6장. 불확실성의 처리

Contents

불확실성(Uncertainty)

자율주행 자동차

모든 법을 따르면서, 효율적이고 안전한 방법으로
A에서 B로 이동

- 불확실성의 예
  - 사고로 인한 교통 상황의 악화
  - 갑자기 나빠진 날씨
  - 도로에 갑자기 공이 튀어 나오는 것 같은 어떤 사건들
  - 자동차 카메라로 돌진해서 날아오는 쓰레기 조각
  - 센서 데이터의 노이즈(noise)
  - 감지 오류(sensing error)

불확실성은 왜 발생하나 ? (1/2)

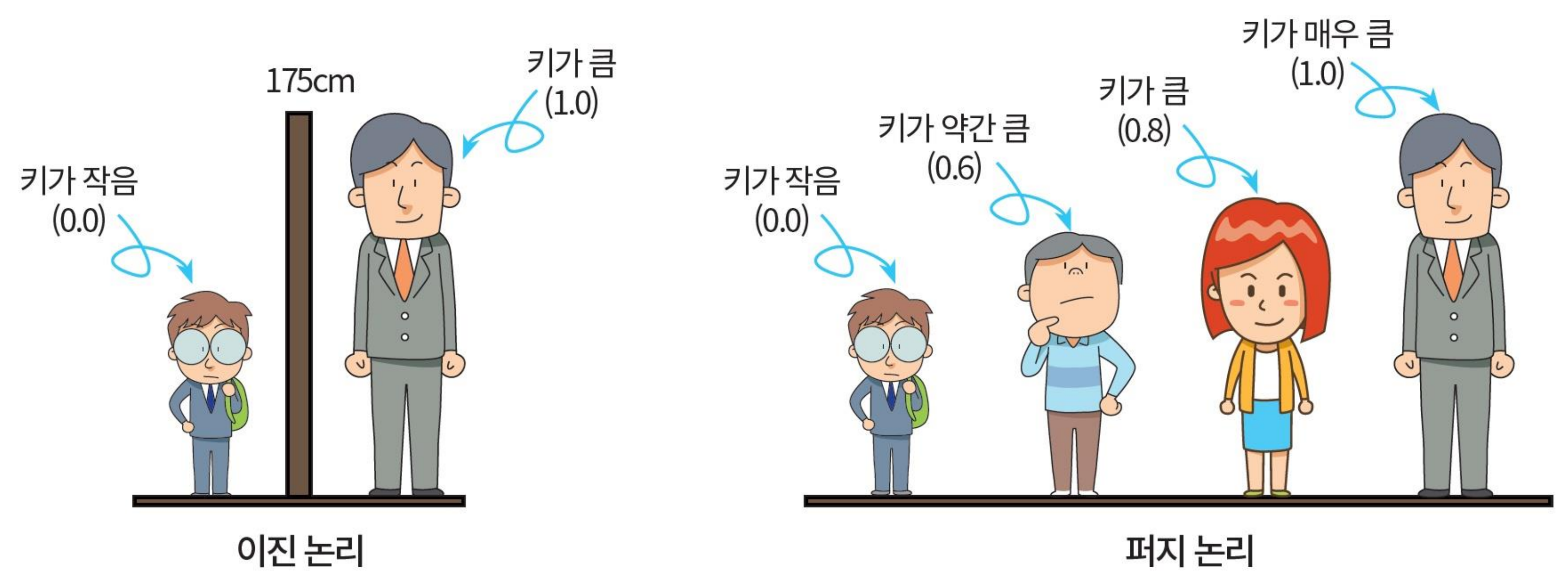
- 데이터의 불확실성
  - 센서 데이터의 오류/오차, 측정할 수 없는 데이터
- 지식의 불완전성
  - 전문가로부터 얻은 지식의 불완전성, 문제 영역 자체의 불확실성, 지식획득의 부실
- 정보 획득의 불완전성
  - 모든 정보를 수집하고 처리하는 것은 불가능.
- 부정확한 언어
  - 자연어의 모호성

불확실성은 왜 발생하나 ? (2/2)

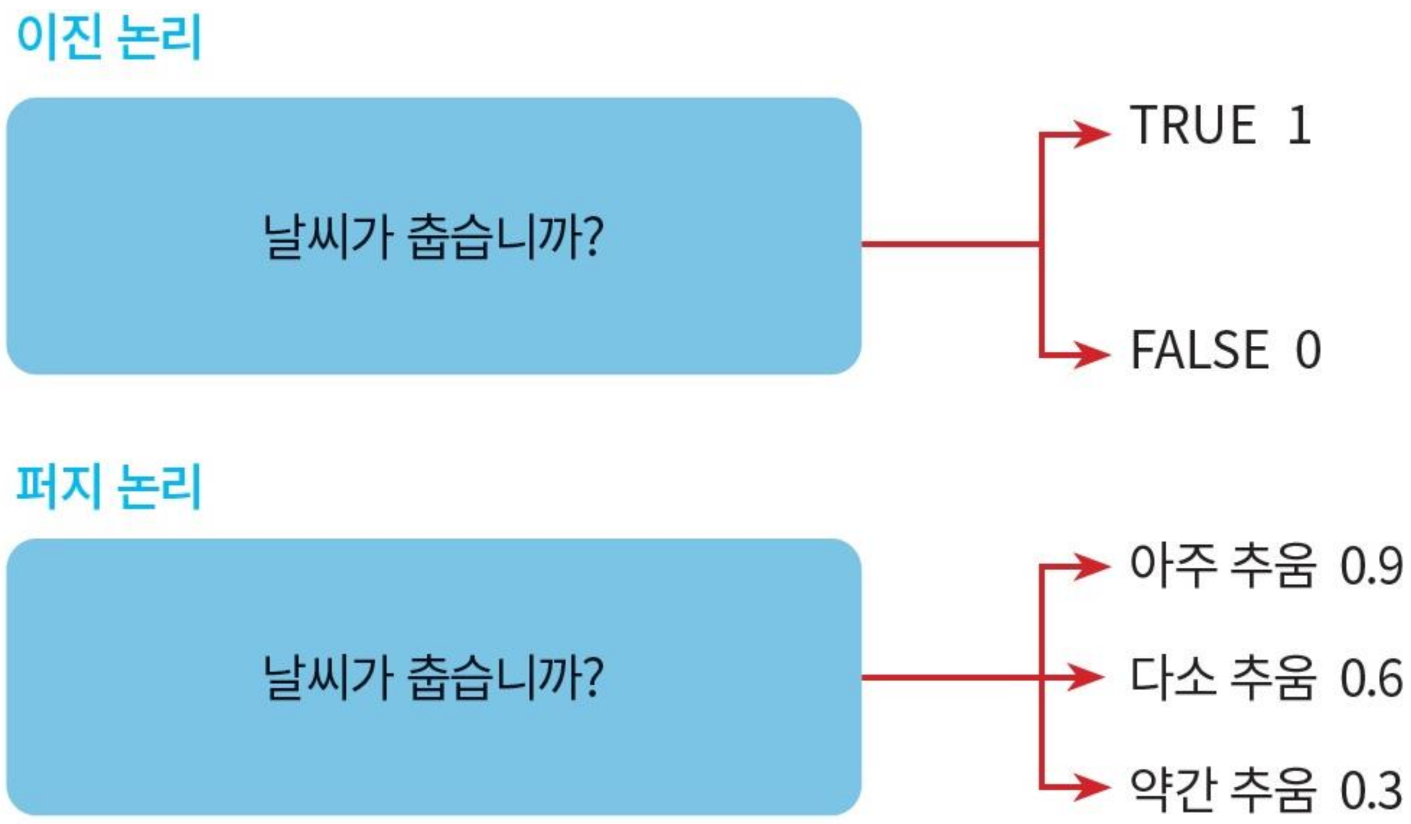
- 실제로는 어떠한 사실이나 지식에 대한 확실성이 100%인 경우는 없음
- 불가피하게 제기되는 불확실성을 표현하고 처리할 수 있는 기능 포함되지 않으면 매우 한정될 수 밖에 없음
  - 통계적 확률을 표현하는 방법인 베이즈 정리(Bayes' Theorem)
  - 실생활에서 자연어의 모호성을 해결하기 위한 퍼지 이론(Fuzzy Theory, 모호성)

