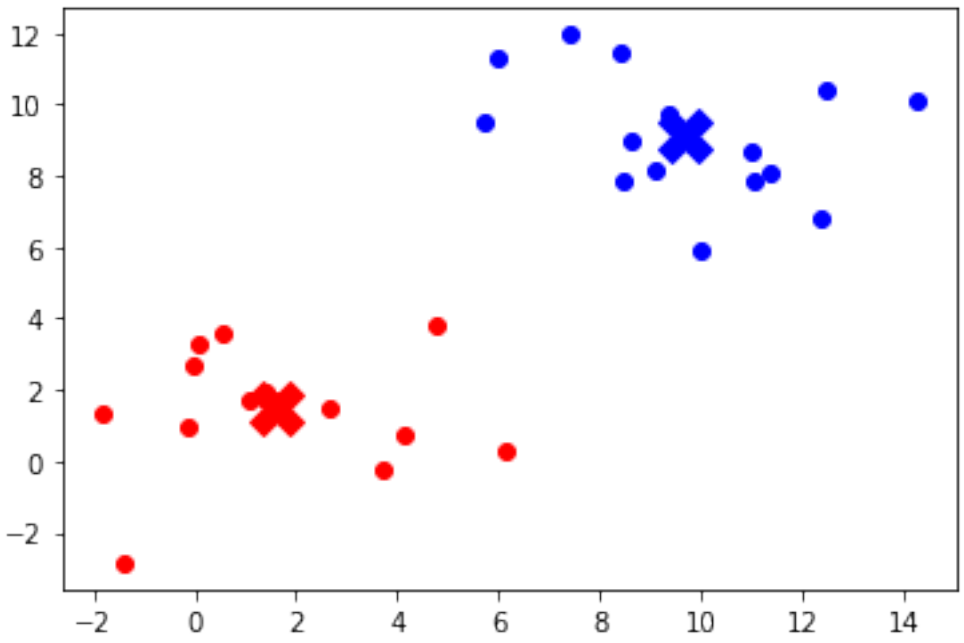
[28]:
```
centroid1, centroid2 = move_centroid(cluster1,cluster2)
```



[29]:
```
cluster1, cluster2 = assignment(points,cluster1,cluster2, plot=True)
```



[30]:
```
centroid1, centroid2 = move_centroid(cluster1,cluster2)
```


12
10
8
6
4
2
0
−2
−2
0
2
4
6
8
10
12
14

[ ]: