

# 결정트리 학습

# Decision Tree

이영석

# 결정트리 학습법(Decision Tree Learning)

- 어디에 쓰는가?
  - 타이타닉호에서 살아남을 사람은 누구일까?
  - 다음 달 번호이동할 가입자는 누구일까?
  - 카드 연체할 사람은 누구일까?
  - 등등 예측관련 서비스
- 왜 ?
  - 쉬워서
  - 계산이 간단해서
  - 실수(수치형)/정수(범주형) 값으로 예측가능

# 좀더 일반적인 설명

- 결정트리 분석 방법 종류
  - 통계학: CARD, CHAID 알고리즘
  - 기계학습: (엔트로피, 정보이득 기반) ID3, C4.5, C5.0 알고리즘
- 지도분류학습의 종류
- 분류트리
  - 0 또는 1
  - 0, 1, 2, 3, 4 범주로 예측
- 회귀트리
  - 실수값 예측
- R
  - tree, rpart, party 패키지
- 관련자료
  - <http://www.dodomira.com/2016/05/29/564/>

# R 패키지

- **tree** 패키지
  - binary recursive partitioning
- **rpart** 패키지
  - CART(classification and regression tree)
  - 엔트로피, 지니 계수 기준 가지치기, 속도가 빠름. 과적합화 위험성
- **party** 패키지
  - unbiased recursive partitioning based on permutation tests
  - p-test 후 significance 기준 가지치기
  - 31개 입력 변수 레벨 제한

# 분석 방법

- 단계
  - 데이터준비 - 트리 만들기 - 가지치기 - 예측 및 모델 평가