



2015년도 제1회 통계적 방법론 공개강좌

과목 소개

Decision Tree의 개관 및 활용

Decision tree는 재귀적으로 데이터를 분할하여 만든 계층적, 축차적 나무구조 모형으로 데이터의 내재된 정보를 의미 있게 시각적으로 표현하는 방법이다. 나무구조 모형의 분할 변수로 반응 값을 예측하고 복잡한 데이터 구조를 쉽게 해석하기 위해 주로 이용된다. 50여년 전 처음으로 AID 알고리즘이 개발된 이후 CART, C4.5, GUIDE, PARTY 등 많은 decision tree 알고리즘이 새롭게 개발되어 다양한 응용 분야에 활용되고 있다. 또한 예측력 향상을 위해 bagging, boosting, random forests 같은 앙상블 기법도 개발되었다. 본 강좌에서는 기본적인 이론과 주요 알고리즘, 그리고 실제 데이터 분석에 활용 사례를 소개한다.

고려대학교 통계연구소