

knn 알고리즘



knn 알고리즘 이란 ?

머신러닝의 종류 중에 하나입니다

머신러닝의 종류는 3가지로 나뉘고

1. 지도 학습
2. 비지도 학습
3. 강화 학습

지도 학습은 분류와 회귀로 나뉩니다

1. 지도 학습

- 분류
- 회귀

1. 비지도 학습

2. 강화 학습

knn 은 이중에 분류에 속합니다

1. 지도 학습

- 분류 : knn
- 회귀

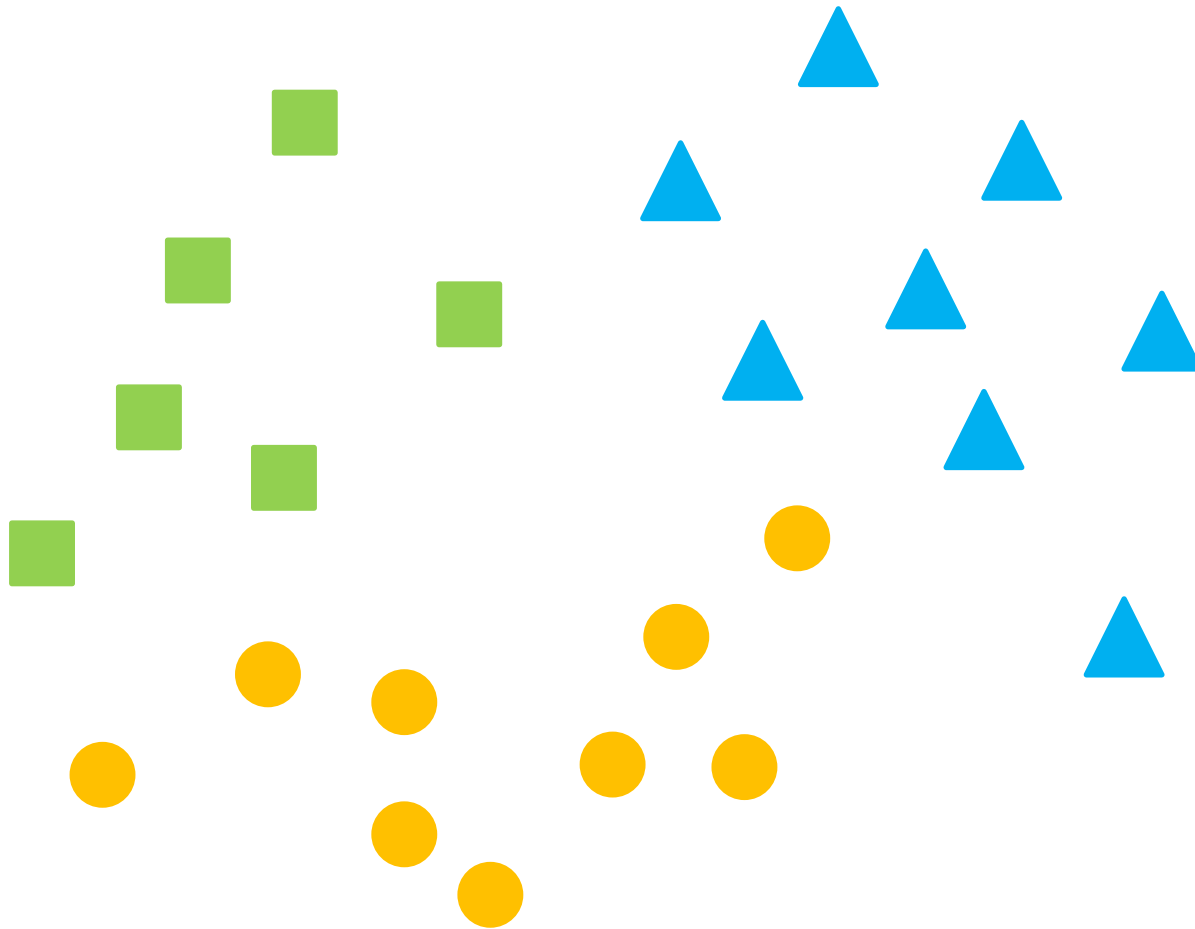
1. 비지도 학습

2. 강화 학습

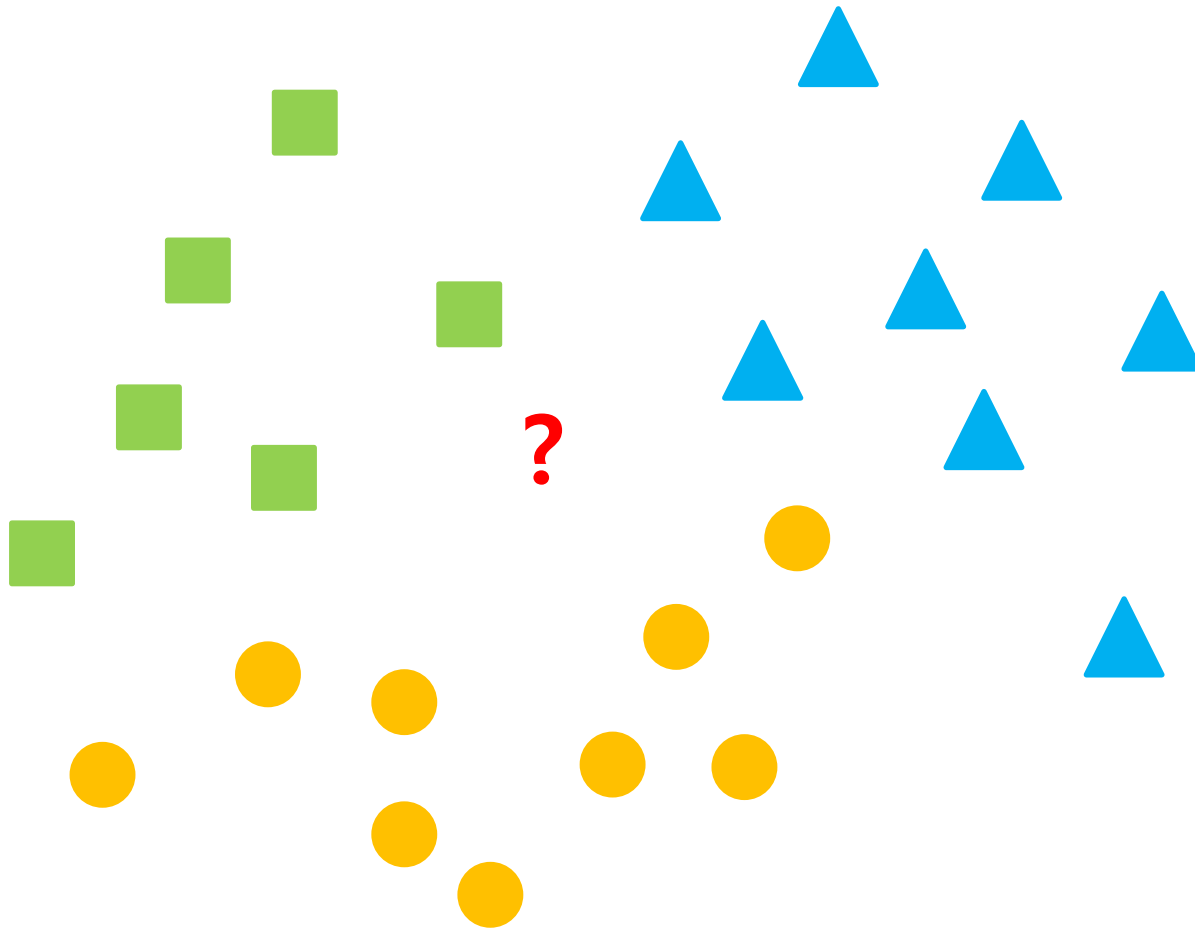
분류문제란
새로운 데이터가 들어왔을때
이 데이터가 기존에 있던 데이터의 그룹중에
어느 그룹에 속한건지를 분류하는것을 말합니다

