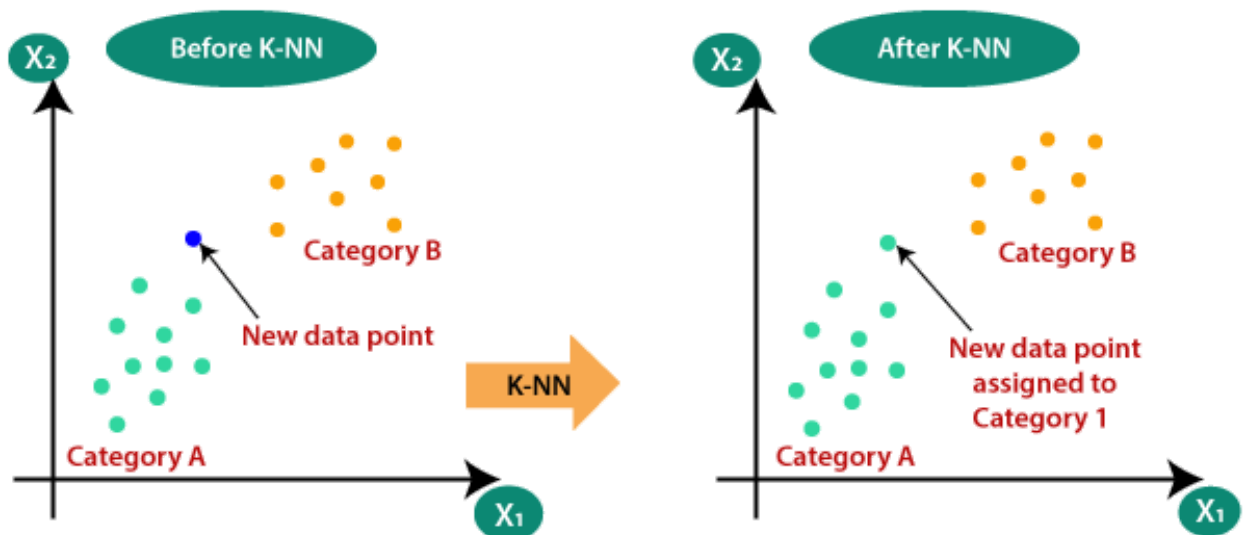


6. KNN

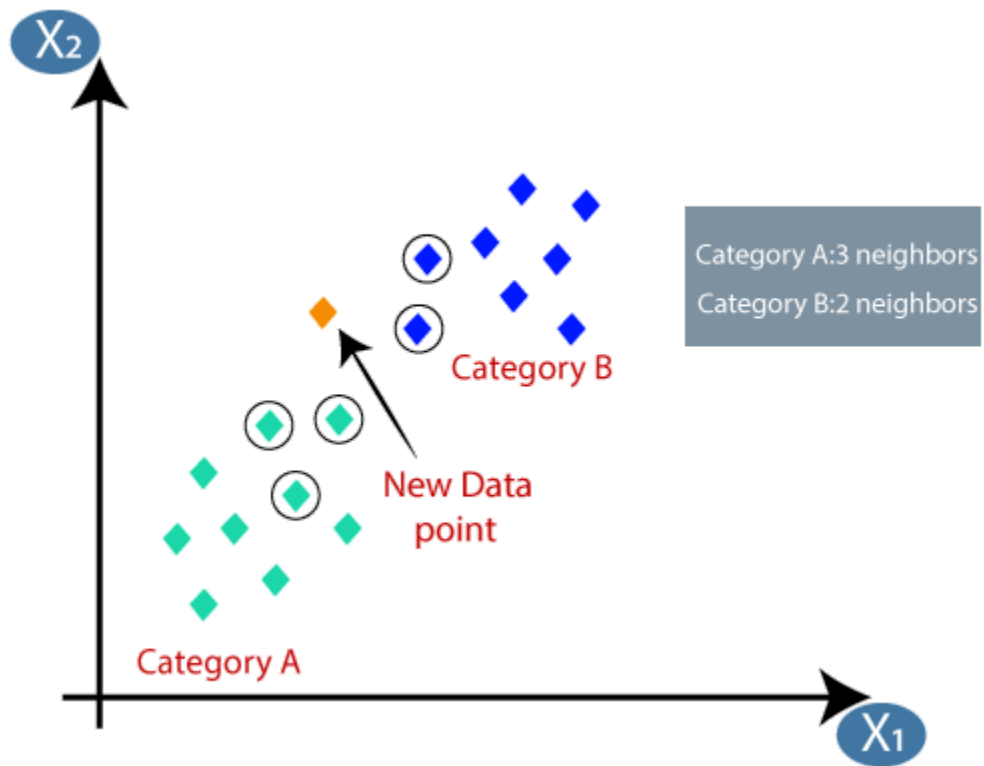
K-nearest neighbor



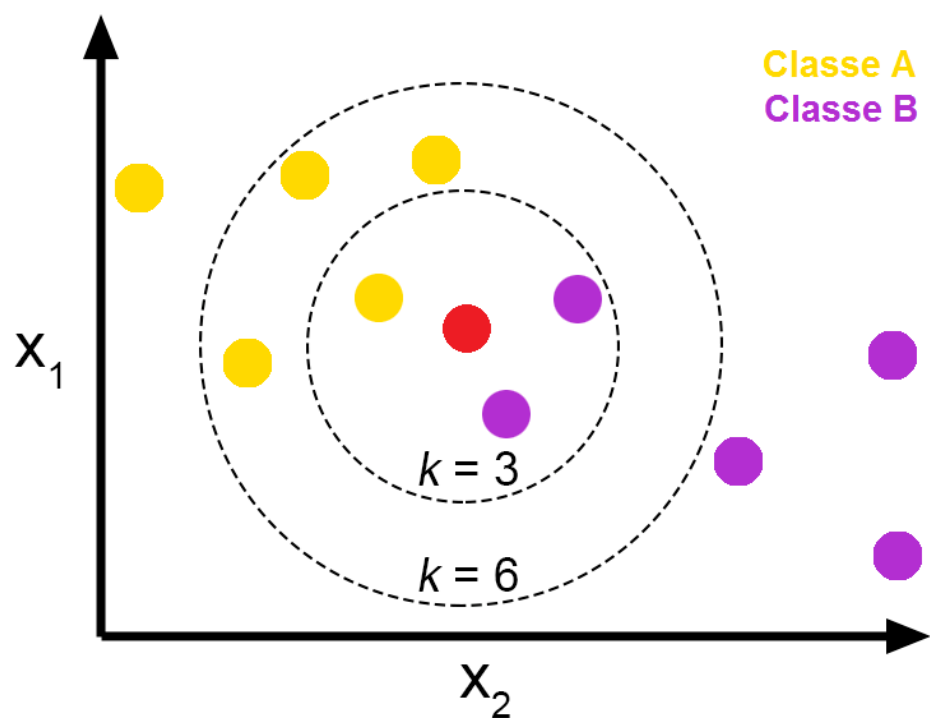
동작

1. New Data와 모든 Train Set사이의 거리 측정
2. 유클리드 거리가 가까운 순으로 K개의 점을 이웃으로 갖는다.
3. K의 개의 점들중 가장 많이 속한 Class 를 찾는다.
