

Machine Learning

Lab_assignment1

2017010146 이시현

1. 주제

Costa Rican Household Poverty Level Prediction (코스타리카 가구빈곤수준 예측)

한 가정이 가지고 있는 140여 개의 속성을 기반으로 그 가정이 가난한 가정인지 부유한 가정인지 판단하는 문제이다.

2. 선택 이유

이전 다른 수업의 프로젝트에서 서울시 아파트 가격 예측 모델을 만드는 프로젝트를 진행한 적이 있다. 아파트 가격에 영향을 미치는 수많은 변수들이 있지만 정성적으로 혹은 정량적으로 아파트 가격에 영향을 미치는 변수가 어떤 것이 있을까 고민하는 과정이 매우 즐거웠다.

선정한 주제도 가난한 가정인지 부유한 가정인지 판단하는 문제로 이에 영향을 미치는 변수는 무수히 많을 것이다. 많은 변수들 중에서 모델의 예측력을 높이고, 합리적 근거를 갖는 변수를 선택하고 가공하는 과정에 집중할 것이다. 모델의 예측력을 높이는 문제에서 더 나아가 선후관계에 대해서 설명할 수 있는 분석을 하고 싶다.

3. 계획

2주차

데이터 수집

데이터 이해를 위한 탐색과 시각화

머신러닝 알고리즘을 위한 데이터 준비

3주차

모델 선택과 훈련

모델 세부 튜닝