

핸즈온 머신러닝 2장

머신러닝 처음부터 끝까지

- 캘리포니아 인구조사 데이터를 이용해 캘리포니아 주택 가격 모델 만들기

1. 비즈니스의 목적?

2. 현재 솔루션은 어떻게 진행되는지?

- 지도학습

- 다중 회귀

- 단변량 회귀

성능 지표

MAE vs RMSE

평균 절댓값 오차 (Mean absolute error, MAE)

$$E(|\hat{\theta} - \theta|)$$

평균 제곱근 오차 (Root Mean square error, RMSE)

$$\sqrt{E((\hat{\theta} - \theta)^2)}$$

론칭 이후

- 모델의 실전 성능 모니터
- 데이터셋을 업데이트하고 모델을 정기적으로 다시 훈련
- 모델의 입력 데이터 품질 평가
- 모델 백업

참고 자료

- 오헬리앙 제롱, *핸즈온 머신러닝 2판* (한빛미디어, 2020), 67-125
- sklearn.impute.KNNImputer, *scikit-learn 0.24.2 documentation*, 2021년 5월 23일 접속,
<https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html>
- KNNImputer: A robust way to impute missing values (using Scikit-Learn), *Analytics Vidhya*, 2020년 6월 14일 수정, 2021년 5월 23일 접속,
<https://www.analyticsvidhya.com/blog/2020/07/knnimputer-a-robust-way-to-impute-missing-values-using-scikit-learn/>