

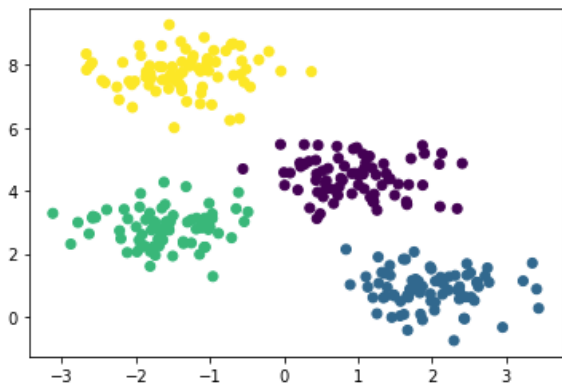
Kmeans 실습

Kmeans 실습 #1 (분류)

- 데이터셋: 합성 데이터
- 학습/시험 데이터: X
- 학습/시험 데이터 라벨: Y

① 데이터 만들기

```
1 from sklearn.datasets.samples_generator import make_blobs
2 import matplotlib.pyplot as plt
3
4 # DATA 1: 합성데이터 (Synthetic Data)
5 X, Y = make_blobs(n_samples=300, centers=4, cluster_std=0.60, random_state=0)
6 plt.scatter(X[:,0], X[:,1], c = Y);
7
```



Kmeans 실습 #1 (분류)

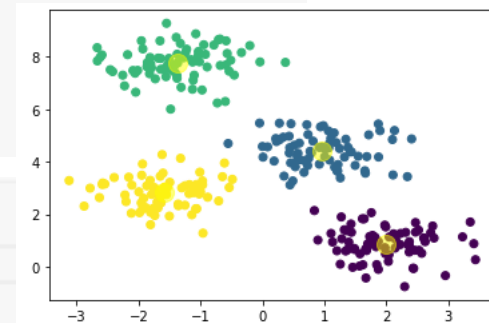
- 데이터셋: 합성 데이터
- 학습/시험 데이터: X
- 학습/시험 데이터 라벨: Y

① Kmeans API 호출 [\[Manual\]](#)

n_cluster : 군집의 개수 K, random_state = 0 : 초기값 선정에 영향을 줌

```
1 from sklearn.cluster import KMeans
2
3 kmeans = KMeans(n_clusters=4, random_state=0)
4 kmeans.fit(X)
5
```

```
1 Y_pred = kmeans.predict(X)
2 |
3 plt.scatter(X[:, 0], X[:, 1], c=Y_pred)
4 centers = kmeans.cluster_centers_
5 plt.scatter(centers[:, 0], centers[:, 1], c='Yellow', s=200, alpha=0.5);
```

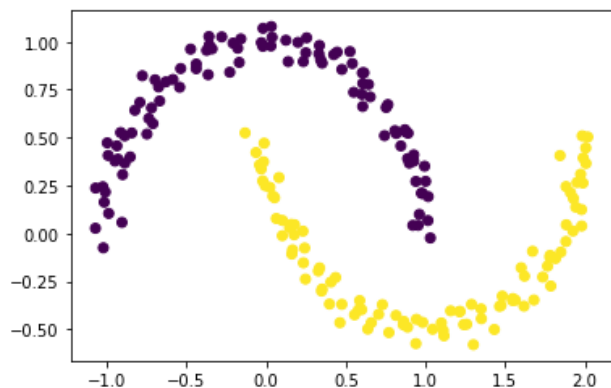


Kmeans 실습 #2 (분류)

- 데이터셋: 합성 데이터
- 학습/시험 데이터: X
- 학습/시험 데이터 라벨: Y

① 데이터 만들기

```
1 from sklearn.datasets import make_moons
2 import matplotlib.pyplot as plt
3
4 X, Y = make_moons(n_samples = 200, noise=.05, random_state=0)
5 plt.scatter(X[:, 0], X[:, 1], c=Y );
```



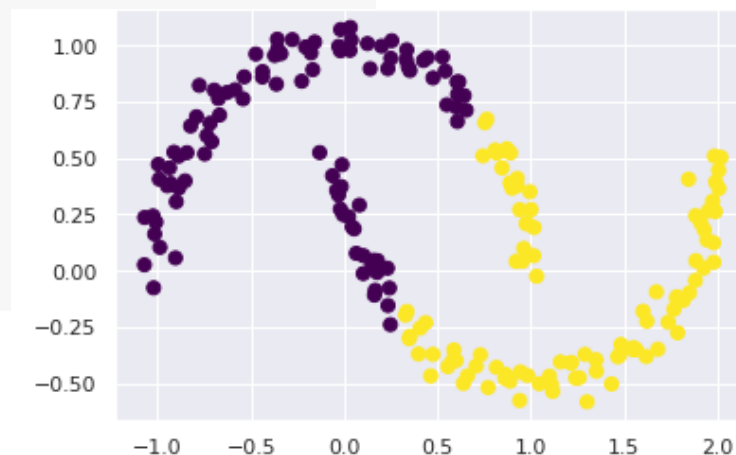
Kmeans 실습 #2 (분류)

- 데이터셋: 합성 데이터
- 학습/시험 데이터: X
- 학습/시험 데이터 라벨: Y

① Kmeans API 호출 [\[Manual\]](#)

n_cluster : 군집의 개수 K, random_state = 0 : 초기값 선정에 영향을 줌

```
1 from sklearn.cluster import KMeans
2
3 # 학습
4 kmeans = KMeans(n_clusters=2, random_state=0)
5 kmeans.fit(X)
6
7 # 테스트
8 Y_pred = kmeans.predict(X)
9 plt.scatter(X[:, 0], X[:, 1], c=Y_pred );
```



Kmeans 실습 #3 (분류)

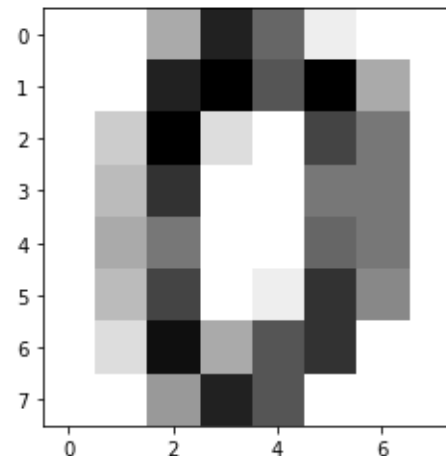
- 데이터셋: 손글씨 데이터 (digits)
- 학습/시험 데이터: X
- 학습/시험 데이터 라벨: Y

① 데이터 만들기

```
1 from sklearn.datasets import load_digits
2 digits = load_digits()
3
4 X = digits.data
5 Y = digits.target
6
7 print(digits.data.shape)
8 print(digits.target)
```

```
↳ (1797, 64)
   [0 1 2 ... 8 9 8]
```

```
1 img = digits.data.reshape(1797,8,8)
2 plt.imshow(img[0],cmap=plt.cm.binary)
```



Kmeans 실습 #3 (분류)

- 데이터셋: 손글씨 데이터 (digits)
- 학습/시험 데이터: X
- 학습/시험 데이터 라벨: Y

① Kmeans API 호출 [\[Manual\]](#)

n_cluster : 군집의 개수 K



```
1 kmeans = KMeans(n_clusters=10)
2 kmeans.fit(X)
```



```
1 centers = kmeans.cluster_centers_.reshape(10, 8, 8)
2 plt.imshow(centers[0], cmap=plt.cm.binary)
```

