

3주차. 데이터의 분류

- * 거리(distance) : $A(a_1, a_2)$, $B(b_1, b_2)$ 사이의 거리는 다음과 같이 계산한다.

$$\text{dist}(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

- * 노름 : 벡터 $\mathbf{a} = (a_1, a_2)$ 에 대하여 \mathbf{a} 의 크기를 다음과 같이 나타내고,
 \mathbf{a} 의 노름(norm)이라 한다.

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2} \quad (\text{벡터의 크기 계산: 노름(norm), 2-norm})$$

- * 내적 : 두 벡터 $\mathbf{a} = (a_1, a_2)$ 와 $\mathbf{b} = (b_1, b_2)$ 의 내적(inner product)은 다음과 같이 정의된다.

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2$$

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta, \quad \text{즉} \quad \cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \quad (0 \leq \theta \leq \pi)$$

- * 코사인 유사도(cosine similarity) : 두 데이터 $\mathbf{a} = (a_1, a_2)$ 와 $\mathbf{b} = (b_1, b_2)$ 의 코사인 유사도는 다음과 같이 계산할 수 있다.

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \left(\frac{\mathbf{a}}{\|\mathbf{a}\|} \right) \cdot \left(\frac{\mathbf{b}}{\|\mathbf{b}\|} \right)$$