**knn 알고리즘의 분류 문제 해결방법은
아주 간단합니다**

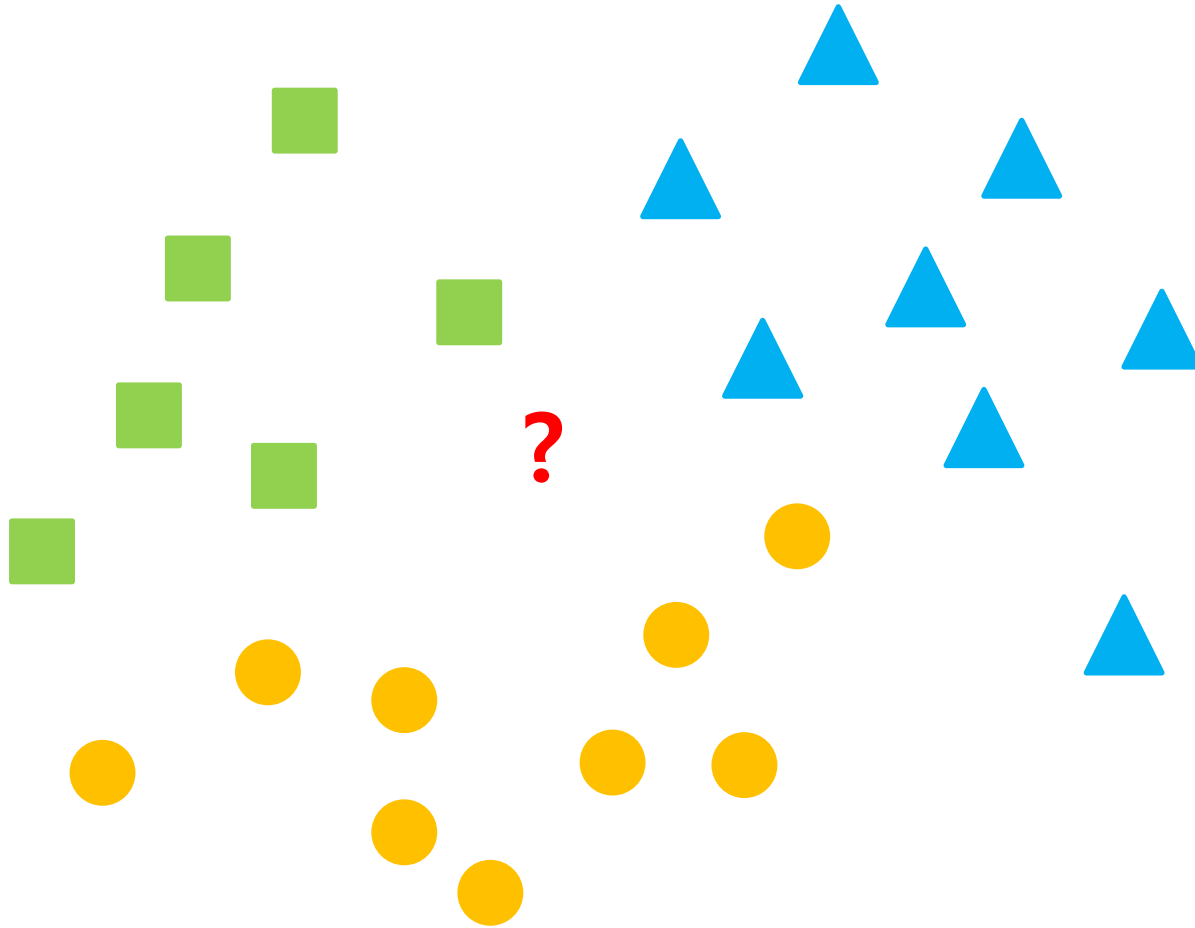
기존에 이렇게 분포되어 있는 데이터에



새로운 ? 가 들어왔을때



knn 은 그냥 새로 들어온 ? 가 ■ 에 더 가까우니
■ 로 분류하겠다는 알고리즘입니다



