Artificial Intelligence

Bayes - for assisting search

Bayes rule : P(A|B) = P(B|A)*P(A)/P(B)

- Purpose:
  - To compute the probability of something being true, given new information
- P (A|B) = P(B|A) * P(A) / P(B)
- P(A|B)
  - "The posterior"
  - The probability of A, given B
  - The probability of A being true, knowing B is true
- P(B|A)
  - "The likelihood"
  - The probability of B being true, knowing A is true
- A - the hypothesis
  - "The prior"
  - The probability of A being true, regardless of everything else
- B - "the marginal likelihood"
  - The new information
  - The probability of B being true, regardless of everything else

Bayes rule : $P(A|B) = P(B|A)*P(A)/P(B)$

- Example 1 : (very) inaccurate test T for disease D
  - Goal: know how to read the notation
- There is a test (T) for a disease (D). The test is not always accurate.
- Odds of someone (randomly selected from the pop) who tested positive with T, to have D?
- Bayes can help: $P(D|T+) = P(T+|D) * P(D)/P(T+)$
- $P(D|T+)$ = the odds someone has D, knowing that he got a positive test (T+)
- $P(T+|D)$ = the odds of someone getting a positive test (T+), knowing that indeed has D
- $P(D)$ = the odds of someone having D, regardless of everything else (for example, regardless of having been tested, or not)
- $P(T+)$ - the odds of someone getting a T+, regardless of everything else (for example, regardless of being sick with D, or not)
- Imagine the following data being available, after a study:
  - 1000 people having D took the test T
  - There were 800 T+
  - $P(T+|D) = 800/1000 = 0.8$
- Imagine that the following data is known, from the overall population:
  - 50 out of every 10000 people have D ("natural occurrence")
  - $P(D) = 50/10000 = 5/1000 = 0.005$
- Further studies found P(T+) regardless of "everything" (including being sick with D)
  - $P(T+) = 100/1000 = 0.1$
- $P(D|T+) = 0.8 * 0.005/0.1 = 0.8 * 0.05 = 0.04$

Bayes rule, example 2

2 approaches : intuitive & P(A|B) = P(B|A)*P(A)/denom

- Example 2: Inaccurate test T for disease D, "intuitive approach" + "Bayesian rule approach"
  - Goal: to observe both an "intuitive" and a Bayes rule based approach
- D has a natural incidence of 1/1000 = 0.001
  - P(D) = 0.001
- There is a test (T) for a disease (D).
  - The test is not perfect
  - The test does NOT produce false negatives
    - Anyone, who has the disease, will test positive
    - P(T+|D) = 1 #everyone who has the disease will test positive
  - The test CAN produce false positives, with a known frequency of 5%.
    - Some, who do NOT have the disease, might test positive
    - P(T+|not D) = 0.05 #5% false positives
- *Odds of someone, randomly selected from the population, having tested positive, to have D?*
  - *Compute P(D|T+)*

P(D|T+) an "intuitive approach"

- What are the odds of someone, randomly selected from the population, having tested positive, to have D?
  - P(D|T+) ?
- Imagine a random selection of 1000 people from the population
  - How many do we expect to have the disease? Natural incidence is 1/1000, so 1 in 1000
    - P(D) = 0.001
  - How many do we expect to NOT have the disease? 999 in 1000
    - P(not D) = 1-P(D) = 1-0.001 = 0.999
  - How many false positives? P(T+|not D)
  - P(T+|not D) = 0.05
    - 0.05 * 999 = 49.95 ~ 50 people will get FALSE T+
  - How many positive tests in total, regardless of correct, or not? P(T+)
  - #(T+) = 1 (from the prior known natural occurrence) + 50 (wrongly estimated) = 51
  - Back to the original question P(D|T+) ?
    - Now we intuitively know P(D|T+) = 1/51 = ~ 0.0196

$P(D|T+)$ by the Bayes rule
$P(D|T+) = P(D)*P(T+|D) / denom$
$denom = P(D)*P(T+|D) + P(not\ D)*P(T+|not\ D)$

- Odds of someone, randomly selected from the population, having tested positive, to have D?
  - $P(D|T+)$?
- Imagine a random selection of 1000 people from the population
  - How many do we expect to have the disease? Natural incidence is 1/1000
    - $P(D) = 0.001$
  - How many do we expect to NOT have the disease?
    - $P(not\ D) = 1-P(D) = 1-0.001 = 0.999$
  - How many false positives?
    - $P(T+|not\ D) = 0.05$
  - Back to the original question $P(D|T+)$?
- $P(D|T+) = P(D)*P(T+|D) / denom$
- $denom = P(D)*P(T+|D) + P(not\ D)*P(T+|not\ D)$
- $P(D) = 0.001$
- $P(T+|D) = 1$
- $P(not\ D) = 1-P(D) = 1-0.001 = 0.999$
- $P(T+|not\ D) = 0.05$
- $denom = 0.001 * 1 + 0.999 * 0.05 = 0.001 + 0.999*0.05 = 0.05095$
- $P(D|T+) = 0.001 *1 / denom = 0.001 / 0.05095 \sim 0.0196$

P(D|T+) by the Bayes rule
P(D|T+) = P(D)*P(T+|D)/P(T+)

- Odds of someone, randomly selected from the population, having tested positive, to have D?
  - P(D|T+) ?
- P(D|T+) = P(T+|D)*P(D)/P(T+)
- Known data
  - P(D) = 0.001 ; P(T+|D) = 1 ; P(not D) = 1-P(D) = 1-0.001 = 0.999 ; P(T+|not D) = 0.05
- P(D|T+) = P(T+|D)*P(D)/P(T+) = 1*P(D)/P(T+) = P(D)/P(T+)
- P(D) is known ; P(T+) must be found
- By "marginalization":
  - P(T+) = P(T+ and D) + P(T+ and not D) # bad idea to formulate this way
  - P(T+) = P(D and T+) + P(not D and T+) # better to formulate this way bc it starts with the prior P(D)
- By the "multiplication rule":
  - P(a and b) = P(a) * P(b|a) # necessary for each parcel of P(T+)
  - P(a and b) = P(b and a) = P(b)*P(a|b)
- So, regarding the 2 parcels from the multiplication rule:
  - P(D and T+) = P(D)*P(T+|D) = 0.001 * 1 = 0.001
  - P(not D and T+) = P(not D) * P(T+|not D) = 0.999 * 0.05
- Meaning
  - P(D|T+) = P(D) / P(T+) = 0.001 / (0.001 + 0.999*0.05) = 0.001 / (0.001 + 0.04995) = 0.019627085

Bayes helping the search for "X"

- Search [Python] game where "thing" X is missing
- A finite set of N search areas/regions are considered as possible locations where X might be
  - R1 .. Rn
- Initial probabilities are assigned to each possible search area
  - P1 .. Pn
- Initial "Search Effectiveness Probabilities" (SEP) will depend on the initial search efforts
  - SEP1 .. SEPn
- A search effort for X @Rn can fail: X might be at the searched area Rn, but NOT be found, even with high SEP
- P(X@Rn) = Pn * (1-SEPn) / sumForAllRegionsN (Pn * (1-SEPn))

References

- https://courses.lumenlearning.com/waymakermath4libarts/chapter/bayes-theorem/