1. Sklearn datasets의 iris data를 불러와서, Pandas DataFrame으로 바꾸세요.
2. 1번의 데이터에 결측치가 있는지 체크하고, 결측치가 존재하면 평균값을 넣으세요.
3. 1번의 데이터의 평균, 표준편차, 4분위 수를 구하세요. Label의 값을 1이면 1, 1이 아니면 0으로 수정하세요.
4. MinMaxScaler()을 사용해 전처리를 한뒤, train set과 test set을 train\_test\_split을 사용하여 test size을 0.25만큼 분리하세요
5. 위에서 만든 train set을 RandomForestClassifier기본 파라메터를 사용하여 학습하고, 정확도(accuracy)를 구하세요.
6. train set을 RandomForestClassifier 기본 파라메터를 사용하여 학습하고, 오차행렬, 정밀도와 재현율, F1 스코어를 구하고 ROC커브를 그리세요.
7. 6번의 결과 중 2종 오류를 범하기 않기 위해서 주의 깊게 봐야하는 것은 무엇인가요?
8. 3번의 데이터를 Get\_dummies를 사용해 iris data의 label을 범주형데이터로 바꿔서 기존 df에 붙이고, 기존의 label은 df에서 삭제하세요.
9. 데이터를 GridSearchCV를 사용해서 RandomForestClassifier을 교차검증과 최적 파라메터 튜닝을 수행하여, 정확도를 다시 구하세요.

파라메터 예시 : params = { 'n\_estimators' : [10, 100],

'max\_depth' : [3, 5, 7,],

'min\_samples\_leaf' : [8, 12],

'min\_samples\_split' : [8, 16]

}

1. 9번에서 학습한 모델 결과의 각 피쳐별 중요도를 그래프로 보이세요.

(rf\_clf.feature\_importances\_)