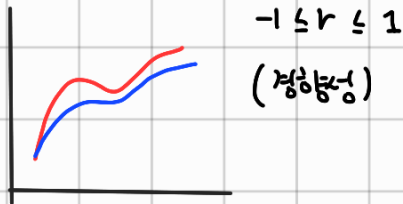


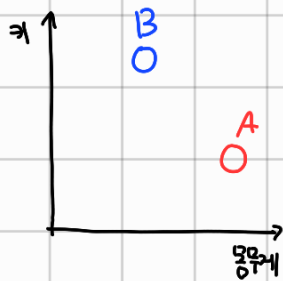
• 11강 코사인과 유사도

인공리듬 회귀에서 많이 씀

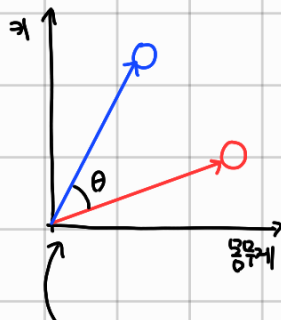
- 상관관계: 연속, 연관된 데이터의 유사도 측정



- 불연속적이고 연관성도 약한 데이터의 유사도 측정 → 다른 기법을 써야한다.



• 연관성X



• θ 가 작으면 유사하다.

• θ 가 작을수록, 같은 방향으로 가고 있다는 뜻

similarity

- 코사인 유사도 (θ 구하는 법)

$$-1 \leq \cos \theta \leq 1$$

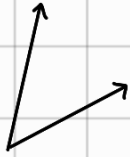
내적: $\vec{A} \cdot \vec{B} = |\vec{A}| |\vec{B}| \cos \theta$ ← 서로의 크기에 $\cos \theta$ 곱하기

$$\cos \theta = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| |\vec{B}|}$$

← 이때, $\cos \theta$ 를 코사인 유사도라고 한다.

- 코사인 유사도 ($-1 \leq \cos \theta \leq 1$)

서로 같은 방향



$$0 < \cos \theta \leq 1$$

• 비슷한 정향성

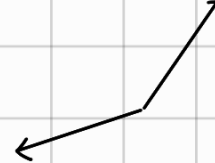
서로 관계 없음



$$\cos \theta = 0$$

• 서로 관계 없음

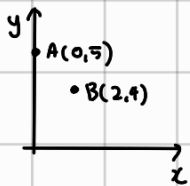
서로 다른 방향



$$-1 \leq \cos \theta < 0$$

• 반대되는 정향성

- 문제풀이) $A(0,5), B(2,4)$ 의 코사인 유사도는?



$$\cos \theta = \frac{A \cdot B}{|A||B|} = \frac{0 \cdot 2 + 5 \cdot 4}{\sqrt{25} \sqrt{4+16}} = \frac{20}{5 \sqrt{20}} = \frac{4}{\sqrt{5}} = \frac{2\sqrt{5}}{5}$$

크기로 스케일링 하기 때문에
크기가 커 비슷해지는 알 수 없다.
다만 정향성은 알 수 있다.

↑
양수이기
같은 방향이다.