COMP 2211 Exploring Artificial Intelligence

Supplementary Notes: A Proof of Bayes' Theorem

Dr. Desmond Tsoi
Department of Computer Science & Engineering
The Hong Kong University of Science and Technology, Hong Kong SAR, China

Bayes' Theorem

- As mentioned in class, Bayes' theorem is defined as:

$$P(B|E) = \frac{P(B)P(E|B)}{P(E)} = \frac{P(B)P(E|B)}{P(E|B)P(B) + P(E|NOT\ B)P(NOT\ B)}$$

- In other words, Bayes' theorem gives us the conditional probability of B given that E has occurred as long as we know $P(E|B)$, $P(E|NOT\ B)$, and $P(B) = 1 - P(NOT\ B)$.

A Proof of Bayes' Theorem

- By the definition of conditional probability:

$$P(E|B) = \frac{P(E\ AND\ B)}{P(B)} \quad (1)$$

$$P(E\ AND\ B) = P(E|B)P(B) \quad (2)$$

- By the law of total probability:

$$P(E) = P((E\ AND\ B)\ OR\ (E\ AND\ (NOT\ B))) = P((E\ AND\ B) + (E\ AND\ (NOT\ B))) \quad (3)$$

- By substituting Equation (2) to Equation (3):

$$P(E) = P(E|B)P(B) + P(E|NOT\ B)P(NOT\ B) \quad (4)$$

where on Equation (4), that fact that $E\ AND\ B$ and $E\ AND\ (NOT\ B)$ are mutually exclusive events was used.

A Proof of Bayes' Theorem

- Since $E\ AND\ B = B\ AND\ E$, by Equation (1) and (2), we have:

$$P(B|E) = \frac{P(B\ AND\ E)}{P(E)} = \frac{P(E|B)P(B)}{P(E)} \quad (5)$$

- By merging Equation (4) to (5), we get the following:

$$P(B|E) = \frac{P(E|B)P(B)}{P(E|B)P(B) + P(E|NOT\ B)P(NOT\ B)} \quad (6)$$

Equation (6) is what we wanted to prove. :)