

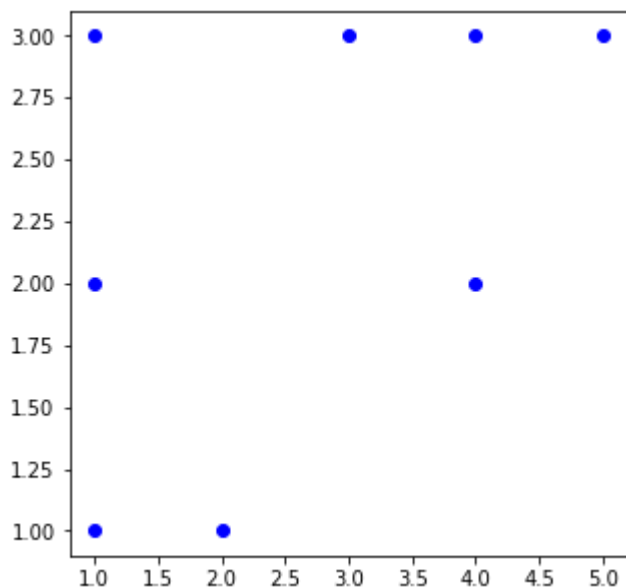
In [1]:

```
# Математика
import numpy as np
# Функция, которая считает расстояния между парами точек из двух массивов, подающихся ей на вход
from scipy.spatial.distance import cdist
# Визуализация
import matplotlib.pyplot as plt
```

In [7]:

```
# Инициализация точек
X = np.zeros((8, 2))
X[0] = 1, 3
X[1] = 3, 3
X[2] = 4, 3
X[3] = 5, 3
X[4] = 1, 2
X[5] = 4, 2
X[6] = 1, 1
X[7] = 2, 1

# Инициализация плоскости
plt.figure(figsize=(5, 5))
plt.plot(X[:, 0], X[:, 1], 'bo');
plt.show()
```



In [9]:

```
# Инициализация центроидов
np.random.seed(seed=42)
centroids = np.random.normal(loc=0.0, scale=1., size=4)
centroids= np.zeros(4)
centroids[0]= 2.5
centroids[1]= 2.5
centroids[2]= 2.0
centroids[3]= 1.75
centroids = centroids.reshape((2, 2))

cent_history = [centroids]
```

In [10]:

```
cent_history
```

Out[10]:

```
[array([[2.5 , 2.5 ],
        [2.  , 1.75]])]
```

In [11]:

```
centroids
```

Out[11]:

```
array([[2.5 , 2.5 ],
        [2.  , 1.75]])
```

In [12]:

```

for i in range(2):
    # Считаем расстояния от наблюдений до центроид
    distances = cdist(X, centroids)
    print(distances)
    # Смотрим, до какой центроиде каждой точке ближе всего
    labels = distances.argmin(axis=1)
    print(labels)
    # Положим в каждую новую центроиду геометрический центр её точек
    centroids = centroids.copy()
    centroids[0, :] = np.mean(X[labels == 0, :], axis=0)
    centroids[1, :] = np.mean(X[labels == 1, :], axis=0)

    cent_history.append(centroids)
    print(cent_history)