불확실성의 처리 - 퍼지이론

- 퍼지 논리(Fuzzy logic)
  - 불확실성과 모호성을 다루기 위한 수학적 이론
  - 명확하게 정의될 수 없는 지식을 표현하는 방법 ⇒ 불확실하고 부정확한 정보를 처리
  - 0.0에서 1.0까지의 진리값을 가짐



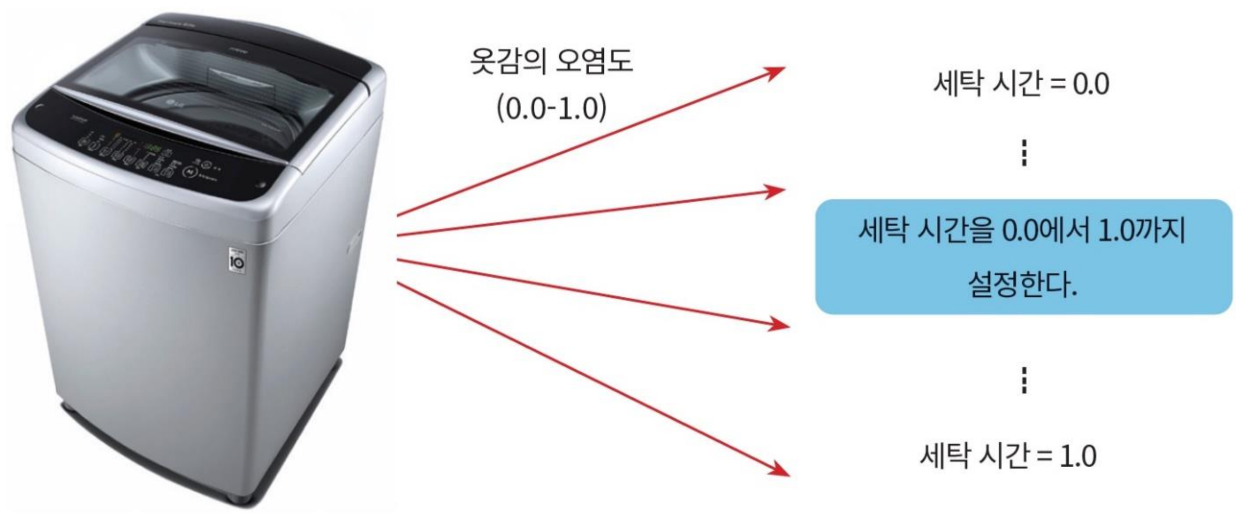
이진 논리 vs. 퍼지 논리

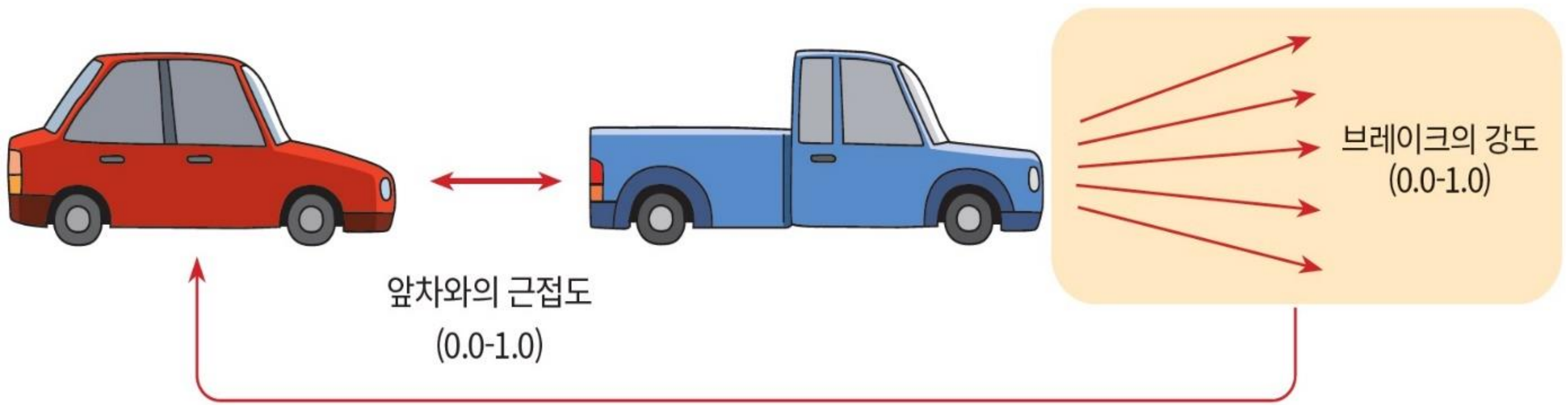


퍼지 집합의 소속도 (퍼지명제의 진리도)

퍼지 논리

- 퍼지논리는 주로 제어분야에서 사용됨





퍼지 집합 (1/3)

