

REPORT

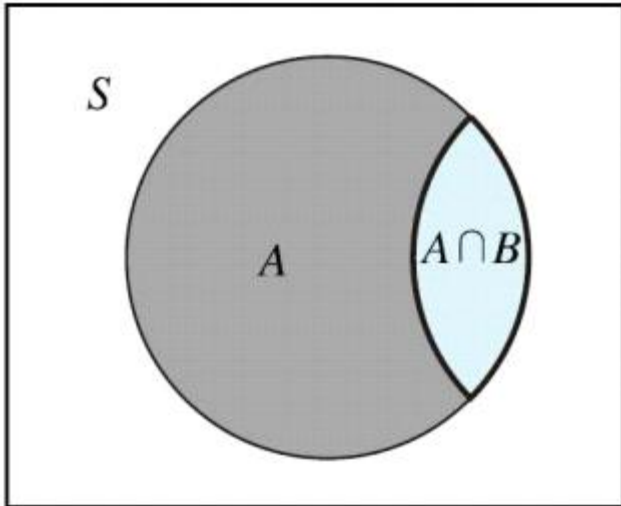
환자의 병명 판정하기



과목명	확률론
담당교수	유 훈 교수님
학과	융합전자공학과
학년	1학년
학번	201910906
이름	이학민
제출일	2019.12.03

1. 조건부 확률(Conditional Probability)

조건부 확률이란 0보다 큰 확률을 가진 사건 A가 이미 발생했을 때, 사건 B가 발생할 확률을 의미한다. $P(B|A)$ 로 나타내고 벤 다이어그램을 이용하면 다음 식을 도출할 수 있다.

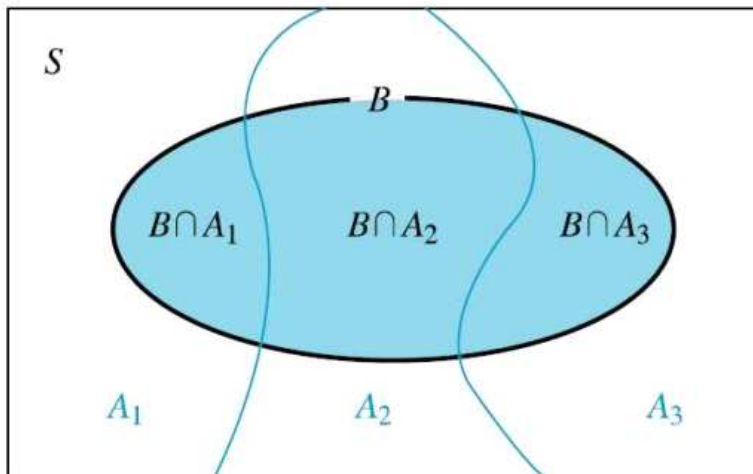


$$P(B|A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0$$

위에서 유도한 식을 정리하면 $P(A)P(B|A) = P(A \cap B)$ -----①임을 알 수 있다.

2. 베이즈 정리(Bayes Probability)

베이즈 정리에서 전확률 공식을 구하는 과정은 다음과 같다.



$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$ 임을 벤 다이어그램을 통해 알 수 있다.

이때, 정리한 식을 식①을 이용하여 다시 써보면

$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$ 이고

이는 $P(B) = \sum_{i=1}^3 P(A_i)P(B|A_i)$ 으로 나타낼 수 있다. 따라서 일반화된 전확률 공식은

$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$ -----②임을 알 수 있다.

문제 풀이에 앞서 각 기호를 한글로 풀어서 해석해보면 다음과 같다.

Prior $P(A_i)$: 사전확률(Prior Probability)로, 각 병명이 확정될 확률.

Likelihood $P(B|A_i)$: 각 병명이 확정되었을 때 심한 기침 or 심한 가래 or 38도 이상의 고열일 확률.

Posterior $P(A_i|B)$: 사후확률(Posterior Probability)로, 과제에서 최종적으로 구해야 할 값, 즉 심한 기침 or 심한 가래 or 38도 이상의 고열일 때, 각 병명으로 판단될 확률.

Posterior $P(A_i|B)$ 의 값을 구하기 위해 위의 식①과 식②를 사용하여 정리해보면 다음과 같다.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)} \text{ -----③}$$

식③을 이용하여 표를 작성하면 문제에서 요구하는 답을 찾을 수 있다.

