Bayesian

Winter Vacation Capstone Study

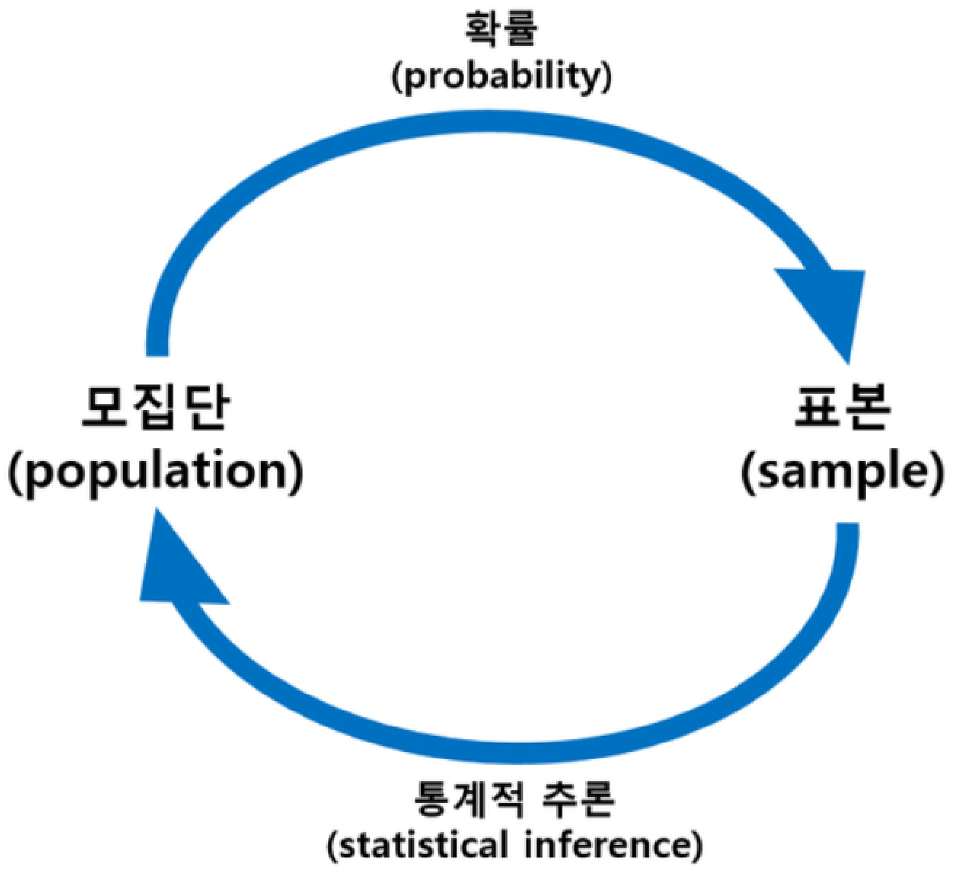
TEAM Kai.Lib

발표자 : 원철황

2020.01.06 (FRI)

확률과 통계

확률 : 오랜시간동안 경험적으로 쌓아온 확률현상에 대한 경험적 인식을 바탕으로 이론화한 것.
통계 : 자료를 수집한 뒤 분석, 해석 및 표현을 다루는 수학의 한 분야.



확률과 통계

1. 수학적 확률 : 어떤 사람이 계산하던지간에 동일한 값으로 계산되는 확률
2. 통계적 확률 : 동일조건&독립적으로 무한반복했을 때 발생하는 확률
3. 주관적 확률 : 관찰자의 주관에 따라 다르게 표현되는 확률

표본공간 (Sample Space): "통계적 실험에서 발생가능한 모든 결과들의 집합 " 으로 정의 가능하며 이를 S로 나타낸다.

표본공간 (Sample Space) 의 예시

S = {앞앞, 앞뒤, 뒤앞, 뒤뒤}

S = {앞앞앞, 앞앞뒤, 앞뒤앞, 뒤앞앞, 앞뒤뒤, 뒤앞뒤, 뒤뒤앞, 뒤뒤뒤}

$S = \{(x, y) \mid x^2 + y^2 \leq 9\}$

사건 (Event): "표본공간의 부분집합"이라 한다. 사건은 표본공간 밖에서 만들어질 수 없다.

정품을 T, 불량품을 F 라고 하면,

표본공간 S = {TTT, TTF, TFT, FTT, TFF, FTF, FFT, FFF} **불량품이 하나라도 존재하는 사건을 A라 하면**

사건 A = {TTF, TFT, FTT, TFF, FTF, FFT, FFF}

확률과 통계

1. 전사건(total event) : 표본공간 S의 모든 원소를 포함하는 사건.
2. 공사건(null event) : 표본공간 S의 어떤 원소도 포함하지 않는 사건.
3. 여사건(complementary event) : 포본공간 S의 사건 E에 속하지 않는 S의 모든 원소들의 집합인 사건
4. 합사건(union event) : 두 사건 E와 F에 대하여 E 또는 F 중 적어도 한 쪽은 일어나는 사건
5. 곱사건(intersection event) : 두 사건 E와 F에 대하여 E와 F가 동시에 일어나는 사건
6. 배반사건(mutually exclusive event) : 두 사건 E와 F에 대하여 E와 F 중 한 쪽이 일어나면 다른 쪽은 일어나지 않을 때, E와 F는 서로 배반이라 한다.(서로 공통부분이 없다.)

이렇게 사건들은 그 성격에 따라 불리는 이름이 달라진다. 그리고 표본공간의 부분 집합인 사건은 각각마다 가중치(비율, Ratio)를 가지는데 이를 확률이라 한다. 또는 다음과 같이 나타낼 수 있다.

사건(event) A의 확률은 사건 A 안에 있는 모든 원소에 할당된 확률의 합이다.
따라서 사건 A의 확률은 $0 \le P(A) \le 1$ 범위를 가진다. 또 $P(\emptyset)=0$, $P(S)=1$ 이다.
여기서 $\emptyset$는 ´공사건´, S는 ´전사건´을 의미한다.

표본공간의 원소들이 가지는 가중치, 즉 확률이 같다면 사건의 확률을 단순히 전체 원소와 해당 사건이 가지고 있는 원소의 비율로서 나타낼 수 있으나 이는 현실에서는 적용 불가능한 가정이다.

확률과 통계

Q. 빨간색 공 5개, 초록색 공 3개가 들어있는 주머니에서 공 3개를 임의로 꺼낼 때, 초록색 공이 2개 포함될 확률을 구하면?

8개의 공에서 임의로 3개를 꺼내는 방법의 수 : ${}_8C_3=56$

빨간색 공 1개를 꺼내는 방법의 수 : ${}_5C_1=5$

초록색 공 2개를 꺼내는 방법의 수 : ${}_3C_2=3$

따라서 빨간색 공 1개, 초록색 공 2개를 꺼내는 방법의 수는
곱의 법칙을 적용하면,

$${}_5C_1\times{}_3C_2=5\times3=15$$

따라서 구하고자 하는 확률은 $\frac{15}{56}$

확률과 통계

'베이즈 통계학'과 기존 통계학과의 큰 차이점은 기존 통계학에서는 모집단을 변하지 않는 대상으로 규정하지만, '베이즈 통계학'에서는 '모집단을 미리 확정 짓지 않는 것'이 특징이다.

베이즈 정리(bayes' theorem)는 이전의 경험과 현재의 증거를 토대로 어떤 사건의 확률을 추론하는 과정을 보여준다.

n개의 사건 $B_1, B_2, \dots, B_n$은 표본공간 S를 분할하고,

사건 A가 표본공간 S의 임의의 사건이라면,

$$P(B_i|A)=\frac{P(B_i)\cdot P(A|B_i)}{\sum_{i=1}^{n} P(B_i)\cdot P(A|B_i)},\ P(B_i)>0,\ P(A)>0$$

이고, 이것을 '베이즈 정리'라고 한다.

증명)

$$P(B_i|A)=\frac{P(B_i\cap A)}{P(A)}=\frac{P(B_i)\cdot P(A|B_i)}{P(A)}=\frac{P(B_i)\cdot P(A|B_i)}{\sum_{i=1}^{n} P(B_i)\cdot P(A|B_i)}$$

확률과 통계

확률 P(AIB)를 알고 있을 때, 관계가 정반대인 확률 P(BIA)를 계산하기 위해 등장한 개념이며 아래는 그 의미들이다.

P(A) : A의 사전 확률(evidence) - 현재의 증거

P(B) : B의 사전 확률(prior probability) - 과거의 경험

P(A | B) : 사건 B가 주어졌을 때 A의 조건부 확률(likelihood) - 알려진 결과(관찰된 결과)에 기초한 어떤 가설에 대한 가능성(즉, 이 가설을 지지하는 정도)

P(B | A) : 사건 A라는 증거에 대한 사후 확률(posterior probability) - 사건 A가 일어났다는 것을 알고, 그것이 사건 B로부터 일어난 것이라고 생각되는 조건부 확률

'결과를 관측한 뒤 원인을 추론할 수 있다.'는 특징을 가지며, 이는 조건부 확률 P(A I B)를 이용하여 다른 조건부 확률 P(B I A)를 구하는 것이다. 베이즈 정리에서는 새로운 경험과 새로운 정보는 새로운 조건을 가져다 준다. 따라서 정보가 증가함에 따라 확률이 수정되는 과정을 설명할 수 있다. 이 특징에 의해 머신러닝에서 분류 및 판단을 할 때 가장 중요하게 사용되는 개념이다.