

## KNN

n\_neighbors : 이웃 수, k의 값

weights: 이웃의 가중치 결정방법, 디폴트 'uniform' 동일한 가중치, 'distance' 거리의 반비례 가중치

metric : 거리측정방식, 'euclidean', 'manhattan' 방식

algorithm : 거리 기반으로 어떻게 탐색할 것인가?

'auto' : 가장 적절한 것 골라서 데이터 차원, 패턴에 따라서 자동으로 탐색하는 것

'ball\_tree' : 필요한 부분 빠르게 탐색하기 위한 방법

'kd\_tree' : 데이터 구조, 분할방식으로 차원 축소하면서 진행하는 방법

'brute' : 완전탐색방식 가장직관적인 방법, 데이터 크면 계산이 증가한다.

가중치에 따라 결과값 달라짐-> 좀 더 매끈

적정 k수가 지나고 난 후에 점점 MSE 증가하는 현상

왜 그럴까?

과적합하여 -> 불필요한 패턴까지 배워서

K가 증가하면, 모델이 학습데이터의 패턴을 충분히 학습을 하지 못하는 경우, 너무 단순한 모델이 된다. 과소적합의 개념 일반화를 못하는 경우

K가 증가하면, 이상치 영향으로 인해서 MSE 값이 올라갈 수 있다.

이상치 탐지 : 혼자 튀는 숫자

-> 이미지 변환 이용... 상자그림, 산점도 등등