Lec. 8 01/23/2012

3.3. Bayes Formula

Allows us to compute $P(F|E)$ from $P(E|F)$.

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(E|F)\cdot P(F)}{P(E)} \quad \Rightarrow$$

Bayes Theorem

$$P(F|E) = \frac{P(E|F)P(F)}{P(E)} = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}$$

Ex. A certain disease affects 1% of population. ($P(P|D^c)$)
A test is false positive in 5% cases, false-negative ($P(P^c|D)$) in 10% cases.
A person is tested positive. What is the probability that the person is sick with this disease?

$D = \{\text{has disease}\}$, $P = \{\text{test positive}\}$, $P(D|P) = ?$

$$P(D|P) = \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|D^c)P(D^c)} = \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.05 \times 0.99} \approx \boxed{0.18}$$

So small! Why? (Almost all positive tests are false-positive — too few people, 1% — are actually sick)

Ex (Example 3k) A missing plane have gone down in one of the three regions with probabilities p_1, p_2, p_3. $(p_1+p_2+p_3=1)$.
A plane can be found upon a search in these regions with probabilities q_1, q_2, q_3.
A search in region 1 was unsuccessful.
What is the prob. that the plane *is* in region 1?

$R_i = \{\text{plane is in region } i\}$, $i = 1, 2, 3$
$U_1 = \{\text{search in region 1 unsuccessful}\}$.

$$P(R_1|U_1) = \frac{P(U_1|R_1)P(R_1)}{P(U_1)} = \frac{P(U_1|R_1)P(R_1)}{P(U_1|R_1)P(R_1) + P(U_1|R_2)P(R_2) + P(U_1|R_3)P(R_3)}$$

$$= \boxed{\frac{(1-q_1)p_1}{(1-q_1)p_1 + 1\cdot p_2 + 1\cdot p_3}}$$

Ex (Bayesian spam filtering - see Wikipedia)

Suppose an incoming e-mail message contains a certain word, e.g. "Transfer". What is the probability that the message is a spam?

$W = \{\text{contains word "TRANSFER"}\}$, $Sp = \{\text{Spam}\}$.

$$P(Sp|W) = \frac{P(W|Sp)\,P(Sp)}{P(W|Sp)\,P(Sp) + P(W^c|Sp)\,P(Sp^c)}.$$

- $P(W|Sp)$ and $P(W^c|Sp)$ can be computed from previous training:

  e.g. $P(W|Sp) = P(\text{containing "TRANSFER" if it is spam})$

  = fraction of previously received spam messages that contain the word "TRANSFER".

- $P(Sp) = P(\text{a given message is spam})$ is set by the program or user (the larger, the more agressive filtering).