[문제 1]

Event A_i	Condition	Prior $P(A_i)$	Likelihood $P(B A_i)$	$P(A_i)P(B A_i)$	Posterior $P(A_i B)$
A1	단순감기	0.50	0.15	0.08	0.24
A2	호흡기 독감	0.30	0.30	0.09	0.29
A3	폐렴	0.10	0.90	0.09	0.29
A4	알러지	0.05	0.20	0.01	0.03
A5	기관지염	0.05	0.95	0.05	0.15
B1	심한 기침	1.00	2.50	0.31	1.00

[문제 2]

Event A_i	Condition	Prior $P(A_i)$	Likelihood $P(B A_i)$	$P(A_i)P(B A_i)$	Posterior $P(A_i B)$
A1	단순감기	0.50	0.10	0.05	0.24
A2	호흡기 독감	0.30	0.15	0.05	0.22
A3	폐렴	0.10	0.80	0.08	0.39
A4	알러지	0.05	0.15	0.01	0.04
A5	기관지염	0.05	0.50	0.03	0.12
B2	심한 가래	1.00	1.70	0.21	1.00

[문제 3]

Event A_i	Condition	Prior $P(A_i)$	Likelihood $P(B A_i)$	$P(A_i)P(B A_i)$	Posterior $P(A_i B)$
A1	단순감기	0.500	0.050	0.025	0.097
A2	호흡기 독감	0.300	0.500	0.150	0.581
A3	폐렴	0.100	0.800	0.080	0.310
A4	알러지	0.050	0.050	0.003	0.010
A5	기관지염	0.050	0.010	0.001	0.002
B3	38도 이상 고열	1.00	1.41	0.26	1.00

의사는 앞서 구한 표를 바탕으로 각 증상에 따라 환자에게 병명을 진단해 줄 수 있다. 확률로만 따져보았을 때 각 병명이 환자에게 해당되는 경우를 높은 순서대로 나열해보면 다음과 같다.

i) 심한 기침을 하는 환자

호흡기 독감, 폐렴(0.29) - 단순감기(0.24) - 기관지염(0.15) - 알러지(0.03)

ii) 심한 가래가 있는 환자

폐렴(0.39) - 단순감기(0.24) - 호흡기 독감(0.22) - 기관지염(0.12) - 알러지(0.04)

iii) 38도 이상의 고열이 나는 환자

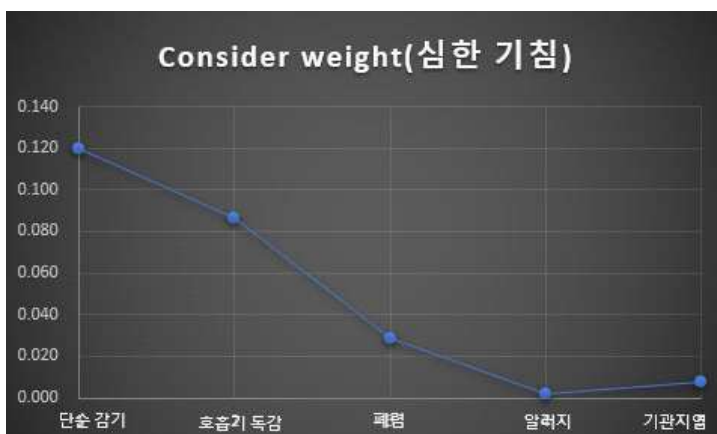
호흡기 독감(0.581) - 폐렴(0.310) - 단순감기(0.097) - 알러지(0.010) - 기관지염(0.002)

단순하게 확률로만 보면 값이 가장 큰 병명을 진단해주는 것이 최선의 선택 같아보이지만 각 증상에 따른 가중치를 부여하게 되면 진단해줄 병명이 달라지게 된다. 사후확률의 가중치는 사전확률, 즉 $P(A_i)$ 의 값과 같고 가중치를 부여하여 최적의 진단을 다시 판단해보면 다음과 같다.

i) 심한 기침을 하는 경우

Event A_i	Condition	Prior $P(A_i)$	Likelihood $P(B A_i)$	$P(A_i)P(B A_i)$	Posterior $P(A_i B)$	Consider weight
A1	단순감기	0.500	0.150	0.075	0.240	0.120
A2	호흡기 독감	0.300	0.300	0.090	0.288	0.086
A3	폐렴	0.100	0.900	0.090	0.288	0.029
A4	알러지	0.050	0.200	0.010	0.032	0.002
A5	기관지염	0.050	0.950	0.048	0.152	0.008
B1	심한 기침	1.000	2.500	0.313	1.000	

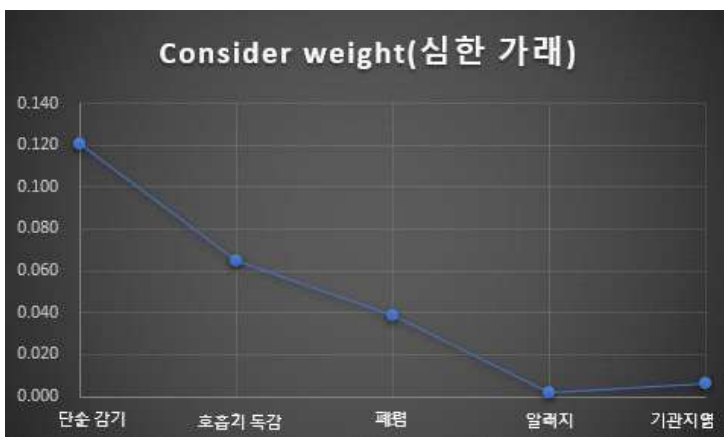
주황색으로 색칠되어있는 셀은 가중치를 적용한 값이다. 각 값은 $Prior P(A_i) \times Posterior P(A_i|B)$ 를 통하여 계산하였다. 이를 그래프로 알아보기 쉽게 나타내보면 다음과 같다.