- 집합과 소속함수(membership function)
  - 집합 : 어떤 조건에 따라 결정되는 요소들의 모임
  - 원소들은 집합에 속하거나 속하지 않거나

$$\mu_A(x) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A \end{cases}$$

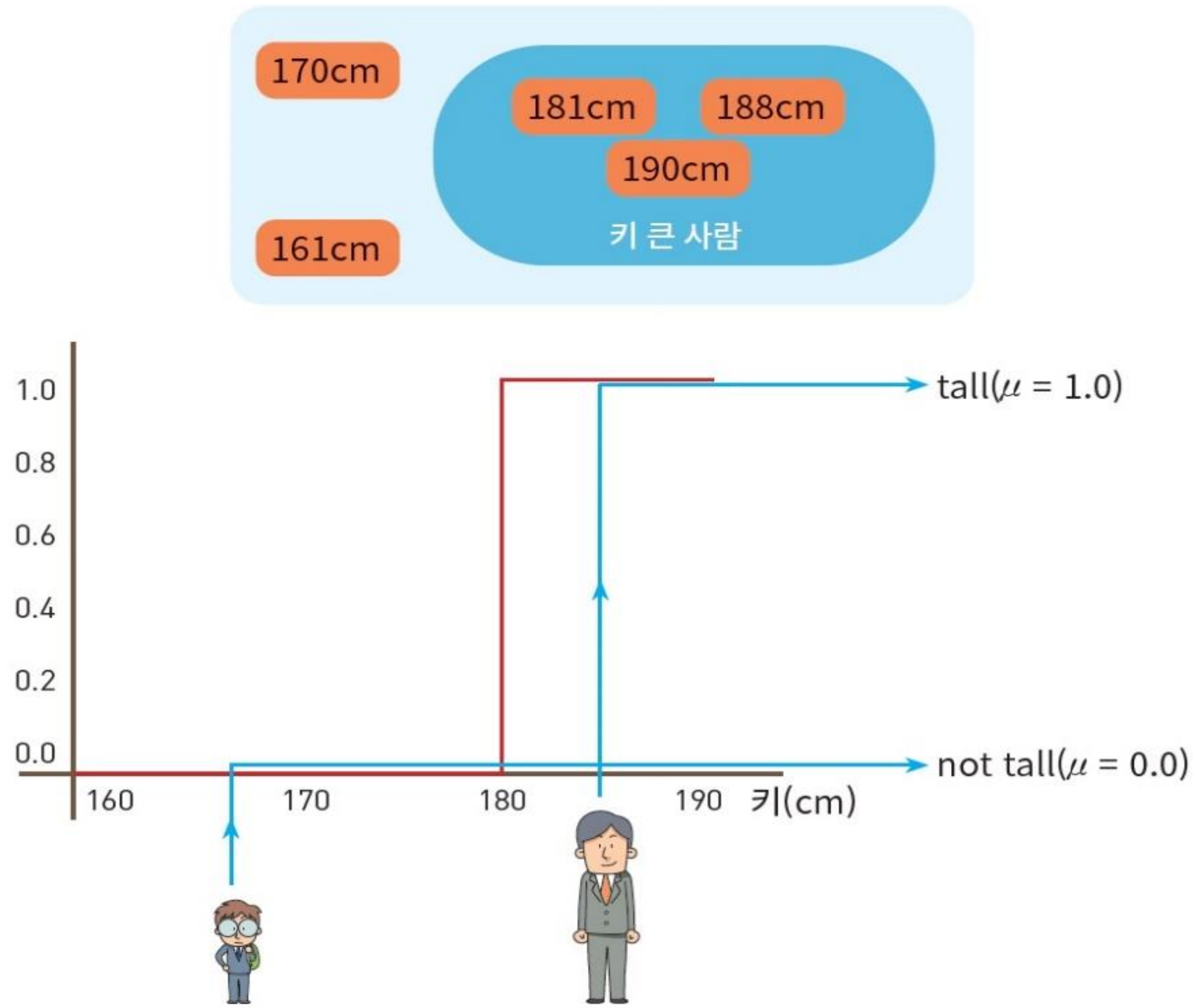
$$\mu_A : A \rightarrow \{0, 1\}$$

- 명제 논리 == 크리스프 집합(Crisp set)
- 퍼지 논리 == 퍼지 집합(Fuzzy set)

퍼지 집합 (2/3)

- 크리스프 집합
  - 기존의 집합이론
  - 속하든지 속하지 않은 것
  - 소속함수로 표현

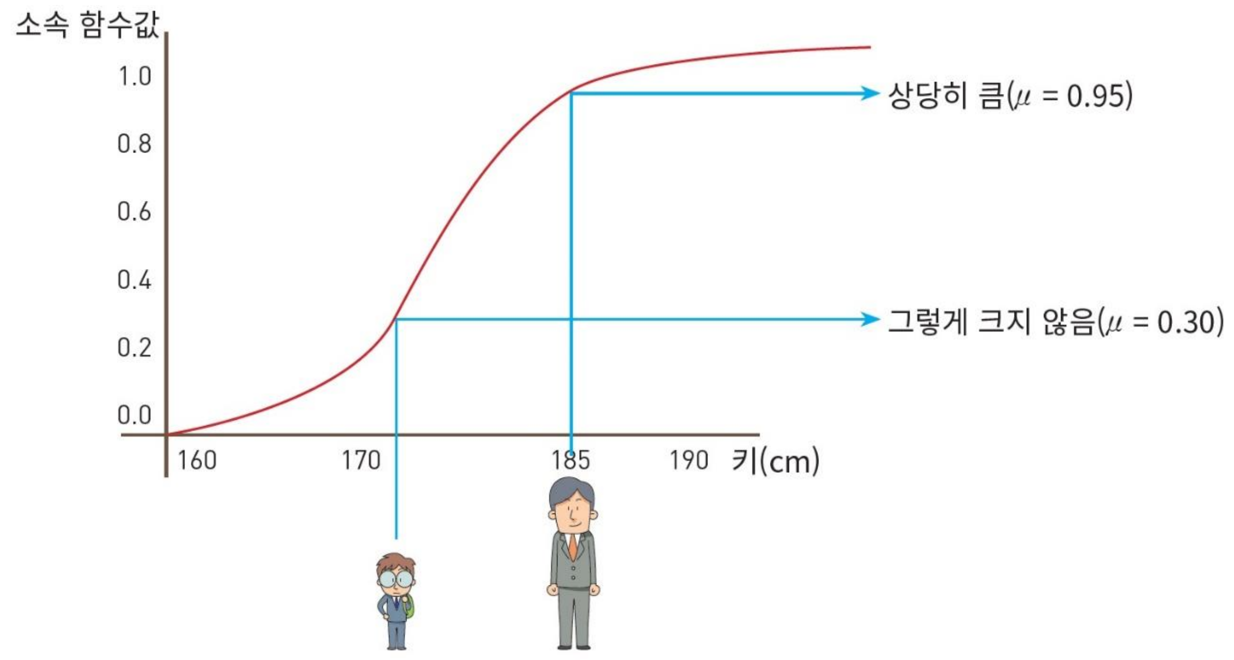
    (Membership function)



퍼지 집합 (3/3)

