통계학 맛보기

통계적 모델링은 적절한 가정 위에서 확률분포를 추정하는 것이 목표

유한한 개수의 데이터만 관찰해서 모집단의 분포를 정확하게 알아낸다는 것은 불가능하므로, 근사적으로 확률분포를 추정할 수 밖에 없다.



평균과 분산을 묶어서 모수라 부르고 특정 확률분포를 따른다고 가정하고 이를 추정하는 방법을 모수적(parametric) 방법론 이라 한다

특정 확률분포를 가정하지 않고 하면 비모수 방법론

비모수 방법론은 모수를 쓰지않는다? -> x(유연함)

텍스트이(가) 표시된 사진

자동 생성된 설명



텍스트이(가) 표시된 사진

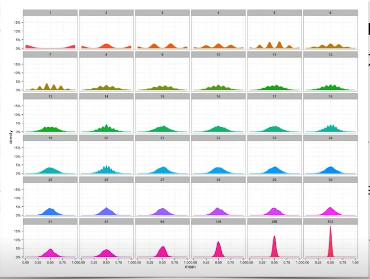
자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

표집분포와 표본분포는 다르다

이항분포의 표본분포는 데이터를 아무리 모아도 정규분포 될 수 없지만 표본평균의 확률분포는 정규분포가 된다. 중심극한정리 n 이 늘어날수록



텍스트이(가) 표시된 사진

자동 생성된 설명

Likelihood 지금 얻은 데이터가 이 분포로부터 나왔을 가능도를 말한다.

**각 데이터 샘플에서 후보 분포에 대한 높이(즉, likelihood 기여도)를 계산해서 다 곱한 것**을 이용할 수 있을 것이다.

텍스트이(가) 표시된 사진

자동 생성된 설명텍스트, 손목시계이(가) 표시된 사진

자동 생성된 설명

전체 표본집합의 결합확률밀도 함수를 likelihood func

위 식의 결과 값이 가장 커지는 theta를 theta^로 본다

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

Log likelihood 미분

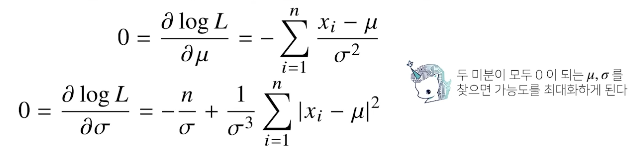
Log함수는 단조증가 따라서 likelihood와 log likeli의 최대값 갖게하는 정의역의 함수 입력값은 동일

최대값을 찾는 방법 중 가장 보편적 -> 미분계수가 0이 되게 하는 theta를 찾자

Mu로 미분하면 우측 시그마만 살아남아

Sigma로 미분하면 둘다 살아

둘다를 0을 만드는 뮤와 시그마를 구하자

텍스트, 화이트보드이(가) 표시된 사진

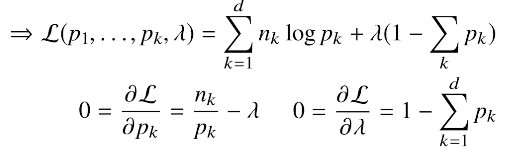
자동 생성된 설명

MLE는 불편추정량을 보장하진 않는다 1/n

텍스트이(가) 표시된 사진

자동 생성된 설명p1…pd 까지를 구하려면?

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

베이즈 통계학 맛보기

텍스트, 클립아트이(가) 표시된 사진

자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

D 우리가 관측할 데이터

Posterior 데이터가 주어져있을때 이것이 hypothesis theta 를 만족할 확률

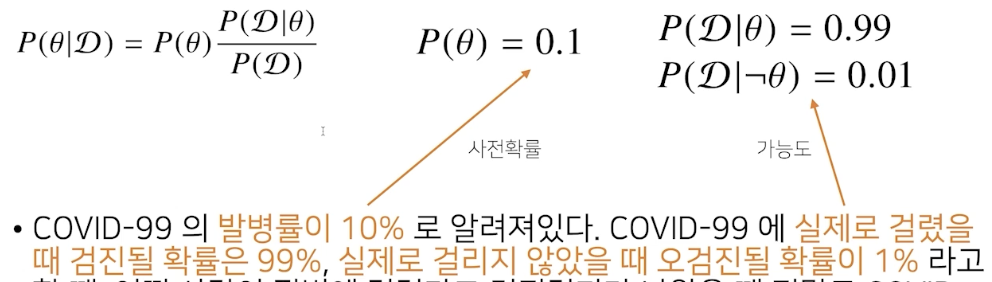
Prior 데이터를 관측하기 전에 가설을 미리 설정한다

Evidence 데이터 자체의 분포

likelihood theta가 주어져있을 때 데이터일 확률

텍스트이(가) 표시된 사진

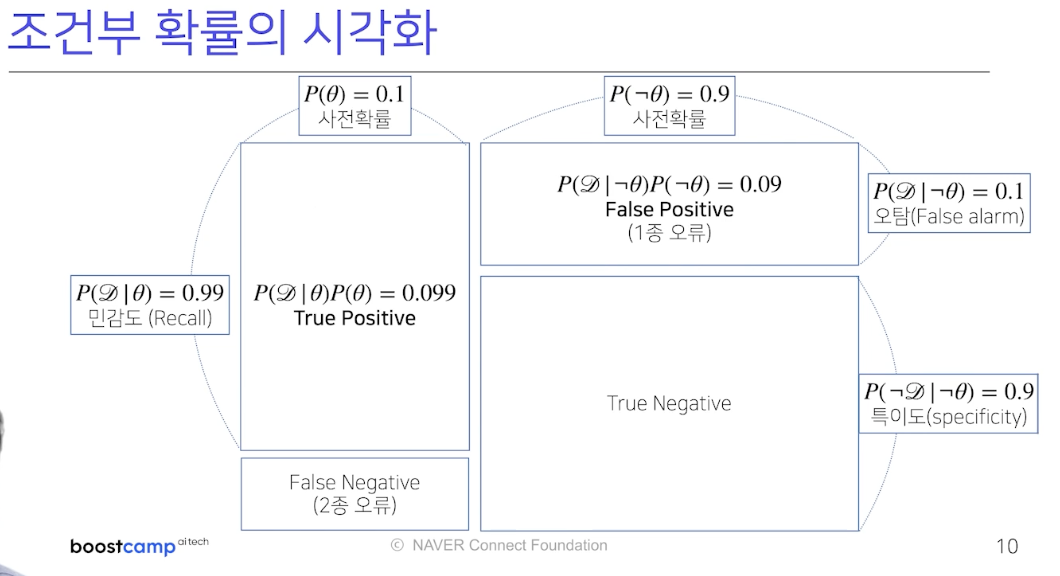
자동 생성된 설명



* Hypothesis: True이다. 즉 실제로 병이 있다.
* Evidence: Positive로 출력되었다. 즉, 병이 있다고 진단 받았다.
* 또한 문제에서 주어진 민감도와 특이도는 각각 True Positive와 True Negative에 해당되는 것이고 각각 P(E|H)=0.99P(E|H)=0.99, P(Ec|Hc)=0.98P(Ec|Hc)=0.98이다.



D D|theta. Theta. D|not theta. Not theta

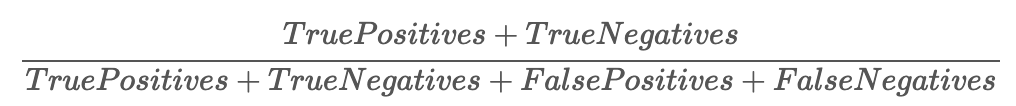


어떤 오류를 줄여할지는 목적에 따라 다름

테이블이(가) 표시된 사진

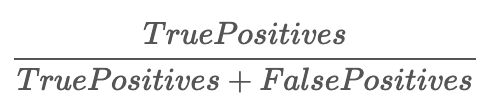
자동 생성된 설명

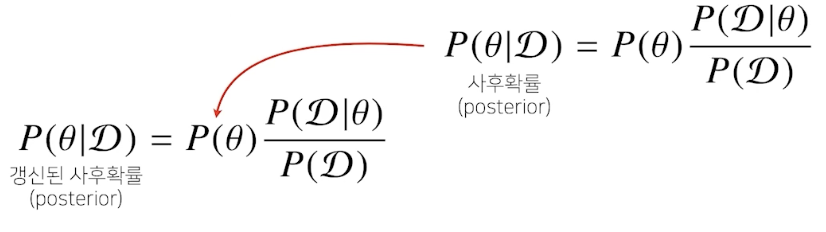
둘이 일치할때 true 예측이 true이면 positive

Acc 

Recall 텍스트이(가) 표시된 사진

자동 생성된 설명

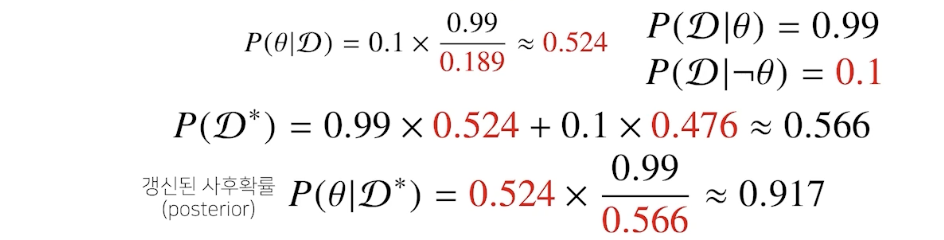
Precision 



사후확률의 갱신

첫번째 과정 거친 posterior가 0.524

이거를 다시 두번째 과정에서 prior로



텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트, 손목시계, 시계이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

전체적으로 봤을때는 b가 더 높은데 각각의 환자케이스로 보면 a가 둘다 높아 왜???

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명