knn 은 k nearest neighbor 의 약자로

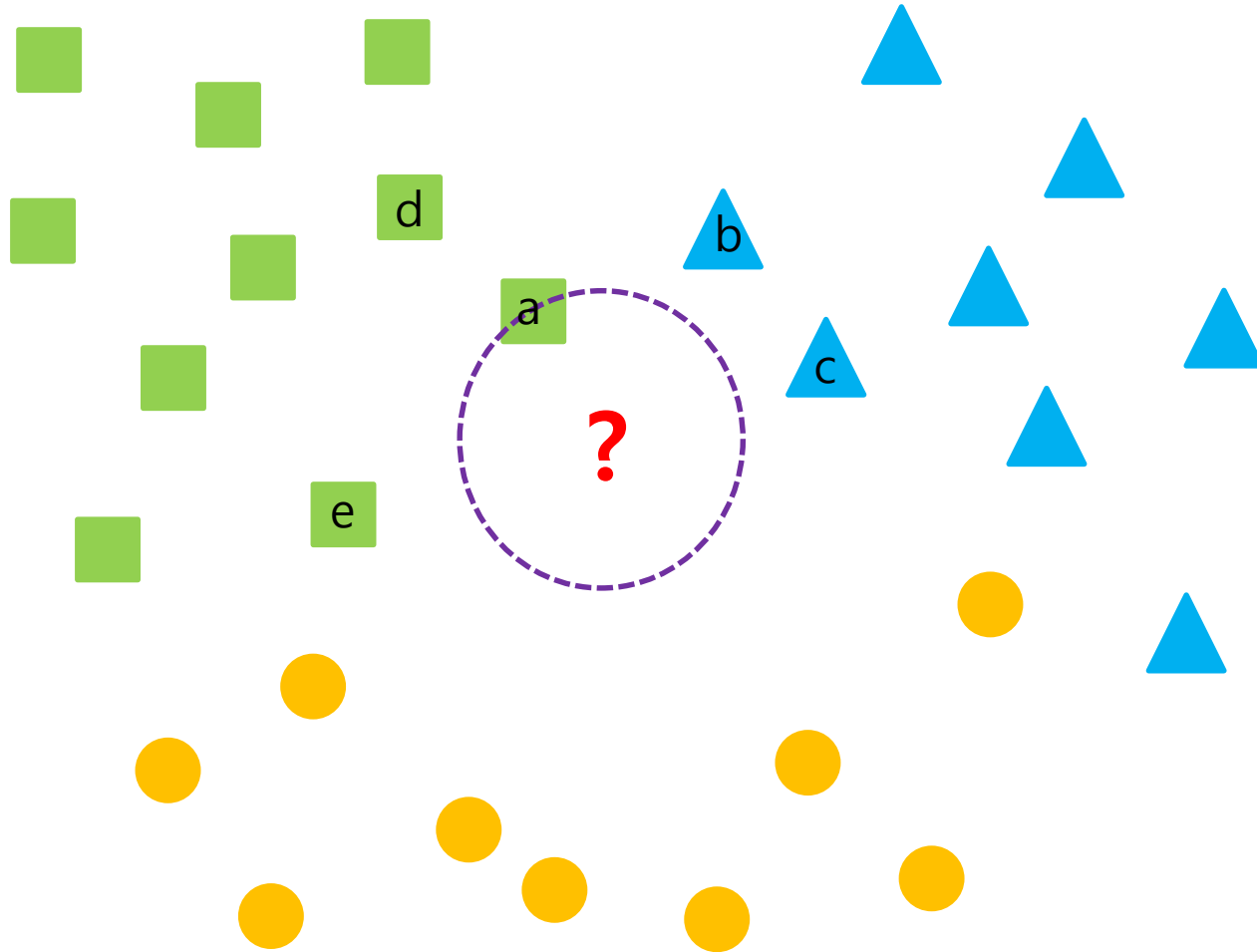
한국어로 k 근접 이웃입니다

여기서 k 는 몇번째로 가까운 데이터를
살펴볼것인가를 정한 숫자로

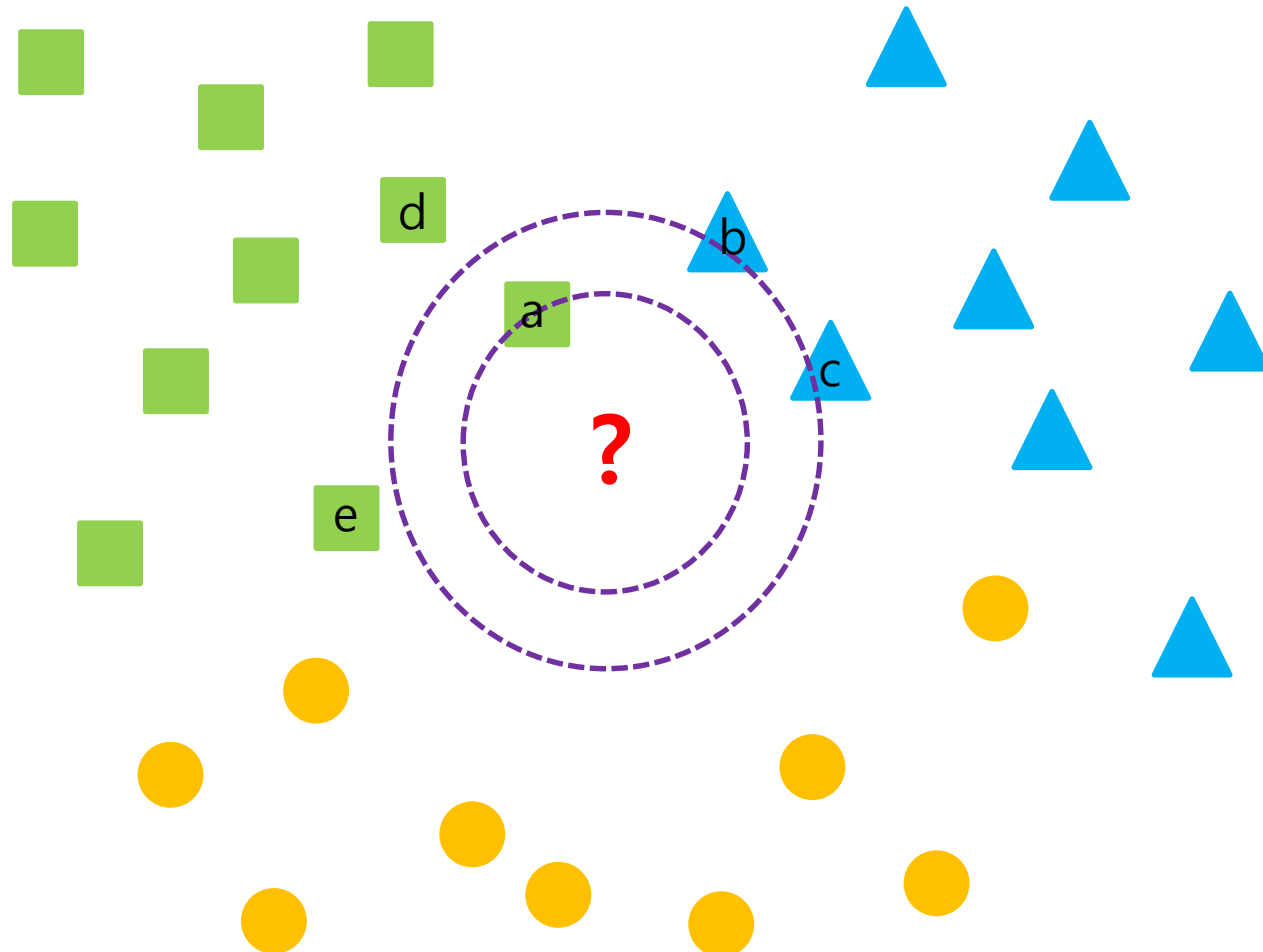
