

chapter10. KNN알고리즘

- Hello Coding 그림으로 개념을 이해하는 알고리즘 -



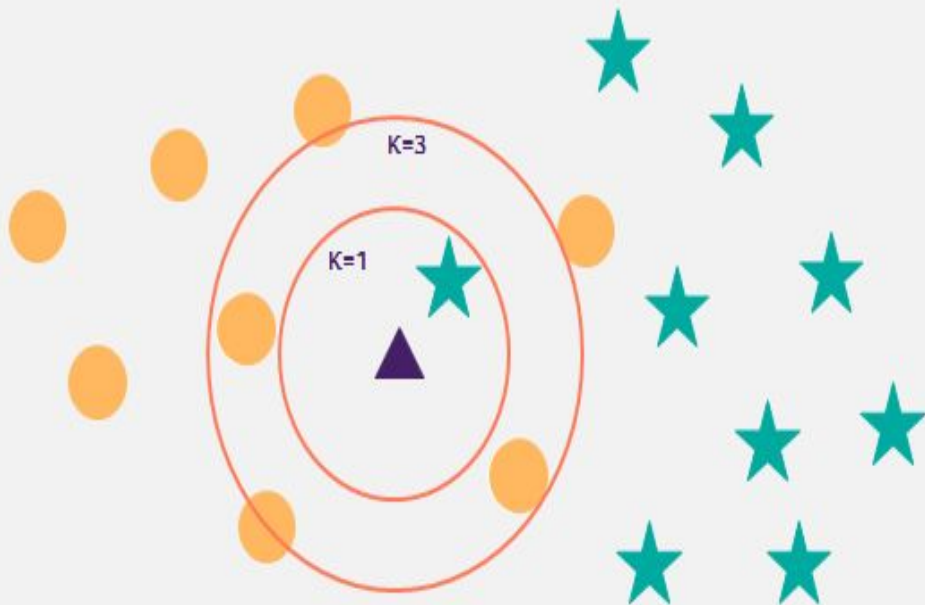
TABLE OF CONTENTS



1. KNN알고리즘
2. 분류
3. 회귀 분석

1. KNN알고리즘

KNN - K-Nearest Neighbor 알고리즘



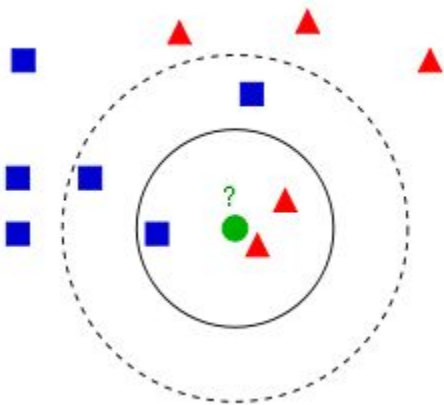
K-Nearest Neighbor, K-최근접 이웃 알고리즘

지도학습 알고리즘에 해당하며 **분류** 또는 **회귀** 모두 사용 가능

특정 공간 내 입력된 데이터와 가장 가까운 K개의 요소를 찾아 더 많이 일치하는 곳으로 분류하는 알고리즘

분류 : 그룹으로 나누는 작업
회귀 : 숫자로 된 반응을 예측

2. 분류(오렌지와 자몽 구별하기)



오렌지와 자몽은 생김새가 비슷합니다.

파란색 네모는 오렌지이고 빨간색 삼각형은 자몽입니다.

초록색 원은 오렌지일까요? 자몽일까요?

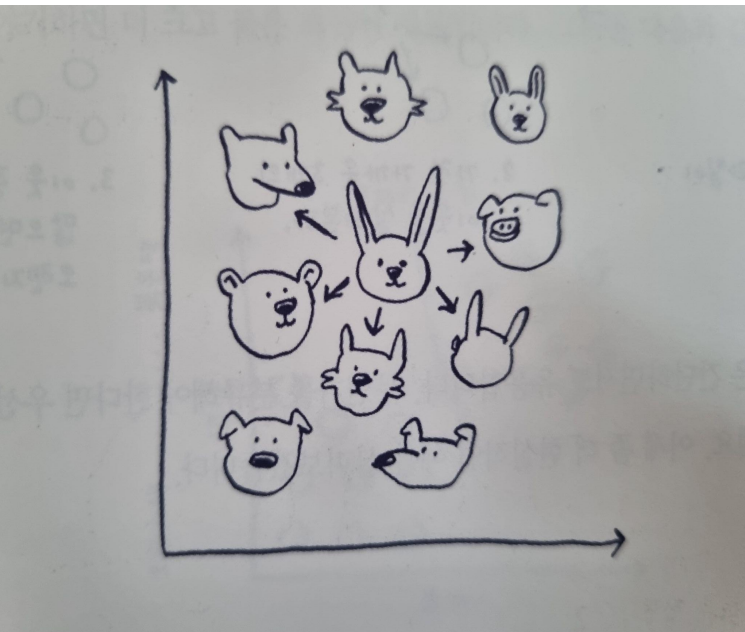
최근접 이웃 데이터를 살펴보겠습니다. ($k=3$)

자몽은 2개이고 오렌지는 1개입니다.