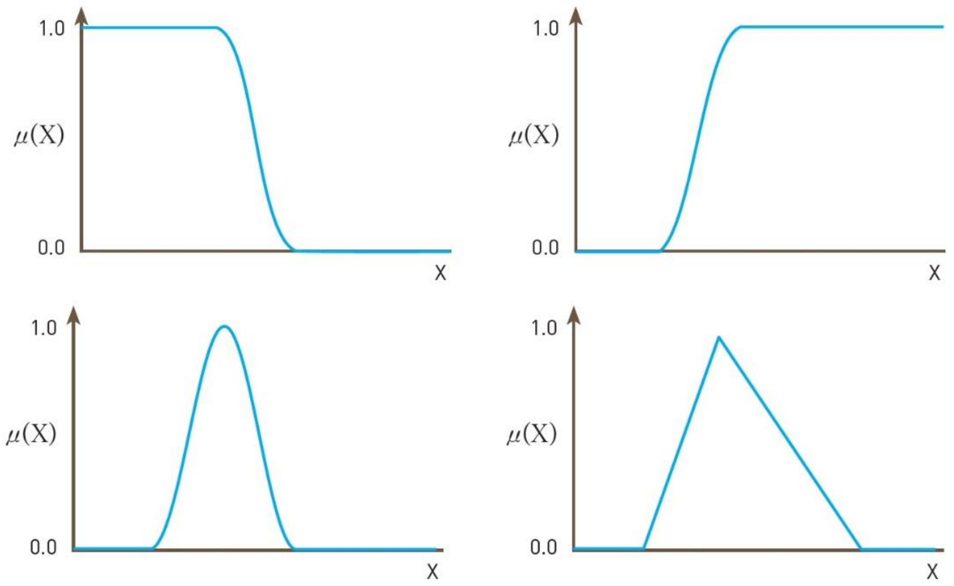
- 퍼지 집합
  - 원소가 집합에 속하는 정도에 따라 소속함수 값이 0과 1사이 값(모든 실수)으로 대응 $\mu_A : A \rightarrow [0, 1]$
  - 예) “키 큰 사람”이라는 집합

{ 0.3/172cm, 0.5/175cm, 0.95/185cm, 1.0/190cm }



퍼지집합의 표현

- 비연속적인 퍼지 집합
  - “키 큰 사람” = { 0.30/170cm, 0.50/175cm, 0.95/180cm, 1.0/190cm }
  - “키 큰 사람” = { (170cm, 0.3), (175cm, 0.5), (180cm, 0.95), (190cm, 1.0) }
- 연속적인 퍼지집합
  - 함수로 표시
  - 시그모이드 함수, 가우스 함수, 선형 함수 등



퍼지 추론 (1/2)

- 기존의 추론

규칙 #1: IF 온도가 높다. THEN 팬의 속도를 증가시킨다.
사실 #1: 온도가 약간 높다.

추론된 사실: ???

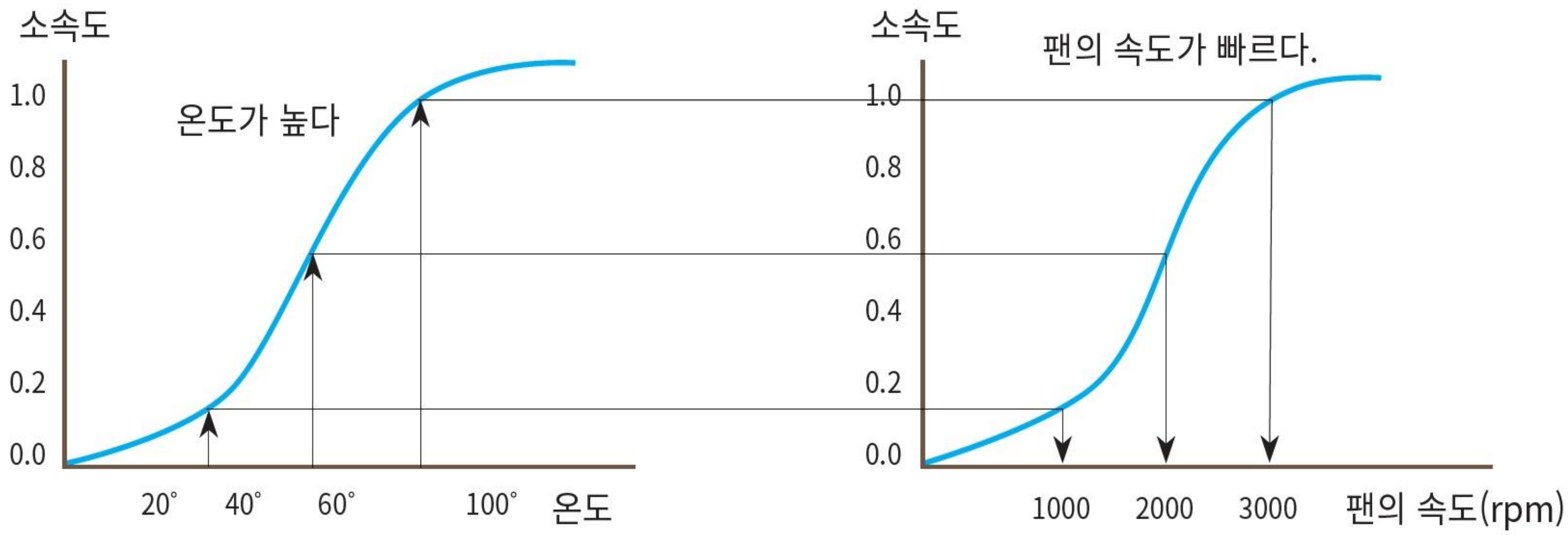
- 퍼지 추론

규칙 #1: IF 온도가 높다. THEN 팬의 속도를 증가시킨다.
사실 #1: 온도가 약간 높다.

사실 #2: 팬의 속도를 약간 빠르게 한다.

퍼지 추론 (2/2)

규칙 #1: IF 온도가 높다.
THEN 팬의 속도를 증가시킨다.



불확실성의 처리 - 베이즈정리

- 베이즈 추론(Bayesian reasoning)
  - 가능성을 다루는 수학적 도구인 확률이론(probability theory)를 기반으로 함
  - 확률적인 추론에서는 명제의 진리값이 참(1) 또는 거짓(0)이 아니고, 0.0 ~ 1.0 사이의 값을 갖게 됨
  - 확률적인 추론에서는 가장 확률이 높은 결론이 선택 됨

확률 (1/3)

- 불확실한 상황에서 추론하기 위한 최선의 접근방식
- 믿음의 정도를 다루는 수학적 도구 ⇒ 일상생활의 위험을 <u>수량화</u> 하고 비교하는데 사용할 수 있음
  - 제한속도를 초과하면 자동차가 충돌할 확률은 ?
  - 모기지 이자율이 향후 5년 이내에 5%가 오를 확률은 ?
  - AI가 X-선 이미지에서 골절된 뼈를 감지할 수 있는 가능성은 ?