k가 1이면
1번째로 가까운 이웃 데이터

k가 3이면
3번째로 가까운 이웃 데이터라는
뜻입니다

k 가 1이면 ? 의 1번째로 가까운 이웃을 a 로 보고
신규 데이터를 ■ 로 분류하는것입니다

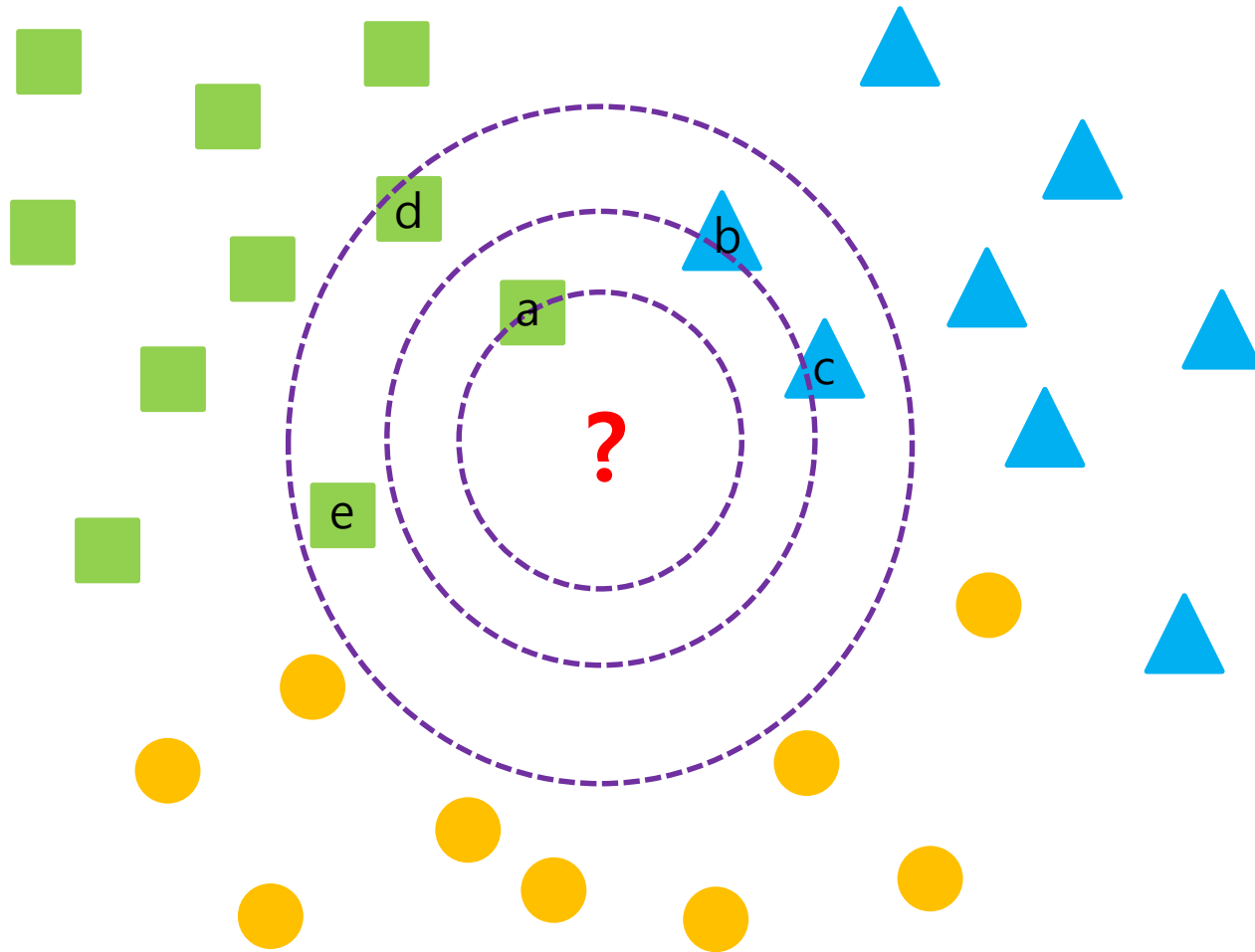