그러면 초록색 원은 자몽입니다.

2. 분류(추천 시스템 만들기)

넷플릭스에서 근무하고 있고 고객들을 위해 영화 추천 시스템을 만들고 싶다고 가정해봅시다.



유사도를 이용하여 고객을 그래프 상에 표현했습니다.

프리얀카라는 고객을 위해 어떤 영화를 추천해야 한다고 가정하죠.

우선 그녀와 취향이 가장 비슷한 다섯 명의 고객을 찾습니다.

저스틴, 제이씨, 조이, 랜스, 크리스라는 고객들이 프리얀카와 취향이 비슷하다면 이 고객들이 좋아하는 영화는 프리얀카도 좋아할 것입니다.

만약 저스틴이 A라는 영화를 좋아한다면 그 영화를 프리얀카에게도 추천합니다.

3. 회귀 분석

빵집에서 근무하고 있다고 가정해봅시다. **오늘 빵을 몇개 만들어야 할지 예측**하고 싶습니다. 분석에 사용될 특징은 다음과 같습니다.

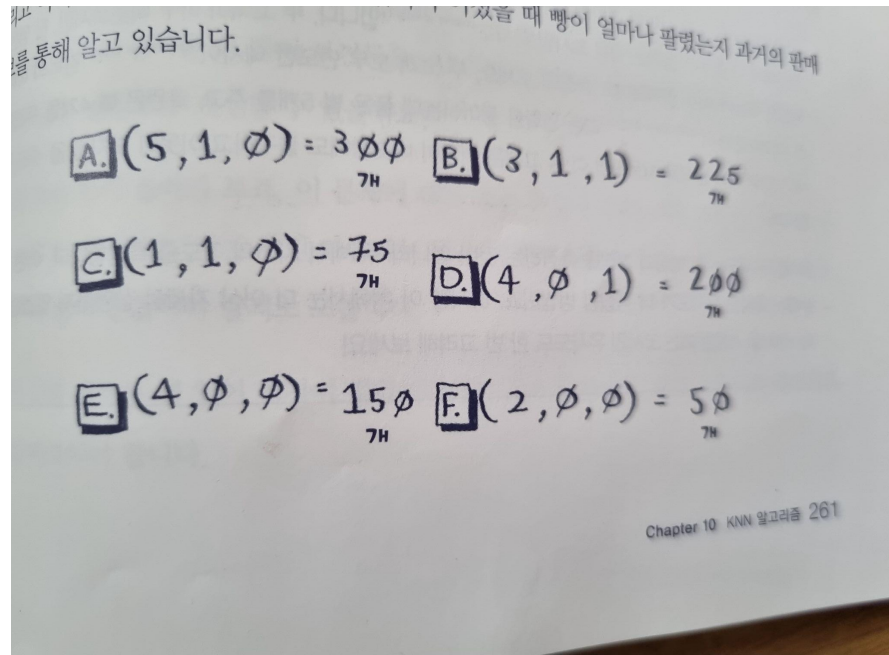
1. 1점부터 5점까지 숫자로 표현할 날씨(1=날씨가 최악, 5=날씨가 최고)
2. 주말 또는 휴일인지?(주말이나 휴일이면 1이고, 평일이면 0)
3. 스포츠 경기가 있는지?(있으면1, 없으면0)

오른쪽은 과거의 판매정보입니다.

오늘은 날씨가 좋은 주말입니다. K=4인 KNN을 사용해보겠습니다.

우선 가장 가까운 4개의 이웃 데이터를 찾습니다.

$$(4, 1, 0) = ?$$



각각의 데이터로부터 거리를 계산했습니다. A, B, D, E가 가장 가깝습니다.

A. 1

B. 2

C. 9

D. 2

E. 1

F. 5

이 날들의 평균을 구해보니 218.75입니다. 여러분은 오늘 빵을 이만큼 만들면 됩니다!