<u>불확실성의 정량화</u>

확률 (2/3)

- 확률(Probability)
  - 특정한 사건(event)가 발생하는 비율
  - 표본공간 S가 n개 사건으로 이루어져 있고, 각 사건이 일어날 가능성이 같다면, 확률은 $1/n$
  - 예) 주사위를 한번 던져서 3이 나올 확률
    - 사건의 개수는 6이므로, $p = {}^{1}/_{6} = 0.166$
    - 3이 나오지 않을 확률은 $q = {}^{5}/_{6} = 0.833$
      - ➢ $p + q = 1$
  - 주사위 실험에서 1과 3이 나오는 사건은 상호배타적

확률 (3/3)

불확실한 추론을 자동화

사례

- 의료진단(Medical diagnosis)
- 스팸 메일 식별

- 확률적 추론에서는 명제의 진리값이 0.0 ~ 1.0 사이의 값
- 확률이 0 : 완전히 불가능한 사건
- 확률이 1 : 완전히 확실한 사건

연습문제 : 확률적 예측

문제	맞음	틀림	결론X
일기예보는 내일 비가 올 확률이 90%라고 했는데, 그 날은 온통 해가 지고 비가 내리지 않는 것으로 나타났다.			
일기예보에서는 내일 비가 올 확률이 0%라고 했는데 그날은 비가 왔다.			
오랫동안 일기예보를 모니터링한다고 가정해 보자. 예보에서 비가 올 확률이 80%인 날만 고려한다. 장기적으로 볼 때 평균적으로 5일 중 3일은 비가 내린다.			
2016년 미국 대통령 선거에서 잘 알려진 정치 예측 블로그는 클린턴에게 71.4%의 승리 확률을 제공했다. 그러나 예상과 달리 도널드 트럼프가 미국의 45대 대통령에 당선됐다.			

맞음 = 예측에 의해 주어진 확률이 정확한 확률
틀림 = 예측이 틀렸음
결론 x = 어느 쪽으로도 결론을 내릴 수 없음

연습문제 : 확률적 예측

문제	맞음	틀림	결론X
일기예보는 내일 비가 올 확률이 90%라고 했는데, 그 날은 온통 해가 지고 비가 내리지 않는 것으로 나타났다.			V
일기예보에서는 내일 비가 올 확률이 0%라고 했는데 그날은 비가 왔다.		V	
오랫동안 일기예보를 모니터링한다고 가정해 보자. 예보에서 비가 올 확률이 80%인 날만 고려한다. 장기적으로 볼 때 평균적으로 5일 중 3일은 비가 내린다.		V	
2016년 미국 대통령 선거에서 잘 알려진 정치 예측 블로그는 클린턴에게 71.4%의 승리 확률을 제공했다. 그러나 예상과 달리 도널드 트럼프가 미국의 45대 대통령에 당선됐다.			V

승률(Odds)

- 불확실성을 나타내는 방법
- 확률 vs. 승률
  - 확률(Probability) : 가능한 모든 이벤트 중에서 특정 이벤트가 발생하는 비율 (항상 0~1 사이의 값)
    - A ratio of PART : WHOLE
    - 두 사건의 확률이 동일하다 = 0.5
  - 승률(Odds) : (Event E가 발생할 확률)/(Event E가 발생하지 않을 확률) (0 ~ ∞)
    - A ratio of PART : PART
    - 승률이 동일하다 = 1번 실패하면 1번 성공 = 1

Contents

확률이 어떻게 변하는가?

사전 확률(Priori Odds)

새로운 정보로 베이즈 규칙 적용

사후 확률(Posterior Odds)

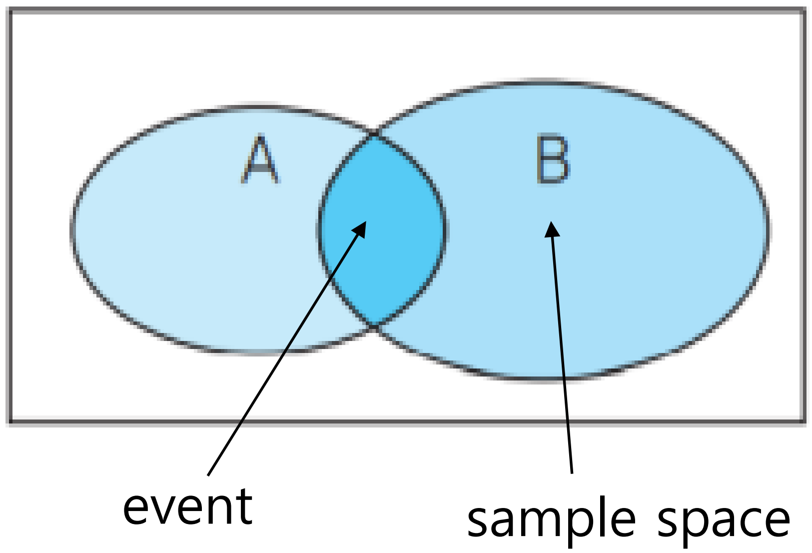
예제

- 오늘 늦게 서울에 비가 올 확률 = 206/365 (=0.56)
  - ➢ 비에 대한 사전 확률 = 206 : 159
  - 아침에 일어나 보니 날씨가 흐림
  - 비 오는 날 구름이 생길 확률 = 9/10
  - 비가 오지 않는 날 구름이 생길 확률 = 1/10
  - 비가 오지 않는 날에 비해 비오는 날에 구름이 생길 확률이 얼마나 더 높습니까? 9배 (가능도, likelihood ratio)
  - 베이즈 규칙 : 사후확률 = 가능도 * 사전확률
  - ➢ 비가 올 사후 확률 = 9 * 206 :159 ⇒ 1854/(1854+159) = 0.92

사전 확률과 사후 확률 (1/2)

- 조건부 확률(Conditional probability)
  - 두 사건 A, B가 있고 상호배타적이지 않다고 가정하면, 사건 B가 발생 했을 때 사건 A가 발생할 확률

$$p(A|B) = \frac{A\text{와 } B\text{가 동시에 발생할 경우의 수}}{B\text{가 발생할 경우의 수}}$$



$p(A)$: 사건 A에 대한 사전확률

$p(B)$: 사건 B에 대한 사후확률

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

사전 확률과 사후 확률 (2/2)

- 조건부 확률(Conditional probability)
  - 결합확률 $p(A \cap B)$: A와 B가 동시에 발생할 확률

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \qquad p(B|A) = \frac{p(A \cap B)}{p(A)}$$

$$p(A \cap B) = p(B|A) \times p(A)$$

$$\Rightarrow \boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}}$$

베이즈 정리 (1/2)

- 조건부 확률을 계산하기 위한 통계학적 정리
- 확률을 이용해 일어나지 않은 미래에 영향을 줄 확률을 구하는 방법

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

베이즈 정리 (2/2)

