K-means Clustering from scratch

In [1]:

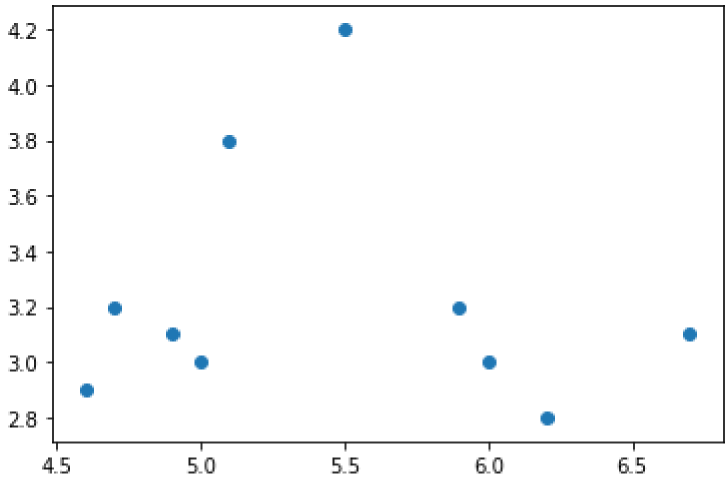
```
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
X = np.array([
    [5.9, 3.2],
    [4.6, 2.9],
    [6.2, 2.8],
    [4.7, 3.2],
    [5.5, 4.2],
    [5.0, 3.0],
    [4.9, 3.1],
    [6.7, 3.1],
    [5.1, 3.8],
    [6, 3]
])

centers = np.array([[1, 1], [2, 2], [3, 3]])
```

In [4]:

```
plt.scatter(*X.T)
plt.show()
```

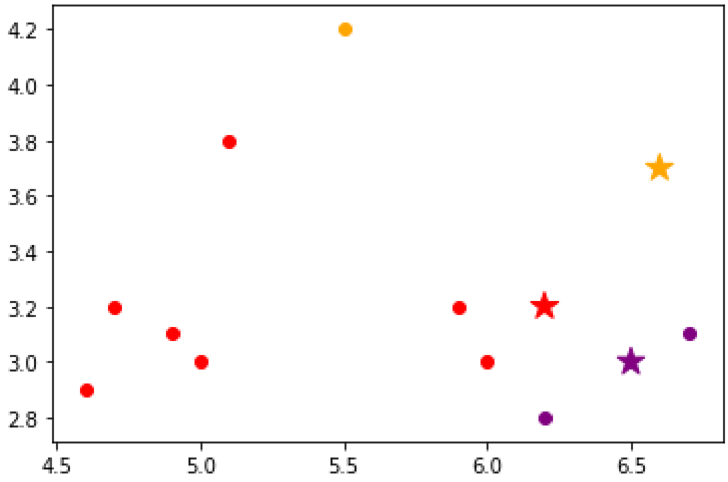


In [5]:

```
def assign_clusters(X, cluster_centers):
    n = X.shape[0]
    K = len(cluster_centers)
    dists = np.zeros(shape = (n, K))
    for i, center in enumerate(cluster_centers):
        dists[:,i] = np.sum((X - center)**2, axis = 1)**(1/2)
    labels = dists.argmin(axis = 1)
    return labels

def compute_cluster_centers(X, cluster_labels):
    K = np.unique(cluster_labels).shape[0]
    cluster_centers = np.array([X[cluster_labels == k].mean(axis = 0) for k in range(K
)])
    return cluster_centers
```

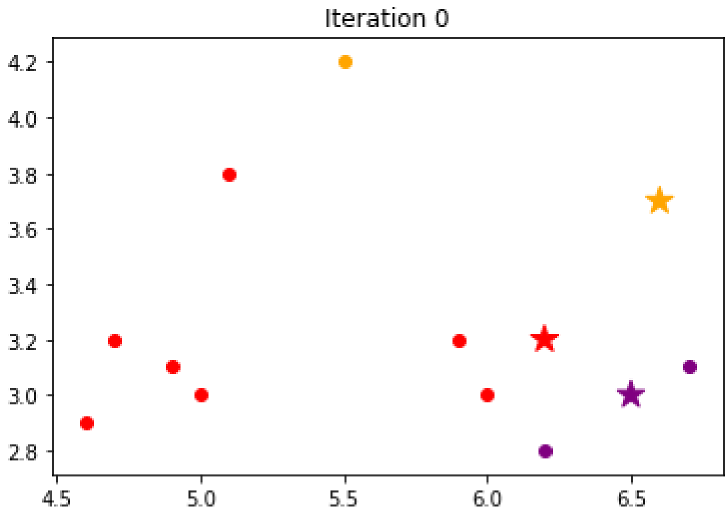
In [6]:

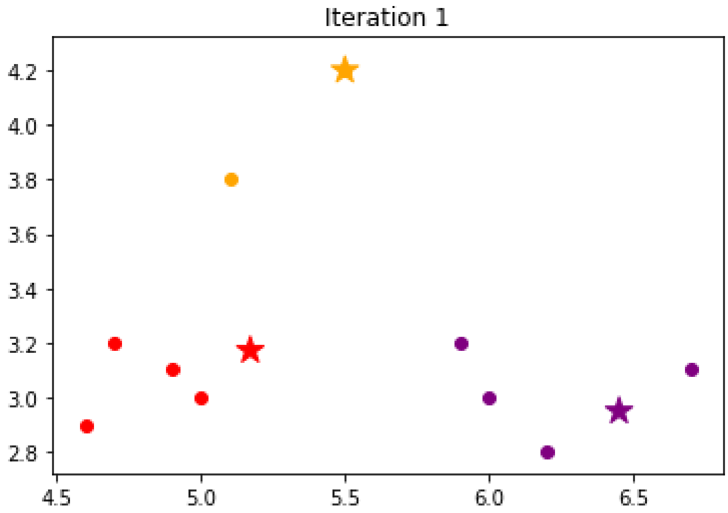
```
colors = ['red', 'purple', 'orange']
cluster_centers = np.array([[6.2, 3.2], [6.5, 3.0], [6.6, 3.7]])
cluster_labels = assign_clusters(X, cluster_centers)
plt.scatter(*X[cluster_labels == 0].T, c = "red")
plt.scatter(*X[cluster_labels == 1].T, c = "purple")
plt.scatter(*X[cluster_labels == 2].T, c = "orange")
plt.scatter(*cluster_centers.T, c = colors, s = 200, marker = "*")
plt.show()
```

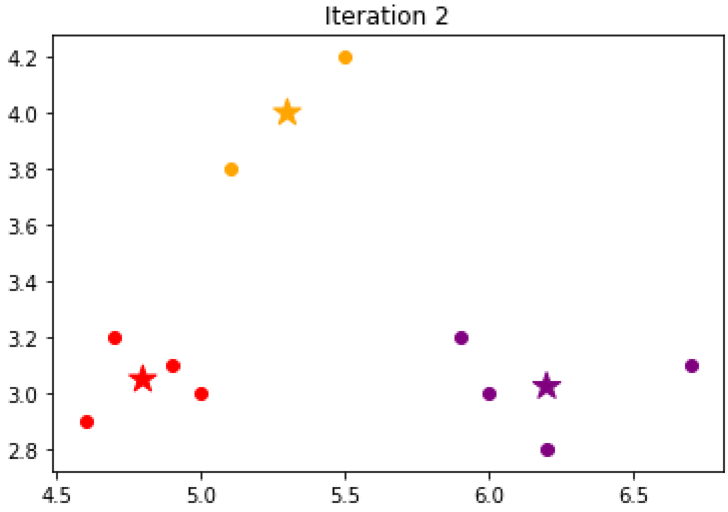


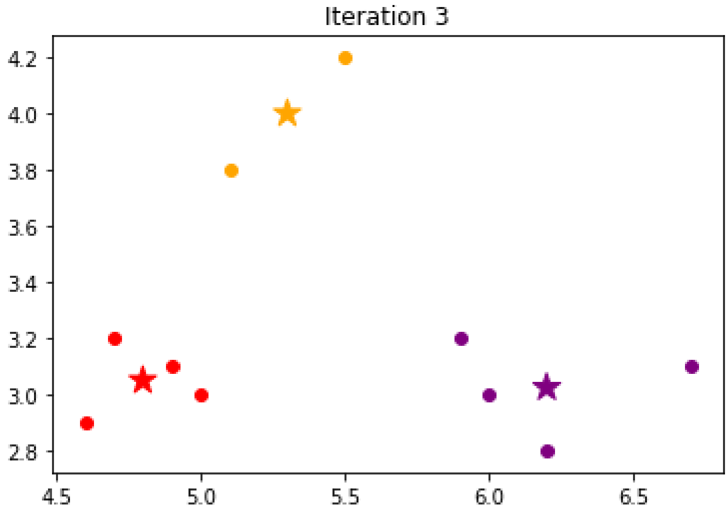
In [7]:

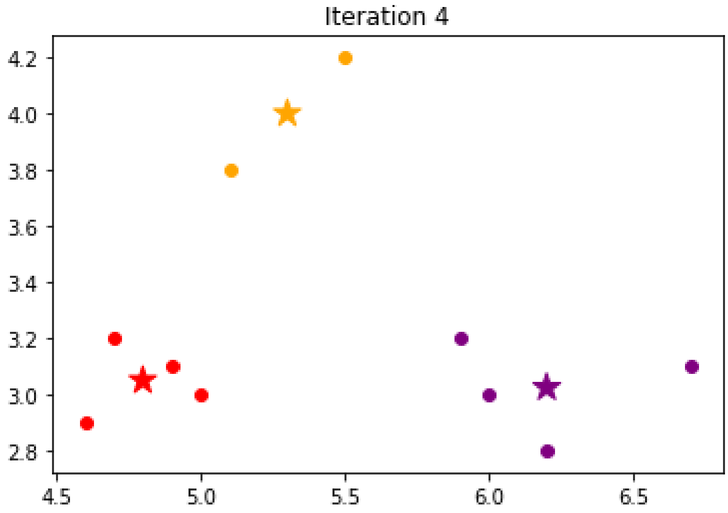
```
colors = ['red', 'purple', 'orange']
cluster_centers = np.array([[6.2, 3.2], [6.5, 3.0], [6.6, 3.7]])

for t in range(5):
    cluster_labels = assign_clusters(X, cluster_centers)
    plt.scatter(*X[cluster_labels == 0].T, c = "red")
    plt.scatter(*X[cluster_labels == 1].T, c = "purple")
    plt.scatter(*X[cluster_labels == 2].T, c = "orange")
    plt.scatter(*cluster_centers.T, c = colors, s = 200, marker = "*")
    plt.title(f"Iteration {t}")
    cluster_centers = compute_cluster_centers(X, cluster_labels)
    plt.show()
```


Iteration 0


Iteration 1


Iteration 2


Iteration 3
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
4.5
5.0
5.5
6.0
6.5


Iteration 4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
4.5
5.0
5.5
6.0
6.5