

12-03_Kmeans

December 6, 2021

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

```
[2]: from sklearn.datasets import load_iris
data = load_iris()
```

```
[4]: data.keys()
```

```
[4]: dict_keys(['data', 'target', 'frame', 'target_names', 'DESCR', 'feature_names',
'filename'])
```

```
[9]: data['data'].shape
```

```
[9]: (150, 4)
```

```
[5]: X = data['data'][:,(0,2)]
X
```

```
[5]: array([[5.1, 1.4],
[4.9, 1.4],
[4.7, 1.3],
[4.6, 1.5],
[5. , 1.4],
[5.4, 1.7],
[4.6, 1.4],
[5. , 1.5],
[4.4, 1.4],
[4.9, 1.5],
[5.4, 1.5],
[4.8, 1.6],
[4.8, 1.4],
[4.3, 1.1],
[5.8, 1.2],
[5.7, 1.5],
[5.4, 1.3],
[5.1, 1.4],
[5.7, 1.7],
```

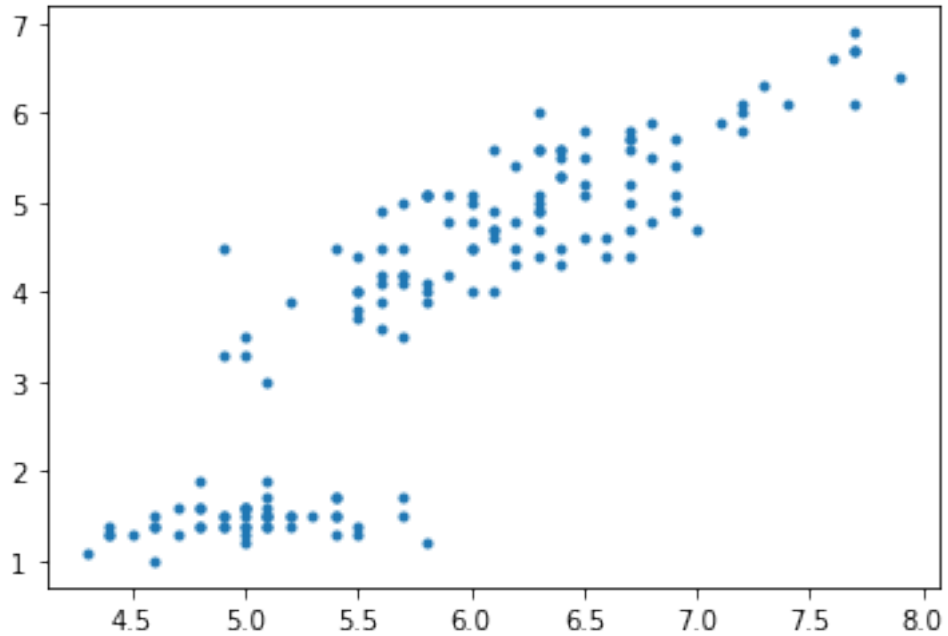
[5.1, 1.5],
[5.4, 1.7],
[5.1, 1.5],
[4.6, 1.],
[5.1, 1.7],
[4.8, 1.9],
[5. , 1.6],
[5. , 1.6],
[5.2, 1.5],
[5.2, 1.4],
[4.7, 1.6],
[4.8, 1.6],
[5.4, 1.5],
[5.2, 1.5],
[5.5, 1.4],
[4.9, 1.5],
[5. , 1.2],
[5.5, 1.3],
[4.9, 1.4],
[4.4, 1.3],
[5.1, 1.5],
[5. , 1.3],
[4.5, 1.3],
[4.4, 1.3],
[5. , 1.6],
[5.1, 1.9],
[4.8, 1.4],
[5.1, 1.6],
[4.6, 1.4],
[5.3, 1.5],
[5. , 1.4],
[7. , 4.7],
[6.4, 4.5],
[6.9, 4.9],
[5.5, 4.],
[6.5, 4.6],
[5.7, 4.5],
[6.3, 4.7],
[4.9, 3.3],
[6.6, 4.6],
[5.2, 3.9],
[5. , 3.5],
[5.9, 4.2],
[6. , 4.],
[6.1, 4.7],
[5.6, 3.6],
[6.7, 4.4],

