

Assignment: Decision_Tree (윤세영 교수님 연구실)

담당조교:김정훈 (junghunkim@kaist.ac.kr)

-데이터 셋 설명

Data set: Credit.csv

Input features: ['Age','Income','Limit','Cards','Student','Education']

Target features: ['Balance']

-과제설명

실습시간에 구현한 Decision tree 모델을 Bagging 을 이용하여 모델을 확장해 보는 과제입니다.

코드에서 정의된 Subsample 과 Bagging 함수를 추가적으로 이용하면 Bagging 모델을 구현할수 있습니다.

코드의 빈칸을 채워 코드를 완성해주시면 됩니다.(Fill out here 부분을 채워주시면 됩니다)

```
def Bagging(repeat,ratio):
    training_set,test_set=loadDataset('Credit.csv')
    predictions=[]
    random.seed(3)
    for i in range(repeat):
        pred_temp=[] #list for prediction of each model
        train_subset=Subsample(training_set,ratio) #using a bootstrap for training set.

        tree = #####fill out here ### we need to using tree_building function

        for employee in test_set: #loop for each test data
            pred_temp.append(##### get a prediction of each test data

        predictions.append(pred_temp) #append prediction of each model

    predictions_avg= #####fill out here ##### get an average prediction of models

    error = getError(test_set, predictions_avg)
    print('Error(RMSE): ' + repr(error))
```

-제출물

완성된 코드(file.ipynb)와 보고서 한페이지를 제출해 주시면 됩니다

보고서에는 코드부분에서 직접 작성하신 부분에대한 설명과 hyperparameter 인 repeat과 ratio를 임의로 바꿔가며 얻은 5개 정도의 error(RMSE)를 표로 작성해 주시면 됩니다. (단, tree_building 함수의 min_size 는 10으로 고정)

또한 최종적으로 얻은 가장좋은 성능을 보이는 모델에대한 error와 plot 그림(x축: true value, y축: predicted value) 를 보고서에 넣어주시면 됩니다.

-채점

채점은 코드(및 직접 작성한 부분에 대한 설명) 5점과 보고서 결과물 5점만점으로 채점이 됩니다.