

✓ 두 점 사이의 거리 구하는 방법-유클리디안거리/맨하탄/해밍/체비세프/마할라노비스거리

✓ 두 점 사이의 거리를 구하는 방법

- 유사도(Similarity)와 관련이 있음
 - 거리가 가까울수록 해당 데이터가 가지고 있는 특징(feature)이 유사할 가능성이 크기 때문임

유사도의 3가지 조건 :

- 양의성질 (non-negativity) : $d(x,y) \geq 0$
- 대칭성 (symmetry) : $d(x,y) = d(y,x)$
- 유사성 성질 (similarity) : $s(x,y) \leq s(x,x)$

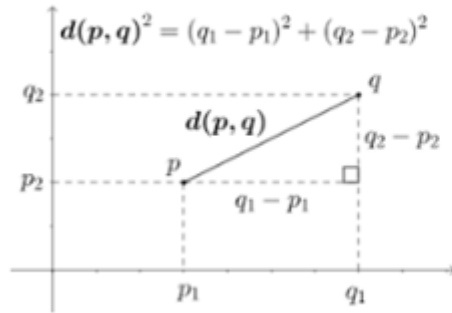
거리의 3가지 조건 :

- 양의성질 (non-negativity) : $d(x,y) \geq 0$
- 대칭성 (symmetry) : $d(x,y) = d(y,x)$
- 동일성 (self-identity axiom) : $d(x,y) = 0$ if and only if $x=y$

✓ 두 점사이 거리를 구하기 위한 대표적인 방법

- 아래 방식들은 데이터의 차원과 요소 개수가 동일해야 함

유클리드 거리(Euclidean Distance)



$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

```

1 def euclidean_distance(A, B):
2     distance = 0
3     for i in range(len(A)):
4         distance += (A[i] - B[i]) ** 2
5     return distance ** 0.5
6
7 print(euclidean_distance(A=[1, 5], B=[6, 15]))
8 #피타고라스 정리와 크게 다르지 않음
9 #다만 차수가 많아져도 계산할 수 있음
10 print(euclidean_distance(A=[1, 5, 7, 9], B=[2, 3, 6, 15]))

```

⇒ 11.180339887498949
6.48074069840786

맨해튼 거리(Manhattan Distance)

맨하탄 거리
(Manhattan Distance) $d_M(x, y) = \sum_{j=1}^n |x_j - y_j|$

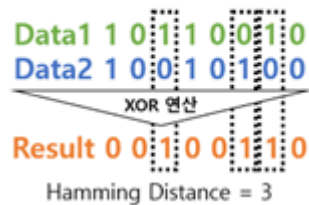


$$d = |(a_1 - b_1)| + |(a_2 - b_2)| + \dots + |(a_n - b_n)|$$

```
1 #택시 거리 또는 시가지 거리로도 불리는데 유클리드 기하학의 거리 공간을
2 #좌표 평면에 표시된 두 점 사이 거리차에 따른 새로운 거리 공간으로 대신하기도 함
3 def manhattan_distance(A, B):
4     distance = 0
5     for i in range(len(A)):
6         distance += abs(A[i] - B[i])
7     return distance
8
9 print(manhattan_distance(A=[1, 5, 7, 9], B=[2, 3, 6, 15]))
```

⇒ 10

해밍 거리(Hamming Distance)



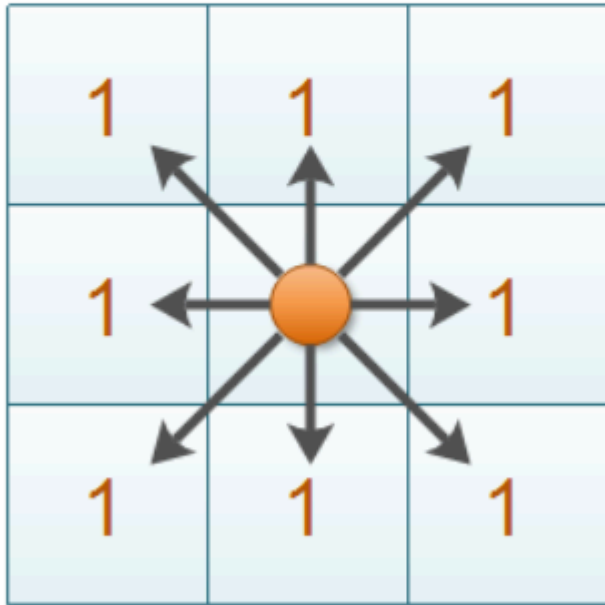
```
1 #해밍 거리는 각 요소마다의 차이가 아닌 정확한지에 대해서만 고려
2 #즉, 같은 길이의 문자열의 같은 위치에서 서로 다른 기호들의 개수를 알 수 있음
3 #예를 들어 Distance라는 단어가 있을 때 Distttce라고 오타가 났다면 이때 해밍 거리는 2가 됨
4 #또는 2진수 0011과 1010이 있을 때 해밍 거리는 2가 되는데 이를 이용해 신호학에서 오류 검출 시에도 이용됨
5 def hamming_distance(A, B):
6     distance = 0
7     for i in A:
8         if A[i] != B[i]:
9             distance += 1
10    return distance
11
12 print(hamming_distance([0, 0, 1, 1], [1, 0, 1, 0]))
```

 2

체비셰프 (Chebychev) 거리

$$d(x, y) = \max_i |x_i - y_i|$$

Chebyshev Distance



$$\max(|x_1 - x_2|, |y_1 - y_2|)$$

```
1 #체비셰프 거리(Chebyshev distance)
2 import math
3 def chebyshev_distance(lat1, lon1, lat2, lon2):
4     x = abs(lon2 - lon1) * math.cos(0.5 * (lat2 + lat1))
```

```
5     y = abs(lat2 - lat1)
6     return max(x, y)
7
8 #체비셰프 거리 계산
9 #예시: 서울(37.5665° N, 126.9780° E)과 도쿄(35.6895° N, 139.6917° E)의 거리
10 seoul_lat, seoul_lon = 37.5665, 126.9780
11 tokyo_lat, tokyo_lon = 35.6895, 139.6917
12 chebyshev_dist = chebyshev_distance(seoul_lat, seoul_lon, tokyo_lat, tokyo_lon)
13 print("Chebyshev Distance:", chebyshev_dist)
```

 Chebyshev Distance: 6.091751850239336

마할라노비스 거리

마할라노비스 거리는 두 점 사이의 거리를 측정하는 방법 중 하나이다. 이 거리는 두 점 사이의 거리를 측정하는 방법 중 하나이다.