[5.6, 4.5],
[5.8, 4.1],
[6.2, 4.5],
[5.6, 3.9],
[5.9, 4.8],
[6.1, 4.],
[6.3, 4.9],
[6.1, 4.7],
[6.4, 4.3],
[6.6, 4.4],
[6.8, 4.8],
[6.7, 5.],
[6. , 4.5],
[5.7, 3.5],
[5.5, 3.8],
[5.5, 3.7],
[5.8, 3.9],
[6. , 5.1],
[5.4, 4.5],
[6. , 4.5],
[6.7, 4.7],
[6.3, 4.4],
[5.6, 4.1],
[5.5, 4.],
[5.5, 4.4],
[6.1, 4.6],
[5.8, 4.],
[5. , 3.3],
[5.6, 4.2],
[5.7, 4.2],
[5.7, 4.2],
[6.2, 4.3],
[5.1, 3.],
[5.7, 4.1],
[6.3, 6.],
[5.8, 5.1],
[7.1, 5.9],
[6.3, 5.6],
[6.5, 5.8],
[7.6, 6.6],
[4.9, 4.5],
[7.3, 6.3],
[6.7, 5.8],
[7.2, 6.1],
[6.5, 5.1],
[6.4, 5.3],
[6.8, 5.5],

```
[5.7, 5. ],  
[5.8, 5.1],  
[6.4, 5.3],  
[6.5, 5.5],  
[7.7, 6.7],  
[7.7, 6.9],  
[6. , 5. ],  
[6.9, 5.7],  
[5.6, 4.9],  
[7.7, 6.7],  
[6.3, 4.9],  
[6.7, 5.7],  
[7.2, 6. ],  
[6.2, 4.8],  
[6.1, 4.9],  
[6.4, 5.6],  
[7.2, 5.8],  
[7.4, 6.1],  
[7.9, 6.4],  
[6.4, 5.6],  
[6.3, 5.1],  
[6.1, 5.6],  
[7.7, 6.1],  
[6.3, 5.6],  
[6.4, 5.5],  
[6. , 4.8],  
[6.9, 5.4],  
[6.7, 5.6],  
[6.9, 5.1],  
[5.8, 5.1],  
[6.8, 5.9],  
[6.7, 5.7],  
[6.7, 5.2],  
[6.3, 5. ],  
[6.5, 5.2],  
[6.2, 5.4],  
[5.9, 5.1]])
```

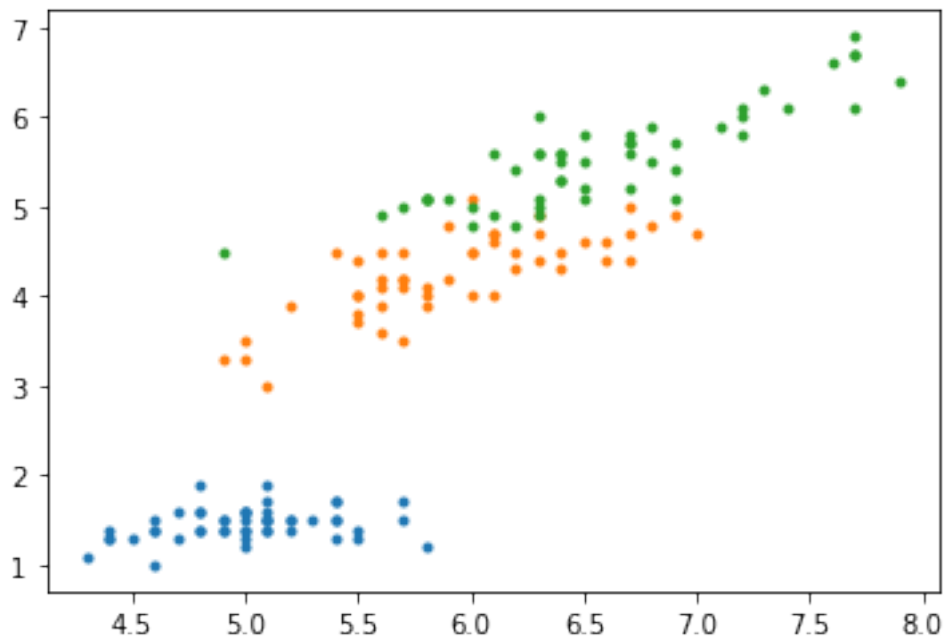
```
[6]: plt.plot(X[:,0],X[:,1],'.')
```

```
[6]: [<matplotlib.lines.Line2D at 0x7f390e68b760>]
```



```
[8]: target = data['target']  
plt.plot(X[target==0,0],X[target==0,1],'.')  
plt.plot(X[target==1,0],X[target==1,1],'.')  
plt.plot(X[target==2,0],X[target==2,1],'.')
```

