

# Machine Problem #1

데이터마이닝

마감: 11월 18일

WEKA를 이용하여 decision tree를 학습하되 10-fold cross-validation을 통해 tree의 size가 최적인 모델을 찾는다 (강의자료 03의 pp. 35 – 39 참조). Tree의 적절한 size를 찾기 위해서는 WEKA에서 ConfidenceFactor의 설정 값을 여러 가지로 바꾸어 가면서 (0.001부터 시작하여 그 값을 증가시켜 가며 실험해 볼 것을 권장함) pruning 정도를 조정해 보면 된다.

대상 데이터로는 UCI Machine Learning Repository에서 제공하는 Car Evaluation Data Set을 이용한다. 학습의 과제는 대상 차량을 구매하는 것이 적합한지 여부를 4 단계로 (unacceptable, acceptable, good, very good) 나누어 예측하는 것이다.

(<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>)

전체 데이터 중 2/3는 model selection 과정까지 포함한 학습에 사용하고, 나머지 1/3은 테스트 용으로 사용한다.

제출할 결과 보고서에는 적어도 아래 내용이 포함되어야 한다.

- 사용 데이터의 특징에 대한 설명
- 최적 모델의 tree 크기 (노드 수)
- 강의자료 03의 38 페이지에 보인 것과 같은 그래프
- 테스트 데이터에서의 분류 정확도

결과 보고서는 한글 또는 워드 파일로 작성하고, PLMS 게시판의 과제 icon을 클릭하여 온라인으로 제출한다.

파일명은 **학번\_이름**의 형식으로 부여하여야 함. (예: **201224278\_홍길동**)