

multiple regression

2022년 6월 3일 금요일 오후 9:09

1. 다중공선성 (Multicollinearity)

- 독립변수들이 강한 선형관계에 있을 때
- 이런 독립변수들은 종속변수를 설명하는데 비슷한 영역을 설명한다.
- 예측의 정확도 하락

2. 다중공선성 확인방법

- VIF(Variance inflation factor)
 - o 분산 팽창요인
 - o 변수들간 correlation 진단 ($VIF_i = \frac{1}{1-R_i^2}$)
 - o 결정계수가 커질수록 VIF 커짐 -> 다중공선성 있는 factor

3. 모든 변수를 넣는 것 보다 중요 변수만 선택하는 것이 좋음

- 변수 갯수가 작아져 관리 하기 쉽고 예측속도 빨라짐

4. 성능지표

- MSE : Mean Squared Error
- RMSE : Root Mean Squared Error
- MAE : Mean Absolute Error
- MAPE : Mean Absolute percentage Error