k 가 3이면 ? 의 3번째로 가까운 이웃을 a,b,c 로 보고
신규 데이터를 분류 하는데 이렇게 그룹이 갈리면



다수결에 의해서 ▲가 2 이고 ■가 1이므로 ▲로
구분합니다

k 가 5이면 ? 의 5번째로 가까운 이웃을 a,b,c,d,e 로 보고 신규 데이터를 분류 하는데

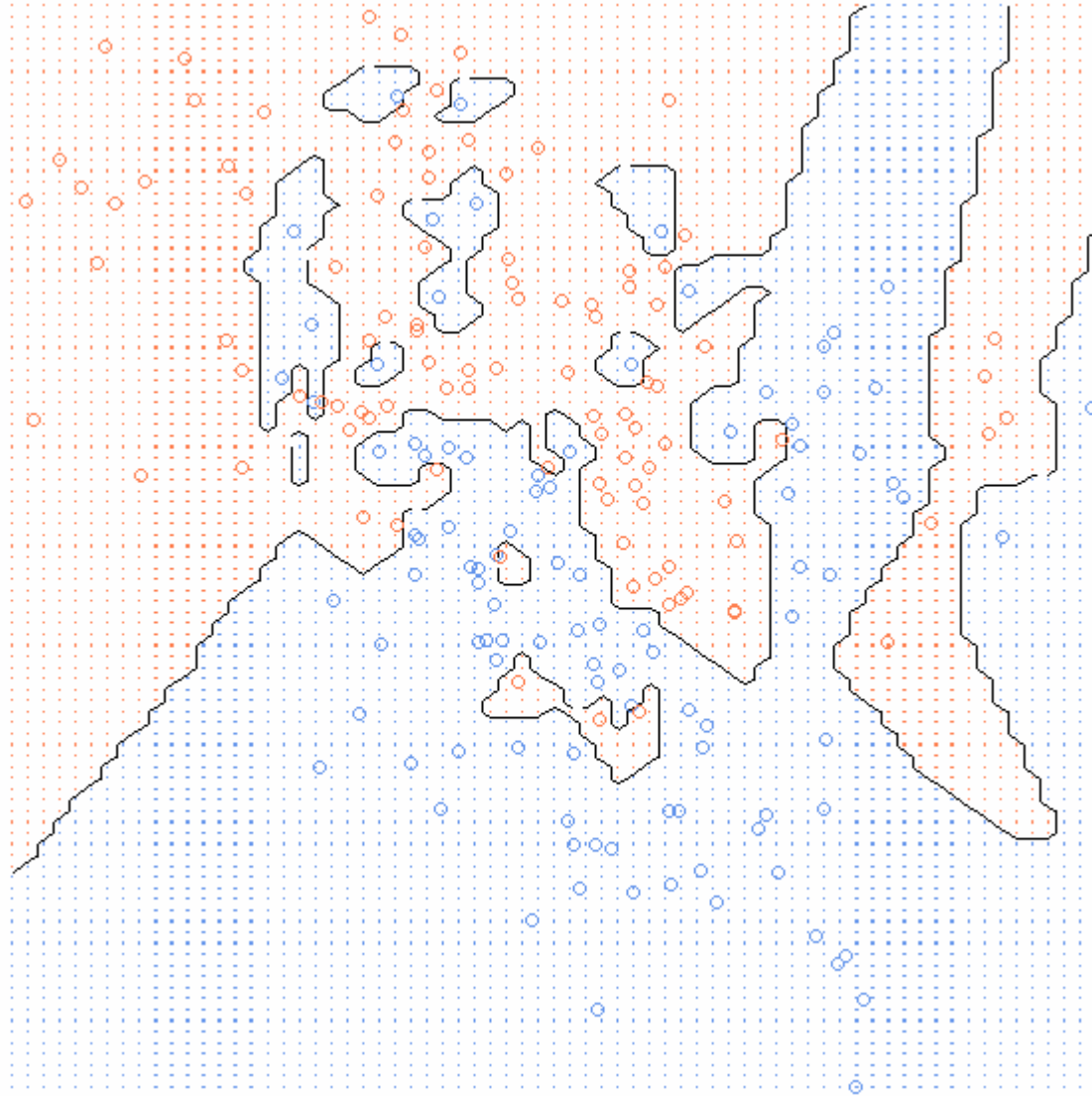


여기서는 ■가 3, ▲가 2이므로 ■로 구분 합니다

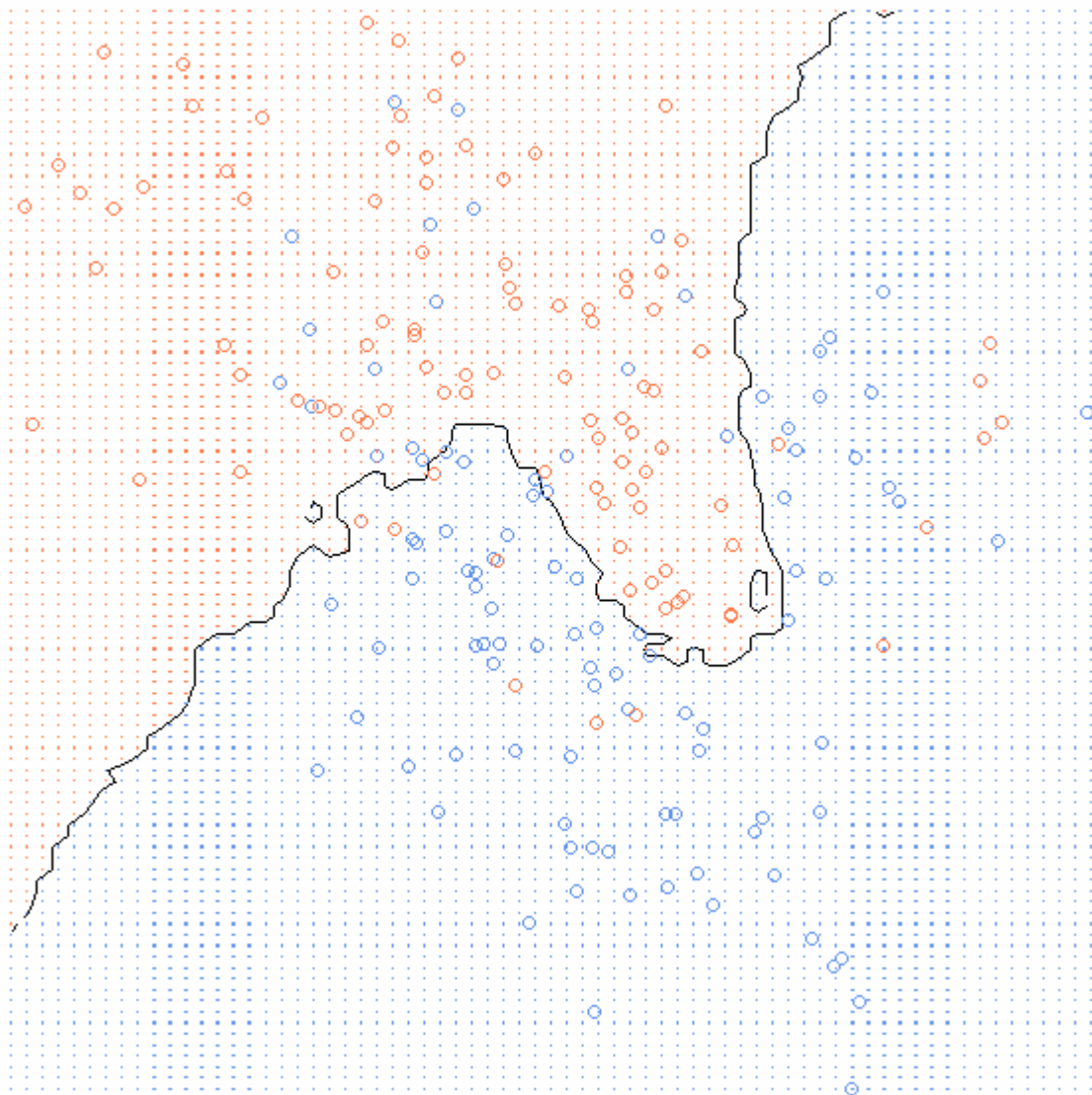
그래서
k의 개수는 의사결정을 해야하기 때문에
항상 홀수를 선택해야합니다