4. NewData를 새로운 Class 에 할당한다.

- K=5 NewData는 녹색으로 할당된다.

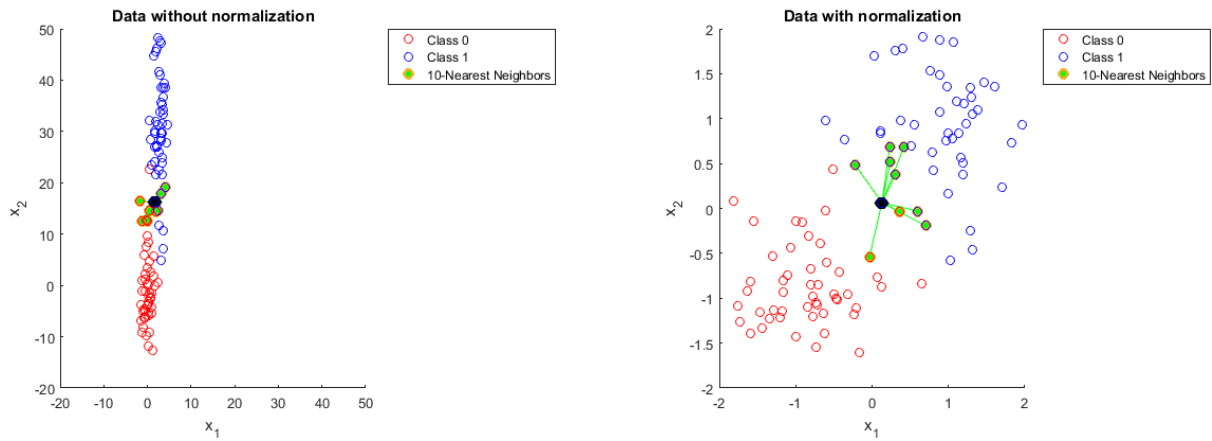


- K 값에 따라 다른 Class 로 할당된다.



변수 값 범위 재조정

- 각 feature 의 단위가 다르기 때문에 거리 측정시에 범위를 재조정 해줘야 한다.
- 물렁도, 크기의 단위는 서로 다르기 때문에 같은 거리 측정 방식을 사용하면 안된다.

