

Statistic

1. 조건부 확률은 무엇일까요?

사건 B가 일어나는 경우에 사건 A가 일어날 확률은 말한다. 사건 B가 일어나는 경우에 사건 A가 일어날 확률은 $P(B|A)$ 로 표기한다. 사건 B가 발생했을 때 사건 A가 발생할 확률은 사건 B의 영향을 받아 변하게 된다.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (\text{단 } P(A) \neq 0)$$

베이즈 정리

→ 어떤 사건이 서로 배반하는 원인 중에 의해 일어났다고 할 때 실제 사건이 일어났을 때 이것이 두 원인 중 하나일 확률을 구하는 정리를 베이즈의 정리라고 한다.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

2. 공분산과 상관계수는 무엇일까요? 수식과 함께 표현해주세요

공분산이란 두 개의 확률변수의 선형적인 상관성의 정도를 나타내는 값이다.

$$\begin{aligned} \cdot \text{확률변수 } X, Y \text{의 공분산: } \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

두 변수가 양의 방향으로 선형적인 관계는 가진다면 공분산이 양의 값으로 커질 것이고, 음의 방향으로 선형적인 관계를 가진다면 공분산이 음의 값으로 커질 것이며 상관성이 없다면 0에 가까운 것이다.

상관계수란 표준화된 두 변수의 공분산 값이다.

$$\text{Corr}(X, Y) = \left(\frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} \right) \quad (-1 < \text{Corr}(X, Y) < 1)$$

상관계수는 두 변수의 공분산을 각 변수의 표준편차로 나눔으로써 표준화 효과를 가지게 된다.
1 또는 -1에 가까울수록 상관성이 큰 것이고, 0에 가까울수록 상관성이 작다.