

## 5.9장

### 회귀 실습-자전거 대여 수요 예측

- 데이터 클렌징 및 가공/ 데이터 시각화
  - **string**으로 되어있는 데이터를 **datetime** 데이터형으로 전환
  - 캐글에서 요구한 성능 평가 방법은 **RMSLE**
- 필요없는 피처 삭제하기
- 평가지표
- 타깃 데이터의 로그 전환, 피처 인코딩과 모델 학습/예측/평가
  - 해당 함수를 사용하는 이유는 **log**함수를 이용할때 데이터 값의 크기에 따라 **overflow/underflow** 오류가 발생하기 쉽기 때문
  - **one-hot encoding**
  - 회귀 트리 모델을 이용하여 예측

○

#### 5.10장 캐글 주택 가격: 고급 회귀 기법

- 주택 가격 **dataset** 다운
  - 자전거 대여 예측 예제와 동일한 **RMSLE**를 기반으로 함
  - 예측 결과 오류가 전체 오류에 미치는 비중이 높으므로 이를 상쇄하기 위해 오류 값을 로그 변환한 **RMSLE**를 이용
- 데이터 사전 처리
  - 정규 분포 형태로 분포하는 것을 볼 수 있음
- 선형 회귀 모델 학습/ 예측/ 평가
  - 예측 평가 지표로 **RMSLE**를 사용할 것. 하지만 이미 **target** 값이 로그 변환됨.  
->예측 결과 오류에 **RMSE**만 적용하면 **RMSLE**가 자동으로 측정됨.
- 회귀 트리 모델 학습/예측/평가
  - **XGBoost**, **LightGBM** 순서로 학습, 예측을 수행
  - 두 모델들의 **feature** 중요도를 시각화
- 회귀 모델의 예측결과 혼합을 통한 최종 예측
  - 두 모델의 예측값이 주어지면 두 모델의 예측값이 차지할 비율을 나눈 다음 합하여 최종 회귀 값으로 예측하는 것
  - 성능 향상
- 스택킹 앙상블 모델을 통한 회귀 예측
  - 스택킹 모델의 핵심은 여러 개별 모델의 예측 데이터를 각각 스택킹 형태로 결합해 최종 메타 모델의 학습 **dataset**과 테스트 **dataset**을 만드는 것
  - 적용할 개별 모델은 릿지, 라쏘, **XGBoost**, **LightGBM**으로 4개이고, 메타 모델은 라쏘 모델을 사용하여 학습하고 예측을 수행