CSE 312

Foundations of Computing II

Bayes' Theorem

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

$$Pr(B \cap A) = Pr(B)\,Pr(A|B)$$

$$= Pr(A)\,Pr(B|A)$$

Definition (Bayes' Theorem)

$$\Pr(F \mid E) = \frac{\Pr(F \cap E)}{\Pr(E)} = \frac{\Pr(E \cap F)}{\Pr(E)} = \frac{\Pr(E \mid F)\Pr(F)}{\Pr(E)}$$

This allows us to use $\Pr(E \mid F)$ to calculate $\Pr(F \mid E)$ which turns out to be useful in a variety of applications.
In particular, it allows us swap $\Pr(\text{model} \mid \text{data})$ for $\Pr(\text{data} \mid \text{model})$.
It also allows us to "update" our beliefs based on new evidence.

Example

- 60% of email is spam
- 90% of spam has a forged header
- 20% of non-spam has a forged header

What is Pr(spam | forged header)? What happens as we vary how much spam there is?

Let S be the event "the email is spam".
Let F be the event "there is a forged header".

$(0.9)\ (0.6)$

$$Pr(S|F) = \frac{Pr(S \cap F)}{Pr(F)} = \frac{Pr(F|S)\,Pr(S)}{Pr(F)}$$

$$Pr(F) = \underbrace{Pr(F|S)\,Pr(S)}_{(0.9)(0.6)} + \underbrace{Pr(F|\bar{S})\,Pr(\bar{S})}_{(0.2)(0.4)}$$

Example

- 60% of email is spam
- 90% of spam has a forged header
- 20% of non-spam has a forged header

What is Pr(spam | forged header)? What happens as we vary how much spam there is?

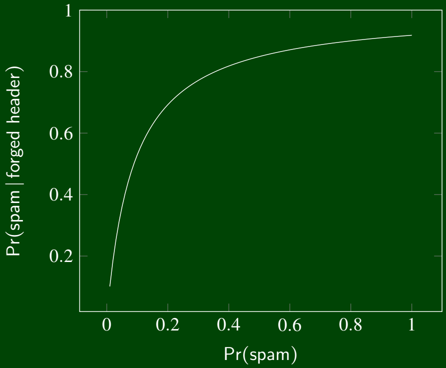
Let S be the event "the email is spam".
Let F be the event "there is a forged header".

$$\begin{aligned}
\Pr(S \mid F) &= \frac{\Pr(S \cap F)}{\Pr(F)} = \frac{\Pr(F \mid S)\Pr(S)}{\Pr(F)} = \frac{(0.9)(0.6)}{\Pr(F)} \\
&= \frac{(0.9)(0.6)}{\Pr(F \mid S)\Pr(S) + \Pr\left(F \mid \overline{S}\right)\Pr\left(\overline{S}\right)} \\
&= \frac{(0.9)(0.6)}{(0.9)(0.6) + (0.2)(0.4)} \\
&\approx 0.871
\end{aligned}$$

Example

- 60% of email is spam
- 90% of spam has a forged header
- 20% of non-spam has a forged header

What is $\Pr(\text{spam} \mid \text{forged header})$? What happens as we vary how much spam there is?



Spam Example 2

Example

- 60% of email is spam
- 10% of spam has the word "Viagra"
- 1% of non-spam has the word "Viagra"

What is $\Pr(\text{spam} \mid \text{viagra})$?

Let V be the event "message contains the word 'Viagra' "
Let S be the event "message is spam"

Example

- 60% of email is spam
- 10% of spam has the word "Viagra"
- 1% of non-spam has the word "Viagra"

What is $\Pr(\text{spam} \mid \text{viagra})$?

Let V be the event "message contains the word 'Viagra' "
Let S be the event "message is spam"

Spam Example 2

Example

- 60% of email is spam
- 10% of spam has the word "Viagra"
- 1% of non-spam has the word "Viagra"

What is $\Pr(\text{spam} \mid \text{viagra})$?