가능도(likelihood)

사전확률
(Priori probability)

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

사후확률
(Posterior probability)

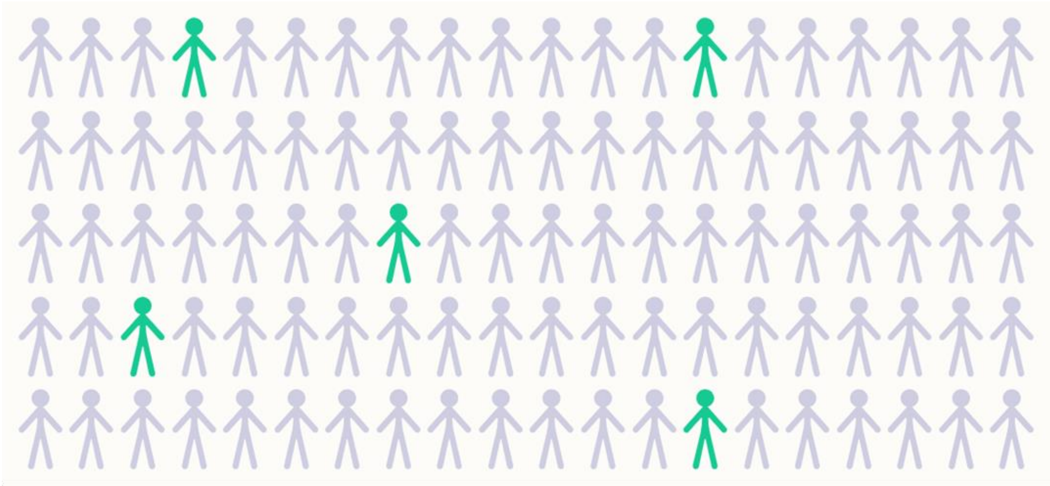
사전확률
(Priori probability)

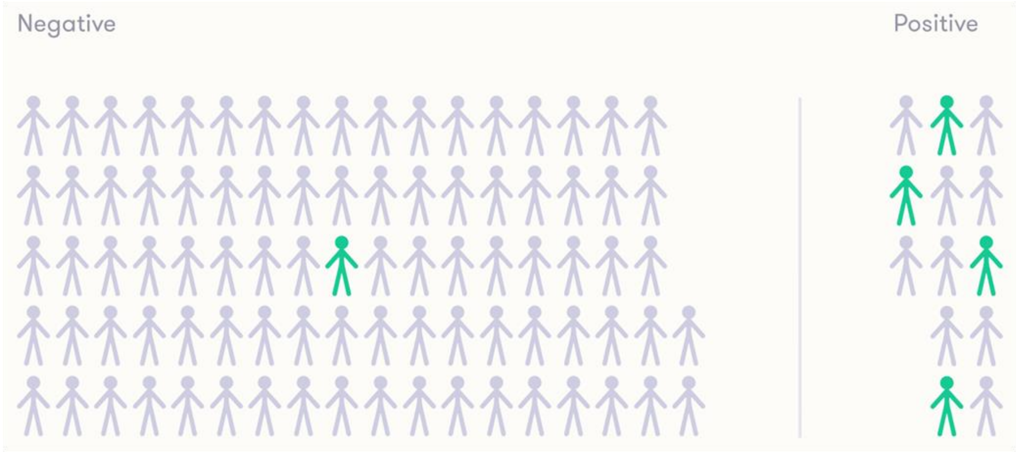
- 가능도(likelihood ratio)
  - 관심 이벤트에 대한 관측 확률을 이벤트가 없는 경우의 관측 확률로 나눈 것

사례 : 유방암 검진

- 민감도(sensitivity) = 80%
- False-positive = 10%
- 100명 중 5명이 유방암에 걸렸다고 가정
  - ➢ 95명중 9.5명(약9명)은 거짓 양성
  - ➢ 86명은 negative

진단 전





진단 후

유방암 걸릴 확률은 ? (1/3)

- 베이즈 규칙 적용하기 (1)
  - 사전확률 : 여성 100명 중 평균 5명이 유방암에 걸리므로,

    5 : 95
  - 가능도 : 암이 있는데 양성이 나올 확률이 암이 없는데 양성으로 나올 확률보다 몇배나 많은가 ?

(80/100) : (10/100) => 8배

유방암 걸릴 확률은 ? (2/3)

- 사후 확률

  $8 * 5 : 95 = 40 : 95 \Rightarrow 40 / (40+95) = 0.3$

양성 검사 결과가 있을 때 유방암 가능성은 약 30%

유방암 걸릴 확률은 ? (3/3)

- 베이즈 규칙 적용하기 (2)

$$p(C|P) = \frac{p(P|C) \times p(C)}{p(P)}$$

  - 민감도(sensitivity) = 80% $\Rightarrow p(P|C) = 0.8$
  - 86명은 negative $\Rightarrow$ 14명이 positive $\Rightarrow p(P) = 0.14$
  - 암에 걸린 사람 5명 $\Rightarrow p(C) = 0.05$

$$p(C|P) = \frac{p(P|C) \times p(C)}{p(P)} = \frac{0.8 \times 0.05}{0.14} \cong 0.29$$

베이즈 정리와 추론 (1/3)

- 전문가 시스템의 규칙

> *IF E is true*
> *THEN   H is true (확률: p)*

- (예)

> *IF 기침을 한다.*
> *THEN   감기이다. (확률: 0.7)*

➢ 규칙의 확률 계산은 어떻게?

베이즈 정리와 추론 (2/3)

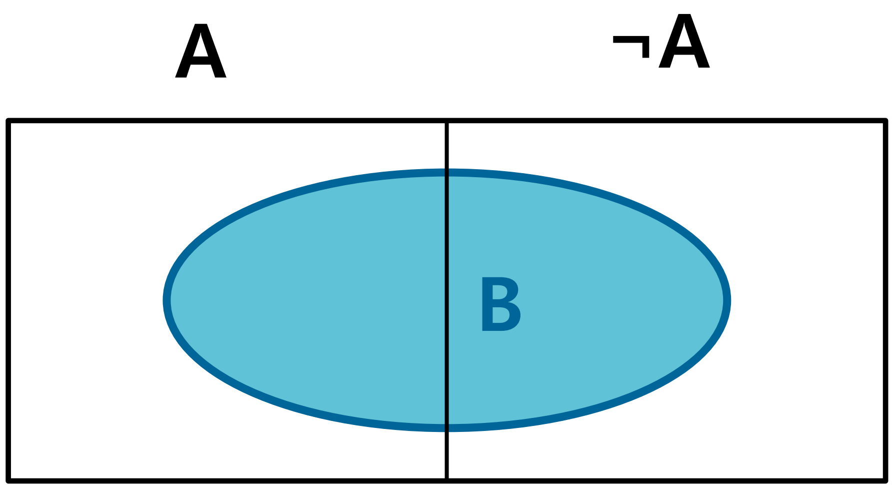
논리연산자 NOT

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)}$$

- $p(H)$ - 가설 H가 참일 확률(사전 확률)
- $p(\neg H)$ - 가설 H가 거짓일 확률(사전 확률)
- $p(E|H)$ - 가설 H가 참일 때, 증거 E가 나타날 확률(사후 확률)
- $p(E|\neg H)$ - 가설 H가 거짓임에도, 증거 E가 나타날 확률(사후 확률)