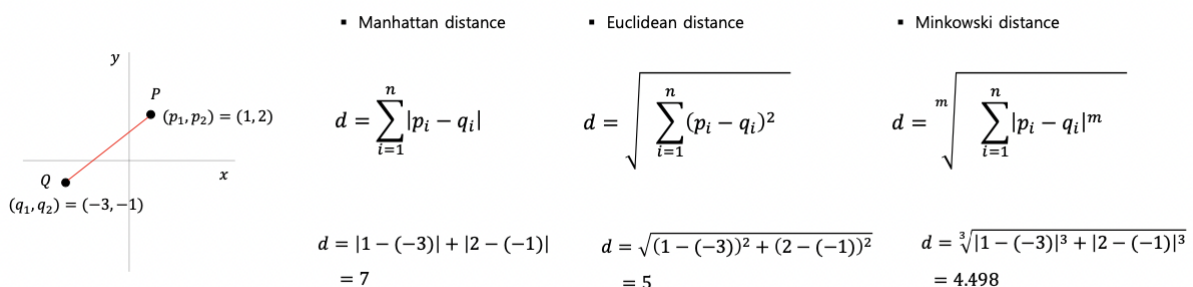
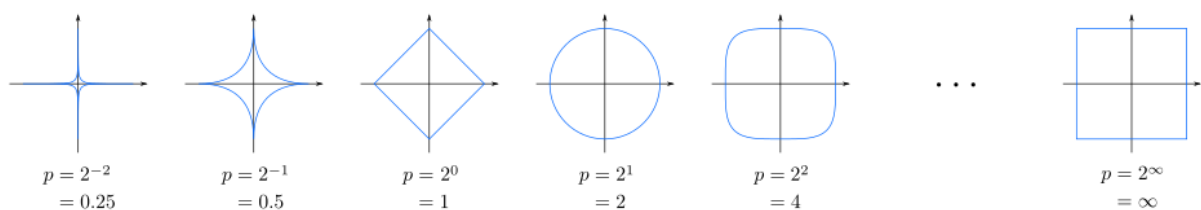


- 최소-최대 정규화(min-max normalization)
- z-점수 표준화(z-score standardization) **(추천)**

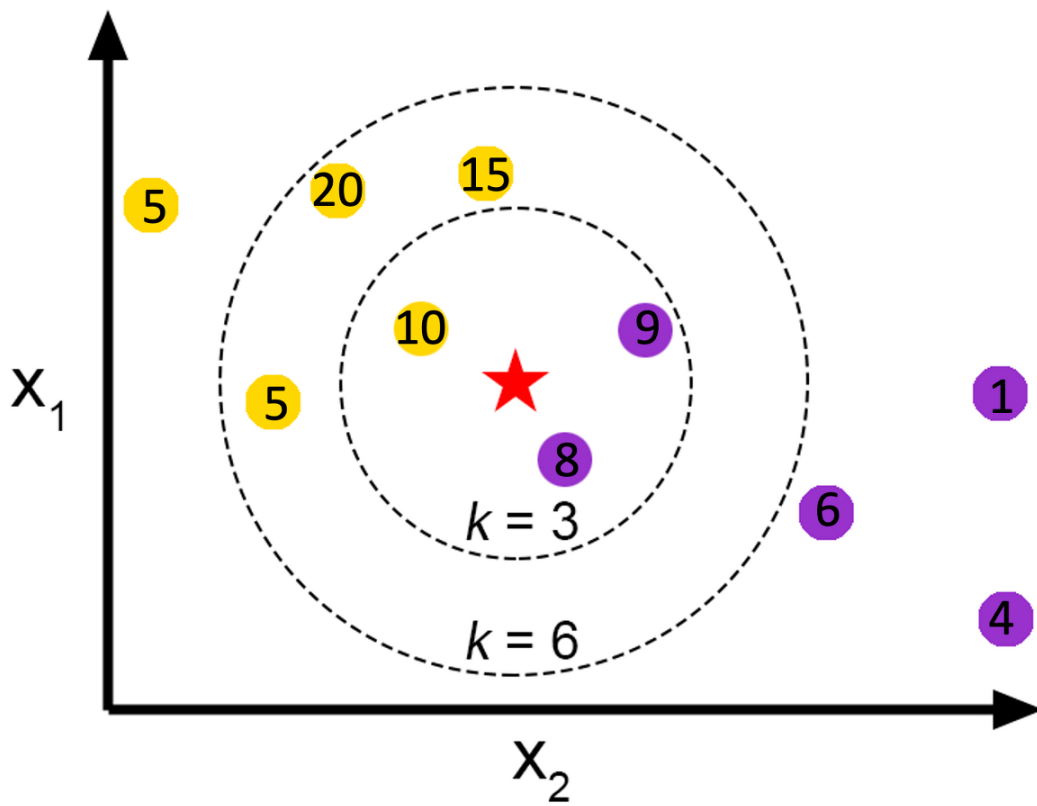
Minkowski 거리

$$D(X, Y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}.$$

- Minkowski 거리는 Manhattan 거리와 Euclidean 거리의 일반화
- $p = 1 = 2^0$ 일 때는 Manhattan 거리
- $p = 2 = 2^1$ 일 때는 Euclidean 거리



회귀



- 주변데이터의 평균을 이용해 예측값을 결정
- 데이터 범위 밖의 새로운 데이터는 예측 불가능