그래프에서 알 수 있듯이 단순 감기가 0.120으로 가장 값이 크다. 따라서 심한 기침만 하는 환자의 경우 **단순 감기**가 최적의 진단임을 알 수 있다. (단순감기 - 호흡기 독감 - 폐렴 - 기관지염 - 알러지 순)

ii) 심한 가래가 있는 경우

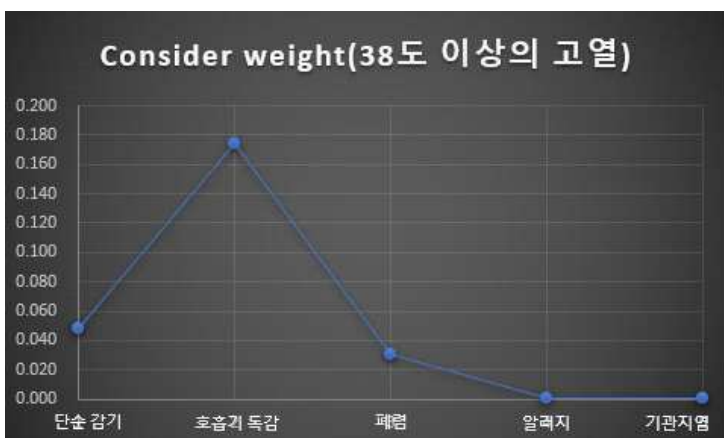
Event A_i	Condition	Prior $P(A_i)$	Likelihood $P(B A_i)$	$P(A_i)P(B A_i)$	Posterior $P(A_i B)$	Consider weight
A1	단순감기	0.500	0.100	0.050	0.241	0.120
A2	호흡기 독감	0.300	0.150	0.045	0.217	0.065
A3	폐렴	0.100	0.800	0.080	0.386	0.039
A4	알러지	0.050	0.150	0.008	0.036	0.002
A5	기관지염	0.050	0.500	0.025	0.120	0.006
B2	심한 가래	1.000	1.700	0.208	1.000	



그래프에서 알 수 있듯이 심한 가래만 있는 환자의 경우 **단순감기**가 0.120으로 가장 높은 값을 가지므로 최적의 진단이라고 할 수 있다. (단순감기 - 호흡기 독감 - 폐렴 - 기관지염 - 알러지 순)

iii) 38도 이상의 고열이 나는 경우

Event A_i	Condition	Prior $P(A_i)$	Likelihood $P(B A_i)$	$P(A_i)P(B A_i)$	Posterior $P(A_i B)$	Consider weight
A1	단순감기	0.500	0.050	0.025	0.097	0.048
A2	호흡기 독감	0.300	0.500	0.150	0.581	0.174
A3	폐렴	0.100	0.800	0.080	0.310	0.031
A4	알러지	0.050	0.050	0.003	0.010	0.0005
A5	기관지염	0.050	0.010	0.001	0.002	0.0001
B3	38도 이상 고열	1.000	1.410	0.258	1.000	



마찬가지 방법으로 38도 이상의 고열만 나는 환자에게는 **호흡기 독감**이 0.174로 가장 높은 값이므로 최적의 진단임을 알 수 있다. (호흡기 독감 - 단순 감기 - 폐렴 - 알러지 - 기관지염 순)

그러나 실제로 의사가 진료하는 환자의 경우 한 가지의 증상만 보이는 경우는 드물다. 만약 심한 기침, 심한 가래, 38도 이상의 고열을 모두 보이는 환자에게 병명을 진단해야 할 경우 3가지 증상의 교집합에 해당하는 병명을 뽑아야 한다. 따라서 가중치가 적용된 각 증상에 따른 병명이 진단될 확률을 모두 곱하여 값을 비교한다. 엑셀로 계산하여 얻어진 값을 정리하면 다음과 같다.

Condition	Product of Consider weight
단순감기	0.000700476
호흡기 독감	0.000980443
폐렴	0.000034430
알러지	0.000000001
기관지염	0.000000004



위의 표와 그래프에서 볼 수 있듯이 호흡기 독감이 0.000980443($0.086 \times 0.065 \times 0.174$)로 가장 높은 값을 가진다. 따라서 심한 기침, 심한 가래, 38도 이상의 고열의 증상을 모두 보이는 환자에게 최적의 진단은 **호흡기 독감**임을 알 수 있다.