

딥러닝 기반 핵심 산업별 빅데이터 분석

<머신러닝&딥러닝 파일럿 프로젝트>

주 제	House Prices: Advanced Regression Techniques	링 크	https://www.kaggle.com/c/house-prices-advanced-regression-techniques
팀 명	D팀	일 자	2018년 11월 29일
팀 장	조민정 <m1nch0@naver.com>	팀 원	고준형, 김효신

1. 과제 개요

사람들이 일반적으로 생각하는 집값에 영향을 미치는 요소(평수, 침실 수, 위치 등) 외에도 어떠한 요소들이 실제 집값에 영향을 끼치는지 알아보려고 한다. 해당 분석 결과는 부동산 구입을 앞둔 사람들이 가격 협상에 나설 때 매도가 조정을 유도할 수 있는 핵심 요인들을 발견하고 이용할 수 있도록 도와줄 것이다.

1) 주제

Ames, Iowa에 있는 주거용 건물들의 79개의 변수들을 분석하고 향후 집값 예측하기

2) 배울 수 있는 내용

고급 선형회귀 스킴 및 텐서플로를 이용한 예측 모델 구성, 예측 오차율을 최소화하기.

3) 주제 선정 이유

- ✓ 조원들의 수준에서 배운 내용을 전부 응용, 실습해보기에 가장 적절한 데이터
- ✓ 향후 프로젝트에서 필요한 예측 개념을 이해하고 머신러닝/딥러닝 기초를 복습할 수 있다.
- ✓ 본 분석의 데이터는 선형회귀모델을 공부하는 데이터 사이언티스트들을 위해 주어졌다,
- ✓ 동일한 집값 예측 데이터인 Boston Housing(<모두의 딥러닝> 예제)에서는 데이터가 이미 충분히 정제되고 가공되어 쉽게 분석이 가능했다면, 본 데이터는 전처리 과정을 실습하고 예측모델에 맞게 가공하는 훈련까지 할 수 있어 선정했다.
- ✓ Boston Housing (<https://www.kaggle.com/schirmerchad/bostonhousingmlnd>)

2. 데이터 설명

1) train.csv

1461 x 81

주요 데이터 필드

SalesPrice(예측해야 하는 것, 집값), LotShape(평수), Kitchen(부엌 수), bedroom(침실 수), Yearbuilt(연식), Foundation(지반), Electronial(전기타입), Neiborhood(시내와의 물리적 거리), HouseStyle(주거타입), SaleCondition(판매 상태) 등 총 81개 칼럼

2) test.csv

3) data_description.txt

각 칼럼에 대한 세부 설명

4) sample_submission.csv

침실 수와 평방미터, 연간 판매량을 통해 도출한 선형 회귀 모델의 결과 샘플

3. 과제 수행 내역

4. 결과 보고

감사합니다