**선형회귀분석(Linear Regression & Ridge & Lasso)**

**회귀 문제**

**Theory**

회귀는 여러 개의 독립변수와 한 개의 종속변수 간의 상관관계를 모델링하는 기법을 통칭합니다. 머신러닝 회귀 예측의 핵심은 주어진 피처와 결정 값 데이터 기반의 학습을 통해 **최적의 회귀 계수**를 찾아내는 것입니다.

**다중선형회귀(Linear Regression, Ridge, Lasso)**

**Theory**

LinearRegression 클래스는 예측 값과 실제 값의 RSS(Residual Sum of Square)를 최소화하는 해를 찾는 추정 방식으로 구현한 클래스입니다. 선형회귀는 입력 피처의 독립성에 영향을 많이 받습니다. **피처 간의 상관관계**가 매우 높은 경우 분산이 매우 커져서 오류에 매우 민감해집니다. 이러한 현상을 다중공선성(Multi-collinearity) 문제라고 하며, 일반적으로 **중요한 피처만 남기고 제거하거나 규제**를 적용합니다. 다중 공선성 문제가 너무 심각할 경우 **PCA**까지 고려해볼만 합니다.

다중선형화귀는 별도로 튜닝할 값이 없습니다. 규제를 적용한 경우 규제 인자에 대한 하이퍼파라미터 튜닝이 필요합니다.

**회귀 평가 지표**

MAE : metrics.mean\_absolute\_error , ‘neg\_mean\_absolute\_error’

MSE : metrics.mean\_squared\_error, ‘neg\_mean\_squared\_error’

R^2 : metrics.r2\_score, ‘r2’

**다항선형회귀(Polynomial Linear Regression)**

**Theory**

PolynomialFeatures(degree=n) -> poly.fit\_transform(X)

**▷** 결측치, 이상치 -> 표준화 or 정규화 or 로그변환 -> 선형회귀