Let V be the event "message contains the word 'Viagra' "
Let S be the event "message is spam"

$$\begin{aligned}
\Pr(S \mid V) &= \frac{\Pr(V \mid S)\Pr(S)}{\Pr(V \mid S)\Pr(S) + \Pr\left(V \mid \overline{S}\right)\Pr\left(\overline{S}\right)} \\
&= \frac{(0.1)(0.6)}{(0.1)(0.6) + (0.01)(0.4)} \\
&\approx 0.9375
\end{aligned}$$

Example

- Child is born with (A, a) gene pair
- Mother has (A, A) gene pair
- Two possible fathers: $M_1 = (a, a)$ and $M_2 = (a, A)$

What is $\Pr(M_1 \text{ is the father} \mid \text{child born with } (A, a)))$?

Let Aa be the event "child born with (A, a)".
Let F_1 be the event "M_1 is the father".
Let F_2 be the event "M_2 is the father".
Say $\Pr(F_1) = p$ and $\Pr(F_2) = 1 - p$

Example

- Child is born with (A,a) gene pair
- Mother has (A,A) gene pair
- Two possible fathers: $M_1 = (a,a)$ and $M_2 = (a,A)$

What is $\Pr(M_1 \text{ is the father} \mid \text{child born with } (A,a))$?

Let Aa be the event "child born with (A,a)".
Let F_1 be the event "M_1 is the father".
Let F_2 be the event "M_2 is the father".
Say $\Pr(F_1) = p$ and $\Pr(F_2) = 1-p$

$$
\begin{aligned}
\Pr(F_1 \mid Aa) &= \frac{\Pr(Aa \mid F_1)\Pr(F_1)}{\Pr(Aa \mid F_1)\Pr(F_1) + \Pr\left(Aa \mid \overline{F_1}\right)\Pr\left(\overline{F_1}\right)} \\
&= \frac{(1)(p)}{(1)(p) + (0.5)(1-p)} \\
&= \frac{2p}{1+p}
\end{aligned}
$$

Example

- An HIV test's "false negative" rate (negative when present) is 2% and its "false positive" (positive when not present) rate is 1%
- 0.5% of the population has HIV

Let P be the event "X tests positive for HIV".
Let H be the event "X has HIV".
What is $\Pr(H \mid P)$?

Example

- An HIV test's "false negative" rate (negative when present) is 2% and its "false positive" (positive when not present) rate is 1%
- 0.5% of the population has HIV

Let P be the event "X tests positive for HIV".
Let H be the event "X has HIV".
What is $\Pr(H \mid P)$?

$$\begin{aligned}
\Pr(H \mid P) &= \frac{\Pr(P \mid H)\Pr(H)}{\Pr(P \mid H)\Pr(H) + \Pr\left(P \mid \overline{H}\right)\Pr\left(\overline{H}\right)} \\
&= \frac{(0.98)(0.005)}{(0.98)(0.005) + (0.01)(0.995)} \\
&\approx 0.330
\end{aligned}$$

Example

- An HIV test's "false negative" rate (negative when present) is 2% and its "false positive" (positive when not present) rate is 1%
- 0.5% of the population has HIV

Let P be the event "X tests positive for HIV".
Let H be the event "X has HIV".
What is $\Pr\left(H \mid \overline{P}\right)$?

Example

- An HIV test's "false negative" rate (negative when present) is 2% and its "false positive" (positive when not present) rate is 1%
- 0.5% of the population has HIV

Let P be the event "X tests positive for HIV".
Let H be the event "X has HIV".
What is $\Pr\left(H \mid \overline{P}\right)$?

$$\begin{aligned}
\Pr\left(H \mid \overline{P}\right) &= \frac{\Pr\left(\overline{P} \mid H\right)\Pr(H)}{\Pr\left(\overline{P} \mid H\right)\Pr(H) + \Pr\left(\overline{P} \mid \overline{H}\right)\Pr\left(\overline{H}\right)} \\
&= \frac{(0.02)(0.005)}{(0.02)(0.005) + (0.99)(0.995)} \\
&\approx 0.0001
\end{aligned}$$