

Build a decision tree, and then classify the test set using it

2018007956 김채아

1. Environment Setting

- A. OS: Windows
- B. Languages: Python

2. How to compile and execute the code

- Command: python [executable file name] [training file name] [test file name] [output file name]

```
PS C:\Users\LG\Desktop\Data_Science\Assignment2> python dt.py dt_train1.txt dt_test1.txt dt_result1.txt
Minimum Gain: 0.06910000000000001
PS C:\Users\LG\Desktop\Data_Science\Assignment2>
```

3. Specification for running the code

- A. dt.py, dt_train.txt, dt_test.txt 모두 같은 경로에 위치하고 있고, 결과 파일 dt_result.txt도 같은 위치에 생성됨
- B. Minimum Gain은 전체 training data의 개수에 0.00007을 곱한 수로 결정됨. 곱해지는 값은 실험적으로 결정된 값이며, 이 값에 따라 classification 정확도가 달라질 수 있음
 - i. Pruning 유/무 Decision tree 비교
(아래는 $\text{len}(\text{training_data}) \times 0.0001$ 을 minimum gain으로 설정한 경우)

1. Pruning 안 한 경우

[illegible]

2. Pruning 한 경우

[illegible]