

Assignment : Supervised learning with decision trees

1. 데이터셋

- breast_cancer dataset

```
from sklearn.datasets import load_breast_cancer
```

2. 구현

- > Data load
- > Data split (Training / Validation / Test) (Hint: train_test_split을 2번 시행)
- > Training (DecisionTreeClassifier)
- > Hyperparameter 2개를 골라 해당 값들을 바꿔가며 validation set에 대한 성능 비교
- > 최적의 hyperparameter 선정 후 해당 모델의 Test Set에 대한 성능 확인
- > Evaluation (Metric: Accuracy)

3. 레포트에 반드시 포함할 것

- input data의 Feature의 수, target class 수, Train/Val/Test 각각의 instance(sample) 개수
- 2개 hyperparameter의 의미
- Hyperparameter 조정에 따른 train / val set에 대한 성능 표 or 그래프
- 선정된 최적 hyperparameter값 조합은 무엇인지
- 선정된 최종 모델의 Test Set에 대한 성능(accuracy 및 confusion matrix도 표현)

4. 과제 제출

- eclass에 ipynb 파일 1개 제출
- ipynb 파일 내에 모든 것이 self-contained 되어 있어야 함
 - Run 후의 결과가 나타나 있어야 함
 - 진행한 내용에 대한 설명 (별도의 보고서를 작성하지 않고 해당 파일에 모든 설명을 적어주세요)