1. ML에서 grid search란?

1) 머신 러닝 모델에서 최적의 하이퍼파라미터를 찾기 위해 사용되는 탐색 방법

2) Grid Search는 하이퍼파라미터 조합을 체계적으로 탐색하여 가장 좋은 성능을 내는 설정을 찾음

2. Random forest의 중요한 하이퍼 파라미터 예

1) 결정 트리의 개수 (n_estimators): 모델이 생성할 트리의 수. 보통은 결정 트리의 개수가 많은게 좋지만 계산 비용 증가. 일반적으로 100-200개로 정함

2) 최대 깊이 (max_depth): 개별 트리의 최대 깊이. 트리가 너무 깊으면 overfitting 위험. 너무 얕으면 모델이 overfitting

3) leaf 노드의 최소 샘플수 (min_samples_leaf): leaf 노드에 있는 최소 샘플 수. 이 값을 크게 하면 overfitting 방지. 너무 작은 값은 overfitting 생김

4) 분할을 위한 최소 샘플수 (min_samples_split): 노드를 분할하기 위한 최소 샘플 수. 기본값은 2. 이 값이 크면 더 간단한 ML 모델. Overfitting 방지

3. Iris dataset에 대하여 random forest 분류기 grid search 예

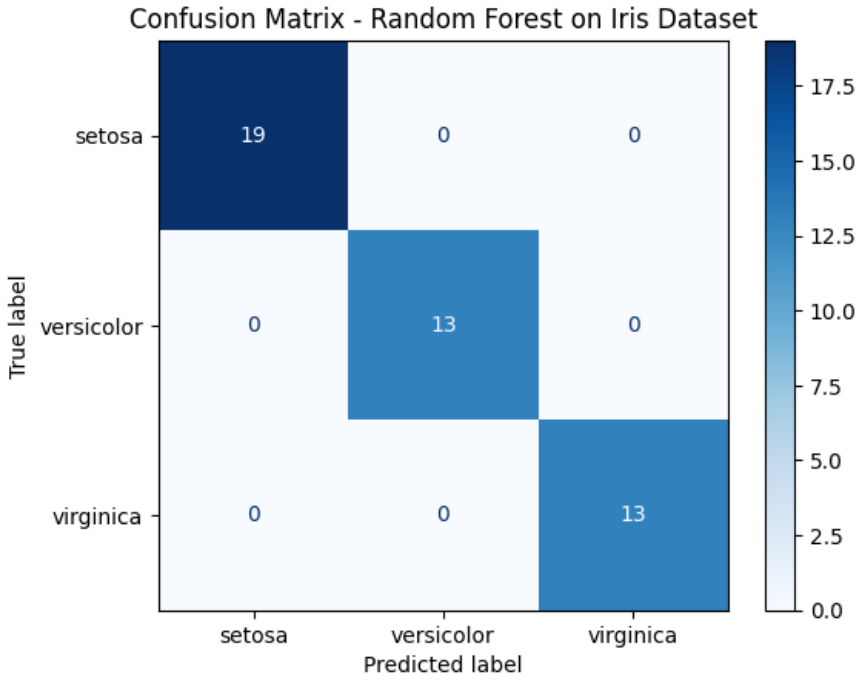
- cross validation으로 grid search를 통해 최적의 hyper parameter 설정

```
# 하이퍼파라미터 그리드 정의
param_grid = {
    'n_estimators': [10, 50, 100, 200],      # 트리의 개수
    'max_depth': [None, 10, 20, 30],         # 트리의 최대 깊이
    'min_samples_split': [2, 5, 10],         # 노드 분할을 위한 최소 샘플 수
    'min_samples_leaf': [1, 2, 4],           # 리프 노드의 최소 샘플 수
}
```

- 총 144개 경우의 수, 5-fold CV 수행 (시간이 꽤 오래걸림)

```
Fitting 5 folds for each of 144 candidates, totalling 720 fits
최적의 하이퍼파라미터: {'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
```

- 결과

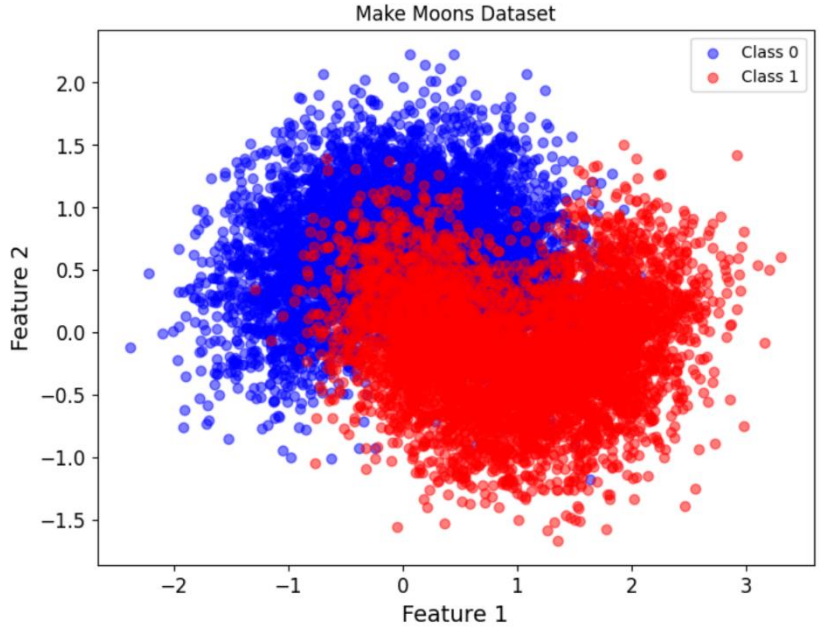


---------- 연습 문제 ----------

Moons 데이터셋에 decision tree를 훈련시키고 tuning까지 실행한 예

[handson-ml2/06_decision_trees.ipynb at master · ageron/handson-ml2 · GitHub]

1. make_moon(n_samples=1000, noise=0.4)를 사용해 dataset을 생성



2. train_test_split()를 사용해 훈련 세트와 테스트 분리. Test 데이터 20%가 되도록 설정

3. Decision tree 분류기의 최적의 매개변수를 찾기 위해 교차 검증과 함께 grid search 수행

DecisionTreeClassifier — scikit-learn 1.5.2 documentation

```
params = {'max_leaf_nodes': list(range(2, 100)),
'min_samples_split': [2, 3, 4]}
```

4 찾은 매개 변수를 사용해 전체 훈련 세트에 대해서 모델을 훈련시키고 테스트 세트에서 성능 측정

정확도가 86.9% 나오는지 확인