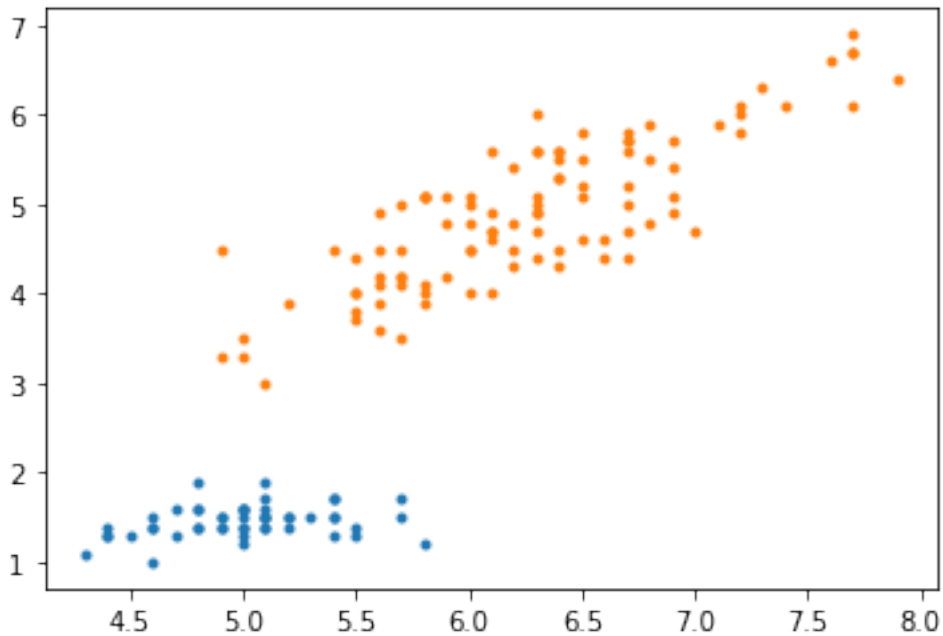
```
[8]: [<matplotlib.lines.Line2D at 0x7f390e570f70>]
```



```
[10]: is_species1 = target==0
```

```
[11]: plt.plot(X[is_species1,0], X[is_species1,1], '.')  
plt.plot(X[~is_species1,0], X[~is_species1,1], '.')
```

```
[11]: [<matplotlib.lines.Line2D at 0x7f390e4cdc10>]
```



```
[28]: m1 = X[np.random.randint(0, X.shape[0]),:]  
m2 = X[np.random.randint(0, X.shape[0]),:]
```

```
[29]: m1, m2
```

```
[29]: (array([4.8, 1.6]), array([6.7, 4.4]))
```

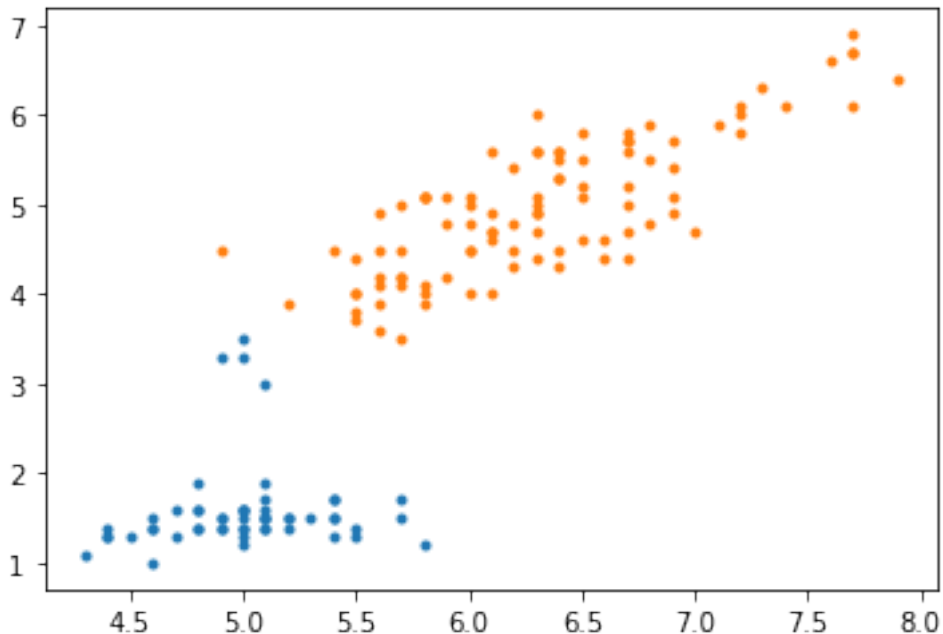
```
[37]: for i in range(1000):  
    in_m1 = np.linalg.norm(X - m1, axis=1) <= np.linalg.norm(X - m2, axis=1)  
    m1 = np.mean(X[in_m1,:], axis=0)  
    m2 = np.mean(X[~in_m1,:], axis=0)
```

```
[38]: m1, m2
```

```
[38]: (array([5.00555556, 1.5962963 ]), array([6.31458333, 4.97395833]))
```

```
[39]: in_m1 = np.linalg.norm(X - m1, axis=1) <= np.linalg.norm(X - m2, axis=1)
plt.plot(X[in_m1,0], X[in_m1,1], '.')
plt.plot(X[~in_m1,0], X[~in_m1,1], '.')
```

```
[39]: [<matplotlib.lines.Line2D at 0x7f390e39ca60>]
```



```
[40]: X = data['data']
```

```
[66]: m1 = X[np.random.randint(0, X.shape[0]),:]
m2 = X[np.random.randint(0, X.shape[0]),:]
assert np.any(m1 != m2)
```