베이즈 정리와 추론 (3/3)

- 전확률 정리



$$p(B) = p(B \cap A) + p(B \cap \neg A)$$

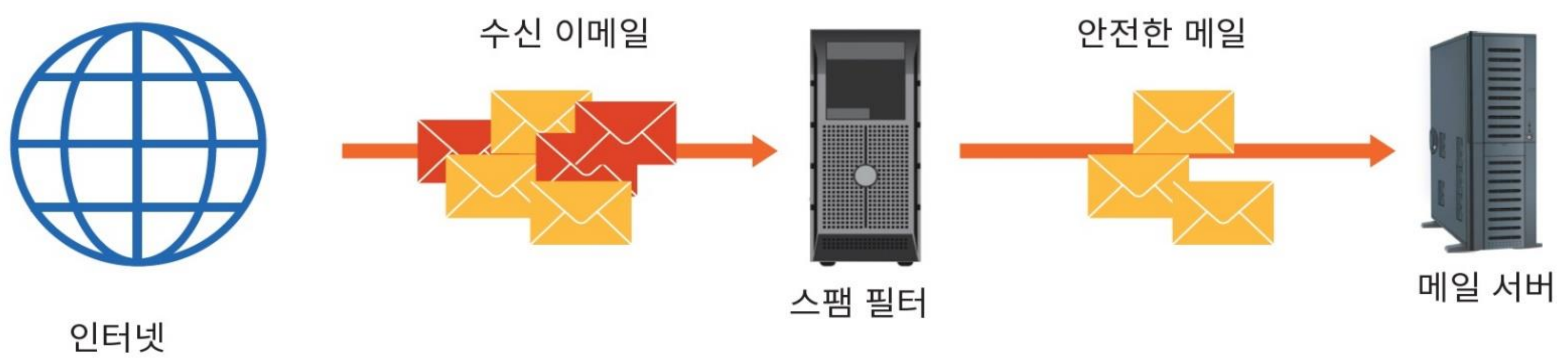
$p(A \cap B) = p(B|A) \times p(A)$이므로

$$p(B) = p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)$$

$$\boxed{p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)}}$$

사례 : 스팸 필터링

- 특정 단어가 들어가 있으면, 스팸일 가능성이 높아짐
  - 메일의 제목에 특정 단어가 들어가 있는 경우를 가정
- 조건부 확률 $p(spam|word)$ 계산하기
  - 전체 메일의 40%가 스팸메일
  - 스팸 메일의 1%가 제목으로 특정 단어를 가짐
  - 정상 메일 중 0.4%가 제목으로 특정 단어를 가짐



스팸 메일일 확률은 ?

- 전체 메일의 40%가 스팸메일 : $p(spam) = 0.4$
- 스팸 메일의 1%가 제목으로 특정 단어를 가짐 : $p(word|spam) = 0.01$
- 정상 메일 중의 0.4%가 제목으로 특정 단어를 가짐 : $p(word|\neg spam) = 0.004$

$$p(spam|word) = \frac{p(word|spam) \times p(spam)}{p(word|spam) \times p(spam) + p(word|\neg spam) \times p(\neg spam)}$$

$$= \frac{0.01 \times 0.4}{0.01 \times 0.4 + 0.004 \times 0.6} = \frac{5}{8} = 0.625$$

약 63%의 확률로 특정 단어가 들어간 메일은 스팸

연습문제 : 좀비 바이러스

- 테스트 키트로 검사했을 때, positive로 나왔다면 이 사람이 좀비일 확률은 ?
  - 세계 인구의 10%가 Z-바이러스에 감염
  - 테스트키트의 정확도 90% 즉, Z-바이러스 감염자이면 90% 확률로 positive로 판정
  - Z-바이러스 감염자가 아니면 70%의 확률로 negative로 판정

연습문제 : 좀비 바이러스

- 세계 인구의 10%가 Z-바이러스에 감염 : $p(Z) = 0.1$
- 테스트키트의 정확도 90% 즉, Z-바이러스 감염자이면 90% 확률로 positive로 판정 : $p(+|Z) = 0.9$
- Z-바이러스 감염자가 아니면 70%의 확률로 negative로 판정 : $p(+|\neg Z) = 0.3$

$$p(Z|+) = \frac{p(+|Z) \times p(Z)}{p(+|Z) \times p(Z) + p(+|\neg Z) \times p(\neg Z)}$$

$$= \frac{0.9 \times 0.1}{0.9 \times 0.1 + 0.3 \times 0.9} = 0.25$$

Contents

Naïve Bayes Classifier

- 베이즈 규칙의 가장 유용한 응용
  - 텍스트 문서와 같은 객체를 둘 이상의 클래스로 분류하는데 사용할 수 있는 기계학습 기술
  - 올바른 클래스가 제공되는 훈련 데이터 세트를 분석해서 훈련
  - 다양한 관측이 주어진 클래스의 확률을 결정하는데 사용
  - 가정 : 특성 값을 표현하는 특징 변수(feature variables)는 모두 서로 독립적임
    - 독립적 = 특성들 사이에서 발생할 수 있는 연관성이 없음