k 의 적절한 개수를 정해주는것도 중요한데

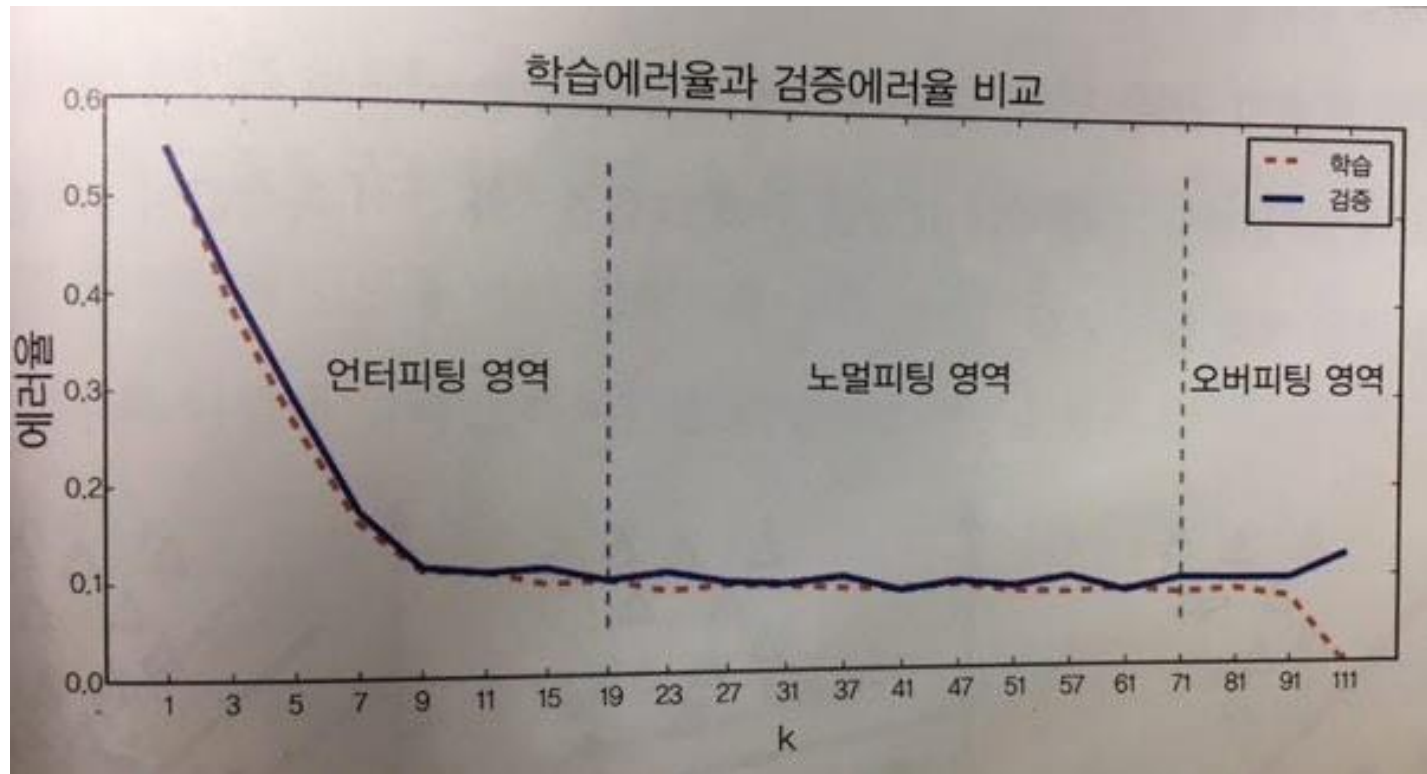
이 경우는 k 가 1일때 인데 너무 오버피팅되고 있습니다



이 경우는 k가 99일때 인데 너무 언더피팅되어 있구요



그래서 적절한 k값을 찾기위한 실험을 해야합니다



참고자료

꼬깰콘의 분석일지 블로그

인공지능, 머신러닝, 딥러닝 입문

위키북스 - 김의중 선생님

사랑하는 자여 네 영혼이 잘됨같이 네가 범사에 잘되고 강건하기를 내가 간구하노라

- 성경 요한삼서 1장 2절