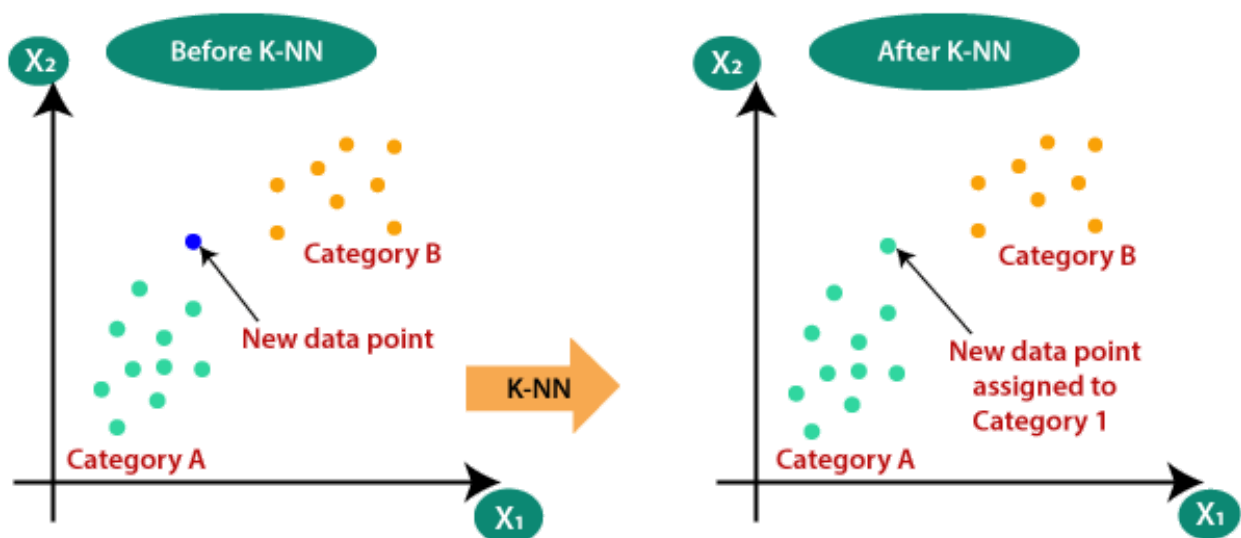


# 5. KNN

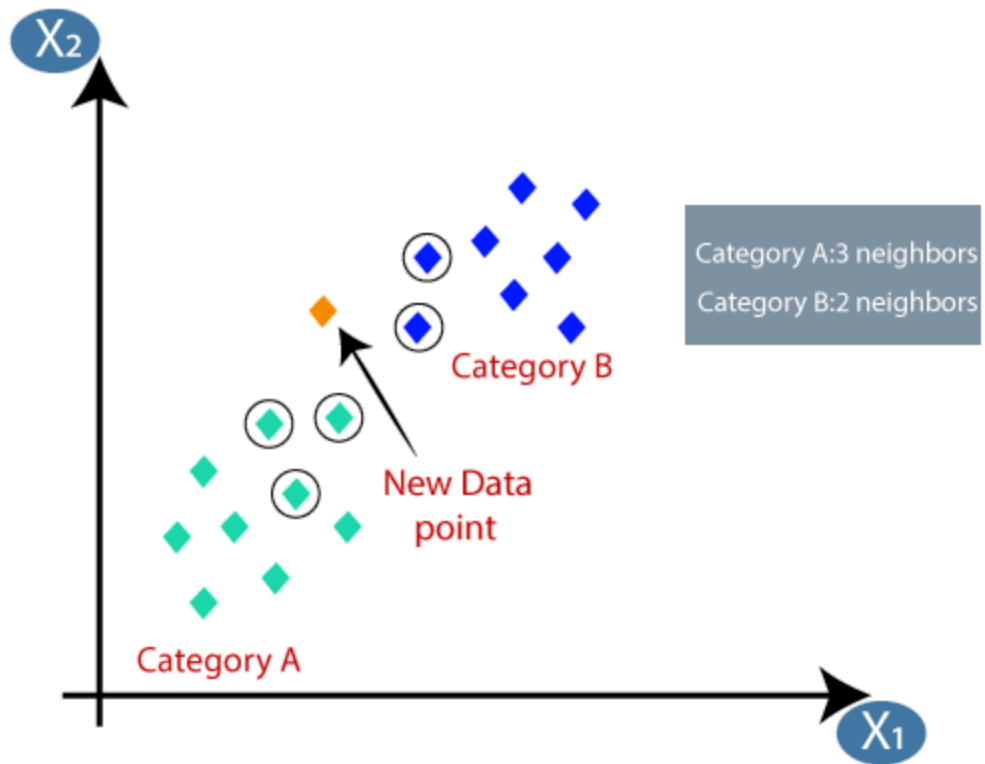
## K-nearest neighbor



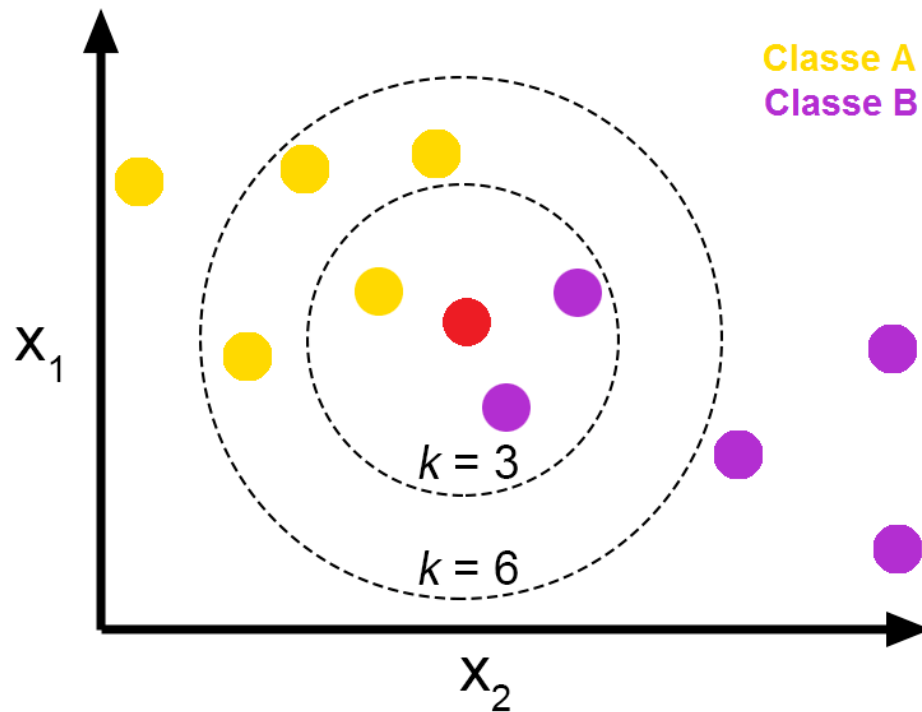
## 동작

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 주황 으로 할당된다.



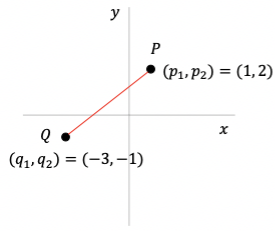
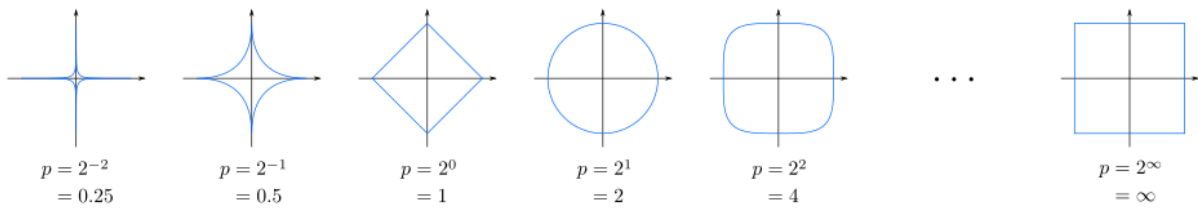
- K 값에 따라 다른 Class 로 할당된다.



## 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



▪ Manhattan distance

$$d = \sum_{i=1}^n |p_i - q_i|$$

$$d = |1 - (-3)| + |2 - (-1)| = 7$$

▪ Euclidean distance

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$$d = \sqrt{(1 - (-3))^2 + (2 - (-1))^2} = 5$$

▪ Minkowski distance

$$d = \sqrt[m]{\sum_{i=1}^n |p_i - q_i|^m}$$

$$d = \sqrt[3]{|1 - (-3)|^3 + |2 - (-1)|^3} = 4.498$$