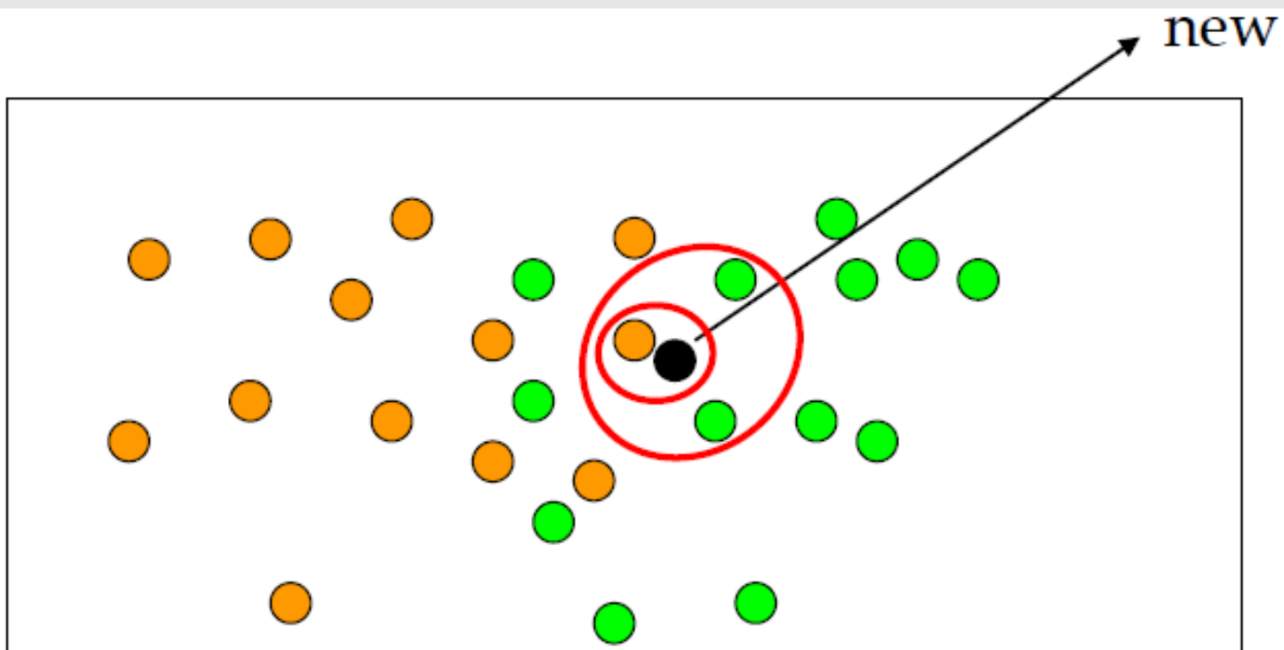


K-Nearest Neighbor (KNN)

KNN 알고리즘, 머신러닝의 분류, 회귀에 쓰이는 대표적인
간단한 알고리즘입니다.

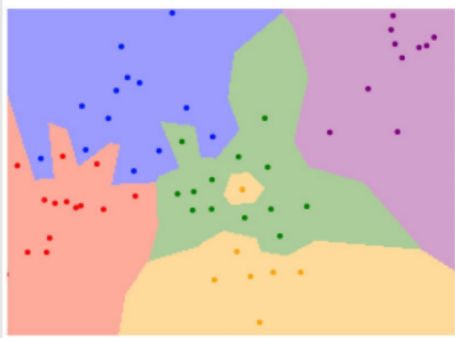
KNN \rightarrow unsupervised learning 정답레이블이 없는 경우
class를 예측할 수 있다.

KNN은 새로운 데이터가 주어졌을 때 기존 데이터 가운데
가장 가까운 k개 이웃의 정보로 새로운 데이터를 예측하는
방법입니다.

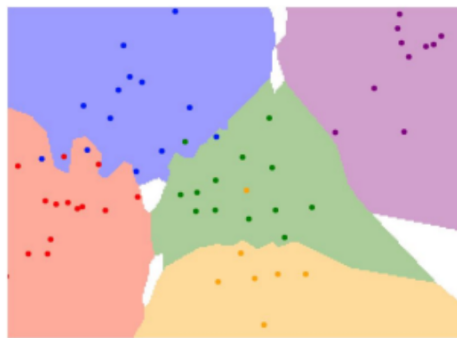


학습이라 할만한 절차가 X.

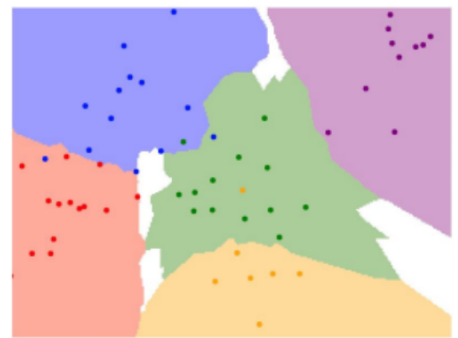
Instance-based learning이라 불리기도 함.