```

```

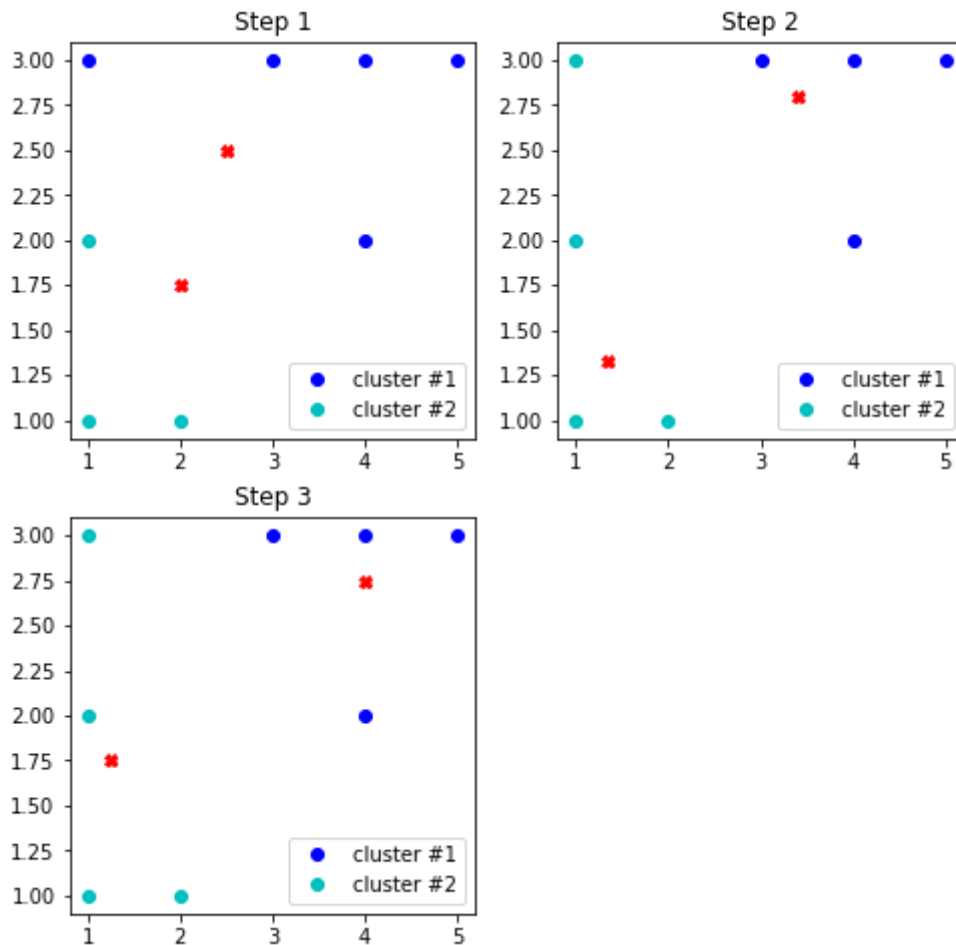
[[1.58113883 1.60078106]
 [0.70710678 1.60078106]
 [1.58113883 2.35849528]
 [2.54950976 3.25      ]
 [1.58113883 1.03077641]
 [1.58113883 2.01556444]
 [2.12132034 1.25      ]
 [1.58113883 0.75      ]]
[0 0 0 0 1 0 1 1]
[array([[2.5 , 2.5 ],
        [2. , 1.75]]), array([[3.4      , 2.8      ],
        [1.33333333, 1.33333333]])]
[[2.40831892 1.69967317]
 [0.4472136  2.3570226 ]
 [0.63245553 3.14466038]
 [1.61245155 4.02768199]
 [2.52982213 0.74535599]
 [1.         2.74873708]
 [3.         0.47140452]
 [2.28035085 0.74535599]]
[1 0 0 0 1 0 1 1]
[array([[2.5 , 2.5 ],
        [2. , 1.75]]), array([[3.4      , 2.8      ],
        [1.33333333, 1.33333333]]), array([[4. , 2.75],
        [1.25, 1.75]])]

```

In [14]:

```
plt.figure(figsize=(8, 8))
for i in range(3):
    distances = cdist(X, cent_history[i])
    labels = distances.argmin(axis=1)

    plt.subplot(2, 2, i + 1)
    plt.plot(X[labels == 0, 0], X[labels == 0, 1], 'bo', label='cluster #1')
    plt.plot(X[labels == 1, 0], X[labels == 1, 1], 'co', label='cluster #2')
    plt.plot(cent_history[i][:, 0], cent_history[i][:, 1], 'rX')
    plt.legend(loc=0)
    plt.title('Step {}'.format(i + 1));
```



In []: