yangyu

April 2, 2018

1 STAT37601 Homework 1

In [2]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

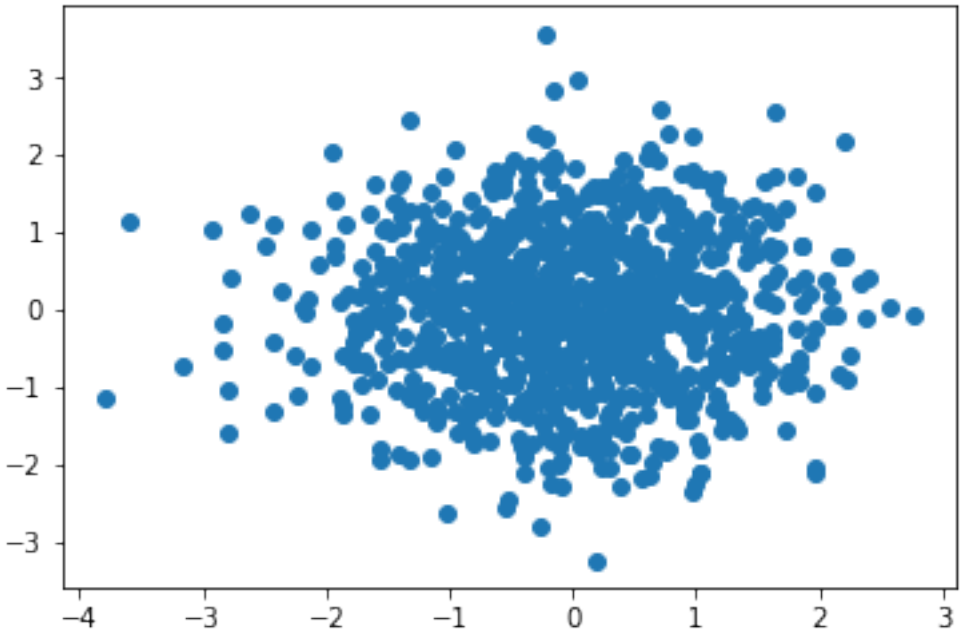
1.1 Problem 4a

In [3]:
```
np.random.seed(123)
mean = (0,0)
cov = [[1,0],[0,1]]
x = np.random.multivariate_normal(mean,cov,1000)
```

In [4]:
```
mean,cov,x
```

Out[4]:
```
((0, 0), [[1, 0], [0, 1]], array([[-1.0856306 ,  0.99734545],
        [ 0.2829785 , -1.50629471],
        [-0.57860025,  1.65143654],
        ...,
        [ 1.84745342,  0.84570124],
        [-1.11992251, -0.35929672],
        [-1.60969508,  0.01357006]]))
```