K = 1



K = 3



K = 5

k 값이 커짐에 따라 분류 경계면이 단순해집니다.

KNN의 Hyperparameter인 k 는 직관적 지역적 특성을 지나치게 반영 (overfitting)

클러스터 모델이 과하게 정규화되는 경향이 있습니다.

KNN은 distance 측정방법에 따라 그 결과가 달라지는 알고리즘입니다.

1. Euclidean Distance (L2 Distance)

특히 마르코프 체인 방법보다
훨씬 더 robust한 방법입니다.

크리틱이웃의 $k(k)$ 와 어떤 거리 척도가 분석에 적합한지
불분명해 데이터 각각의 특성과 맞게 연구자가 임의로
선택해야 하는 단점이 있습니다.

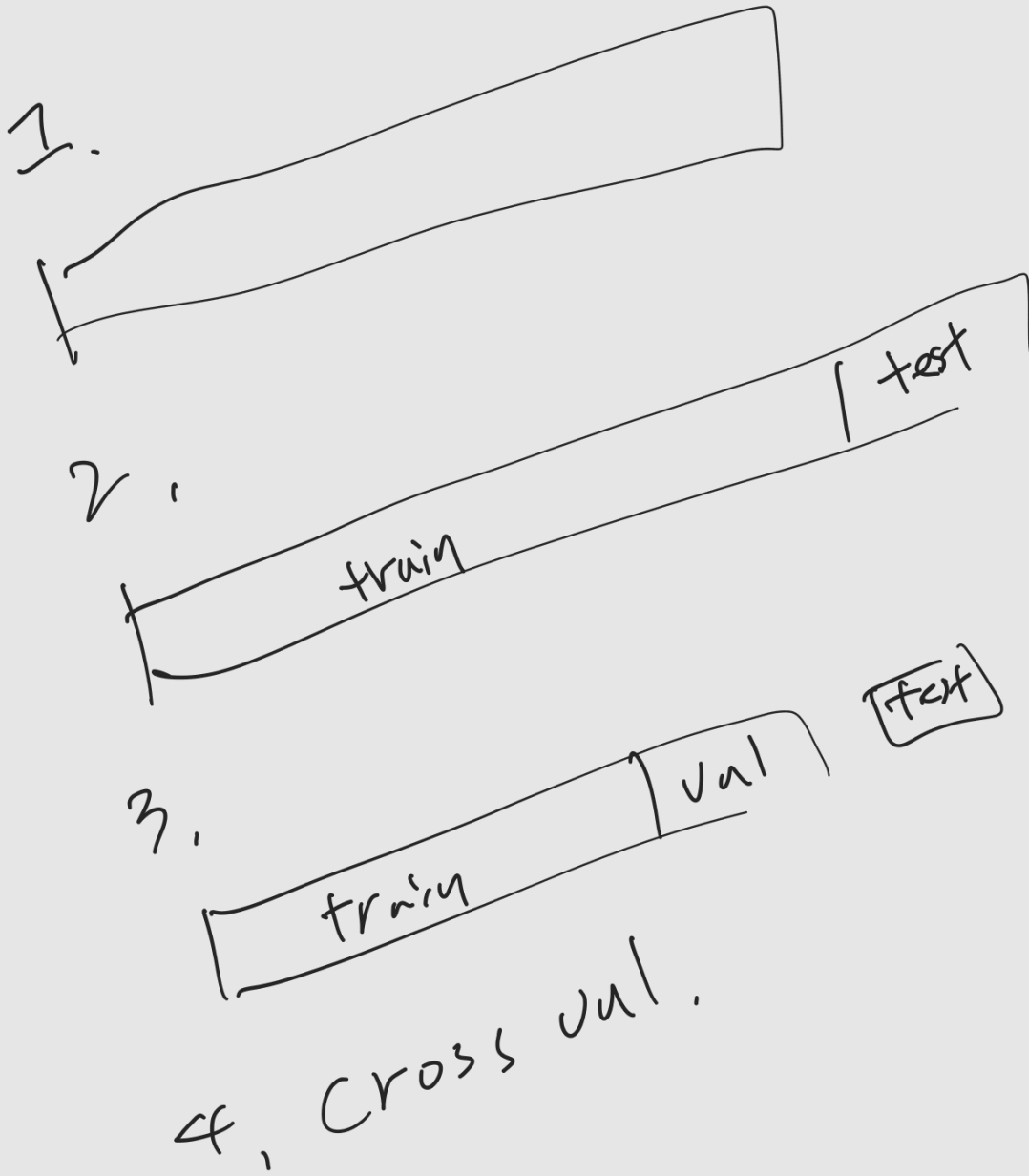
새로운 관측치와 각각 학습데이터 사이의 distance를
계산 선택해야 함으로 계산 시간이 오래 걸립니다.

이미지 X

Hyperparameters

k , distance.

$k \Rightarrow$ coordinate dependency



ImageNet of 1000 classes.

Distance metrics on pixels are

Information

not in

Curse of dimensionality

Parametric App

Image



Array

$32 \times 32 \times 3$



$f(x, w)$



10 numbers

w

Word2Vec.