

## 2016년 봄학기 인공지능 기말과제

### 개요

- 두 개의 public domain dataset을 사용하여
- 4개 이상의 classifier를 사용하여, 각 classifier의 성능을 비교 분석하고
- 그 dataset에 대한 최대한의 모든 분석을 해보도록 한다.
- 단순히 프로그램을 돌린 결과만 제시하는 것이 아니라, 그 dataset에 대한 본질을 알아내기 위하여 도구를 사용하는 것이다. 본인이 가지고 있는 분석력을 최대한 보여주기 바람.

### 마감

- 6월19일(일요일) 밤 12시.
- 블랙보드에 올릴 것 (자세한 제출 방법은 블랙보드에 공지 예정)

### 성적 grading

- 기말고사는 원래 40%였는데, 그것을 기말시험 10%, 기말프로젝트 30%의 비율.

---

### Datasets

- UCI Machine Learning Repository <http://archive.ics.uci.edu/ml/> 에서 두 개의 dataset을 선택한다.
- # attributes 와 # instances 가 충분하여야 하며, data type은 multi-variate이 하나 이상 포함되도록 한다.

### Tools

- 수업시간에 소개한 Weka를 사용하여도 됨.
- 혹은, 다른 데이터 분석, 기계학습 도구를 사용하여도 됨.
- 혹은, 본인 스스로 코드를 만들어서 분석하여도 됨.

### 분석과 평가

- 최소 4개의 서로 다른 classifier가 사용되어 결과의 '비교 분석'이 이루어 져야 함.
- 도구의 분석 결과를 사용함으로써, dataset 자체에 대한 분석을 시도하도록 하여야 함.
- 최종평가는 cross-validation을 사용하여야 함.
- 평가 분석에 overfitting 의 관점이 사용되어야 함.
- zeroR, oneR의 결과를 baseline 으로 사용할 것.

### 보고서

- 한 페이지짜리 실험 요약이 들어가야 함.
- 데이터에 대한 설명, 그리고 본인이 이 데이터셋을 선택한 이유를 설명
- 실험 설계 및 방법. 진행 내용.
- 비교 분석 결과
- 결론.

참고.

- 각 dataset을 보면, 그 dataset을 사용한 분석 논문 목록이 있다. 같은 dataset을 사용한 기존의 논문들을 참조하여, 비교하고 그 실험 계획과 방법, 결과 분석의 내용을 참조해 보는 것을 권장함. 이 경우에는 참고문헌 표시를 명확하게 하고, 자신의 내용과 구분하여 기술할 것.