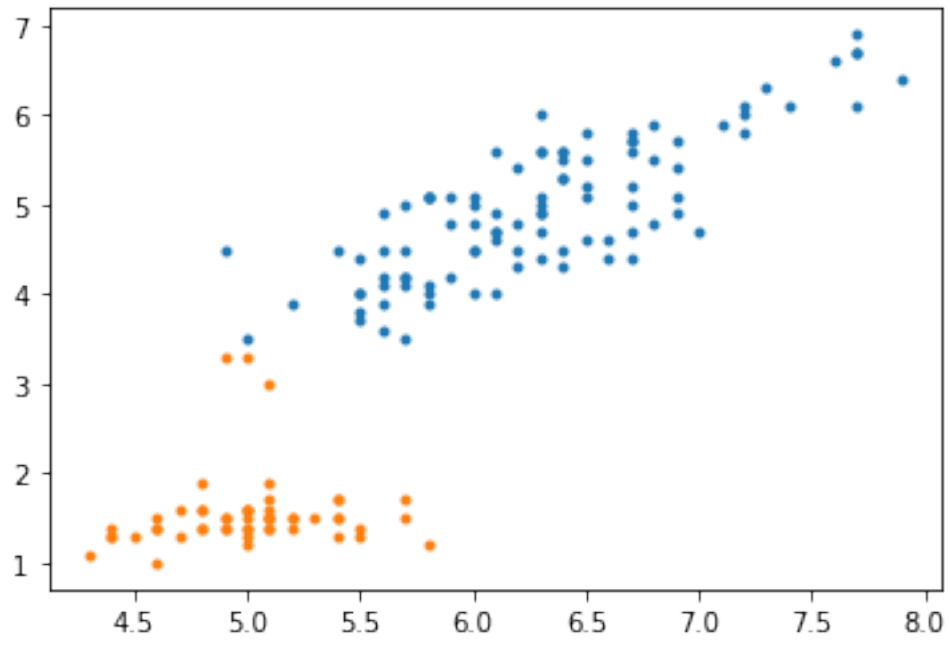
```
[67]: m1, m2
```

```
[67]: (array([6.4, 3.2, 4.5, 1.5]), array([5.2, 4.1, 1.5, 0.1]))
```

```
[70]: for i in range(1000):
    in_m1 = np.linalg.norm(X - m1, axis=1) <= np.linalg.norm(X - m2, axis=1)
    m1 = np.mean(X[in_m1,:], axis=0)
    m2 = np.mean(X[~in_m1,:], axis=0)
```

```
[71]: in_m1 = np.linalg.norm(X - m1, axis=1) <= np.linalg.norm(X - m2, axis=1)
plt.plot(X[in_m1,0], X[in_m1,2], '.')
plt.plot(X[~in_m1,0], X[~in_m1,2], '.')
```

```
[71]: [<matplotlib.lines.Line2D at 0x7f390e1ebd90>]
```



```
[ ]: X = data['data'][:,(0,2)]
```