In [5]:
```
a = x[:,0]
b = x[:,1]
plt.scatter(a,b)
plt.show()
```



1.2 Problem 4b

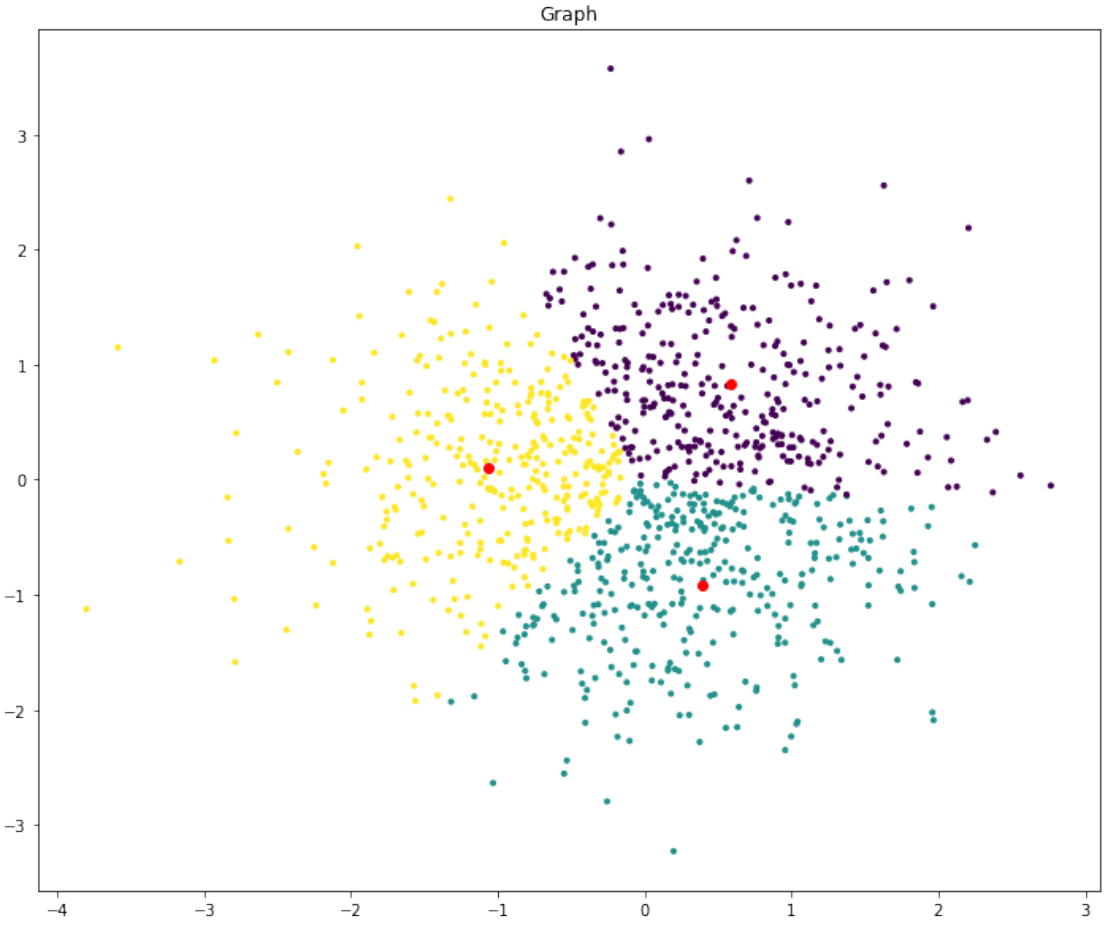
```
In [6]: kmeans = KMeans(n_clusters=3,random_state=0).fit(x)

In [7]: kmeans.cluster_centers_

Out[7]: array([[ 0.58511754,  0.83653477],
               [ 0.38742429, -0.91712467],
               [-1.0642939 ,  0.10313537]])

In [8]: y=kmeans.labels_

In [9]: plt.figure(figsize=(12,10))
        plt.title("Graph")
        plt.scatter(x[:, 0], x[:, 1], marker='.', c=y)
        plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:,1], c='r')
        plt.show()
```



1.3 Problem 4c

```
In [10]: N,D = x.shape
         new_x = np.zeros((N,D))
         for i in range(N):
             new_x[i] = 0.99*x[i] + 0.01*kmeans.cluster_centers_[y[i]]
         new_kmeans = KMeans(n_clusters=3).fit(new_x)
         new_y = new_kmeans.labels_

In [11]: plt.figure(figsize=(12,10))
         plt.title("After Shift")
         plt.scatter(new_x[:, 0], new_x[:, 1], marker='.', c=new_y)
         plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:,1], c='r')
         plt.show()
```

1.4 Problem 4d

```
In [12]: prev = x
         pred = y
         for i in range(50):
             for j in range(N):
                 prev[j] = 0.99*prev[j] + 0.01*kmeans.cluster_centers_[pred[j]]
             kmeans = KMeans(n_clusters=3).fit(prev)
             pred = kmeans.labels_
```

```
In [13]: plt.figure(figsize=(12,10))
         plt.title("After Shift")
         plt.scatter(prev[:, 0], prev[:, 1], marker='.', c=pred)
         plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:,1], c='r')
         plt.show()
```


After Shift
2
1
0
−1
−2
−2
−1
0
1
2