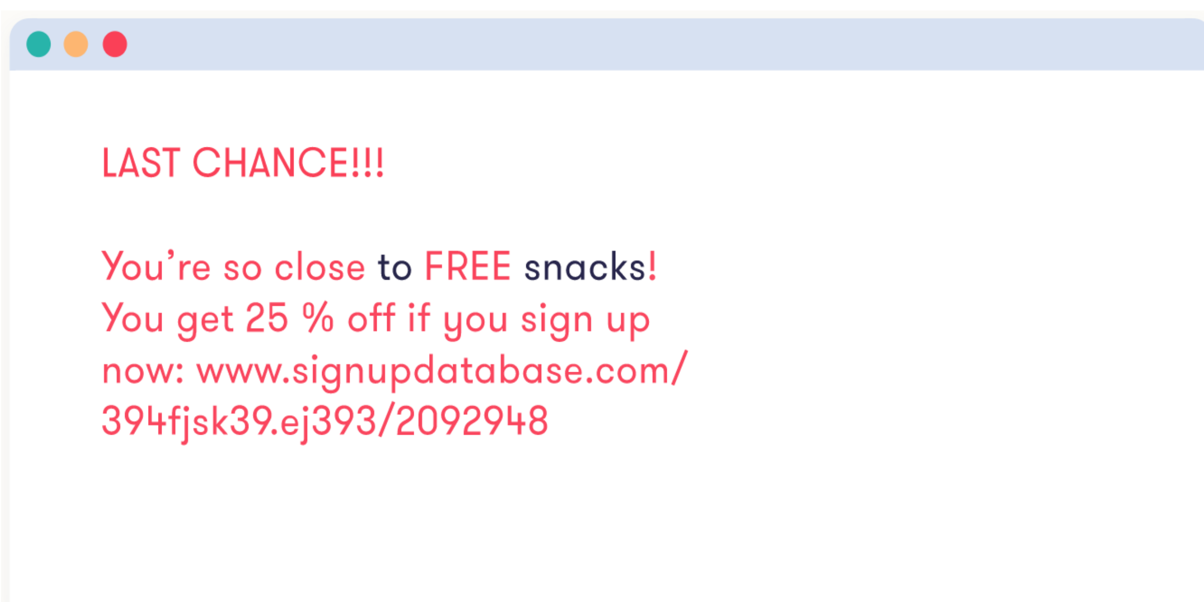
사례 : 스팸 메일 필터

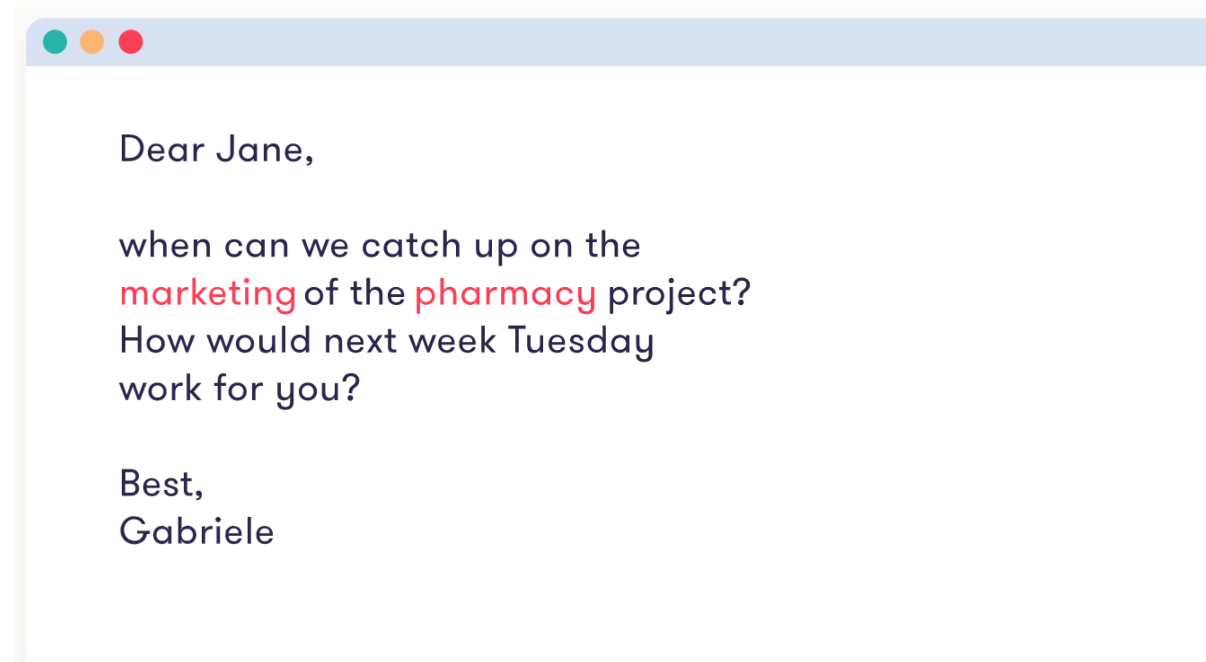
- 클래스 변수는 메시지가 스팸(정크메일)인지 또는 햄 (합법적인 메시지)인지 여부를 나타냄
- 메시지의 단어는 특징변수에 해당하므로, 모델의 특징변수 수는 메시지의 길이에 따라 결정됨
- Naïve 라고 하는 이유
  - 단어의 선택이 그 메시지가 스팸인지 햄인지에만 의존하도록 함
  - 인접한 단어 사이의 종속성이 없고, 단어의 순서가 중요하지 않음 ⇒ 프로세스를 심하게 단순화 시킴

어떻게 훈련하는가 ?

- 올바른 훈련 데이터의 셋을 분석해서

spam : ham





매개변수 추정

- 햄에 대한 스팸의 사전확률 = 1:1

word	spam	ham
million	156	98
dollars	29	119
adclick	51	0
conferences	0	12
total	95791	306438

단어 "million" 스팸 : 156/95791 = 1/614

햄 : 98/306438 = 1/3127

가능도(likelihood ratio) = (1/614)/(1/3127) = 5.1

연습문제

- 햄에 대한 스팸의 사전확률 (spam : ham) = 1:1

word	likelihood ratio
million	5.1

햄에 대한 스팸의 사후확률은 ?

연습문제

- 햄에 대한 스팸의 사전확률 = 1:1

word	likelihood ratio
million	5.1

햄에 대한 스팸의 사후확률은 ?

사전확률이 1:1이므로, 사후확률 = 1 x 5.1 : 1

"million" 단어를 포함한 메시지의 경우,
약 83.6%로 스팸메시지가 나타남

베이즈 정리의 단점

- 베이즈 추론이 정확한 결과를 도출하려면 사전 확률과 같은 확률값이 입력되어야 함

- 어떤 분야에서는 믿을만한 통계 자료가 없어서 사전 확률을 산정할 수 없는 경우도 있음

불확실성의 정량화 (1/2)

- AI 시스템은 블랙박스 특성으로 인해 불확실성을 내포
  - ➢ 신뢰할 수 있는 AI를 위해서는 불확실성을 해결해야 함

- 신뢰할 수 있는 AI의 5대 주요 구성요소 (2022년)
  - 안전(Safety)
  - 투명성(Transparency)
  - 설명가능성(Explainability)
  - 견고성(Robustness)
  - 공정(Fairness)

불확실성의 정량화 (2/2)

- 불확실성 정량화 기술
  - AI 시스템의 활용 시나리오에서 발생 가능한 상황과 그에 대한 객관적 확률을 측정
  - AI 학습에 사용되는 데이터에 내재된 불확실성의 측정 또는 학습된 모델이 확신하지 못하는 영역에 대한 규명 등의 연구들이 이어지고 있음
- 설명가능 AI(XAI, eXplainable AI)를 위해서는 필수적
  - XAI는 머신러닝/딥러닝 모델의 결과값에 대한 이유를 인간이 이해할 수 있도록 블랙박스 성향을 분해・파악하여 설명 가능성을 제공하는 방식 => 신뢰성 향상, 편향 감소