CS109 - lecture 4

Let's derive the 3rd ~~Axioms~~ corollary from last class

$$P(E^c) = 1 - P(E)$$

complement

$$P(E \cup E^c) = P(E) + P(E^c)$$

Is due to $\boxed{\text{axiom 3}}$: $P(E \cup F) = P(E) + P(F)$

BTW:

$\boxed{\text{Axiom 1:}}$ $0 \leq P(E) \leq 1$

$\boxed{\text{Axiom 2:}}$ $P(S) = 1$

$$P(E \cup E^c) = P(S) = 1$$

$$1 = P(E) + P(E^c)$$

$$\text{so} \quad P(E) = 1 - P(E^c)$$

Serendipity

- Population of Stanford is 17000 pple
- You are friends with X people
- You walk into a room and see 268 pple
- What is the probability of seeing someone you know?

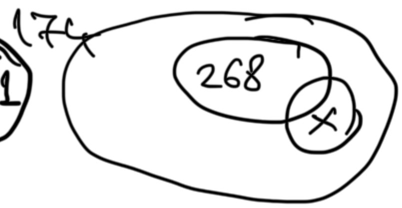
know: 1, 2, 3 . . .

P(seeing someone you) = P(E) Prob(know 1) + … + Prob(know

know in a room) = ———— = ————

$P(S)$ $\binom{17000}{268}$

Attempt 1



$\binom{X}{1}\binom{17k-1}{268-1} \rightarrow$ But just for one!!!

ProTip: Sometimes it is easy to calculate the probability of the Event <u>not happen</u>

$$\text{Sample Space} = \binom{17000}{268}$$

$$\text{Event Space for not happen (do not know anyone)} = \binom{17000-X}{268}$$

Remove the People I know

$X = 100$

$$\text{Prob not seeing someone} = \frac{\binom{17000-100}{268}}{\binom{17000}{268}} = 0{,}2$$

$$\text{Prob of seeing someone} = 1 - 0{,}2$$

↓

Can be see 1 or 2 or 3 or ...

Much easier to calculate

Goal for the Lesson: $\boxed{\text{Conditional Probability}}$

$P(E \text{ and } F)$

$\uparrow \downarrow$ Definition of Conditional Probability

$P(E|F)$

Law of total Probability

Bayes' theorem

$P(E)$ $\qquad$ $P(F|E)$

Conditional Probability

Roll 2 dice (6 sided) so that the sum is 4

- What is the best outcome for $P(D_1)$?

Choices:

A: 1 and 3 tie $\qquad 1 \to 3 \quad 3 \to 1$

B: 1, 2, 3 tie $\qquad 1 \to 3 \quad 2 \to 2 \quad 3 \to 1$

C: 2 is best $\qquad 2 \to 2$

D : other

$|S| = [\,(1,1)\ (1,2)\ (1,3) \cdots (1,6)$
$(2,1)\ (2,2)$
$(3,1)$
$(6,1) \cdots (6,6)\,]$

$= 36$

$E = \{(1,3), (2,2), (3,1)\} \quad |E| = 3$

$P(E) = \frac{3}{36} = \frac{1}{12}$

What is P(E, given F already observed)? $F=2$

Now $S = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)$

$E' = \{(2,2)\}$ So $P(E') = \frac{1}{6}$

So $P(E) \longrightarrow P(E')$
from: to:

the probability increased!

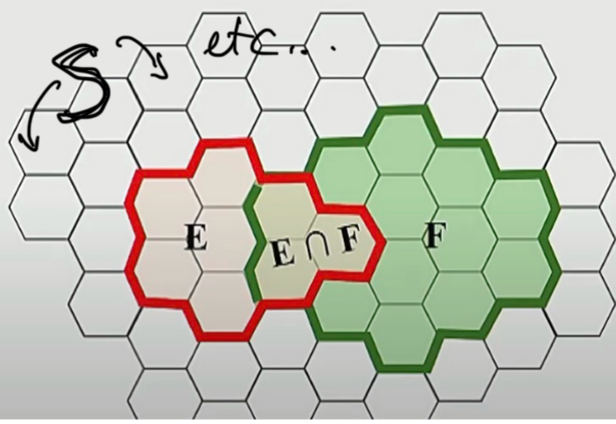
$\frac{1}{12} \rightarrow \frac{1}{6}$

Increased because the Sample

Space decreased more than the Event Space

Conditional Probability, visual intuition

The conditional probability of E given F is the probability that E occurs given that F has already occurred. This is known as conditioning on F.



$$P(E) = \frac{8}{50} \approx 0.16$$

$$P(E|F) = \frac{3}{14} \approx 0.21$$

So with equally likely outcomes:

$$P(E|F) = \frac{\#\text{ of outcomes in } E \text{ consistent with } F}{\#\text{ of outcomes in } S \text{ consistent with } F}$$

Implicit AND

$$= \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

$$P(E|F) = \frac{P(EF)}{P(F)}$$

General Definition of Conditional Probability

$$P(EF) = P(E|F) \cdot P(F)$$

Chain Rule of

If $P(F) = 0$ then

Cond. Prob

$P(E|F)$ = undefined (observed the impossible)

Netflix $\longrightarrow$ "Life is beautiful"

What is the probability that a user will watch "Life is beautiful" $P(E)$

$S = \{\text{watch}, \text{not watch}\}$

$E = \{\text{watch}\}$ $\quad P(E) \neq 1/2$

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\#\text{ people who watched it}}{\#\text{ people on Netflix}}$$

So: $P(E) = \dfrac{10\,234\,231}{50\,923\,123} = 0{,}20$

Now the conditional prob:

"What is the probability that a user watches "Life is beautiful" given they watched "CODA"?

$P(E|F) \qquad P(\overbrace{EF}^{E \text{ AND } F})$

$$P(E|F) = \frac{P(EF)}{P(F)}$$

$$= \frac{\dfrac{\text{\# Users that watched both}}{\text{\# Total users}}}{\dfrac{\text{\# users that watched Coda}}{\text{\# Total users}}}$$

$$= \frac{\text{\# Users that watched both (outcomes)}}{\text{\# Users that watched Coda (outcomes)}}$$

$$\approx 0.42$$

Why all this?

Because Machine Learning is:

Probability + Data + Computers

Notation

And	Or	Given
$P(E \text{ and } F)$	$P(E \text{ or } F)$	$P(E\|F)$
$P(E,F)$	$P(E \cup F)$	$P(E\|F,G)$
$P(EF)$		

$P(E \cap F)$

Different notations are used across the world.

Chain Rule via baby poop

In the morning when she wakes up, a baby has 50% chance of having pooped. The chance that a baby cries if she has pooped is 50%.

· What is the probability that a baby has pooped and cries?

$$P(E \cap F) = P(EF) = \ ?$$

$$P(E|F) = \frac{P(EF)}{P(F)}$$ → General definition of Conditional Probability

$$P(EF) = P(E|F) \cdot P(F)$$ → Chain rule ''

P(cries and pooped) = P(cries given it pooped) · P(it pooped)

$$= 0{,}5 \cdot 0{,}5 = 0{,}25$$

In conditional probabilities the order matters

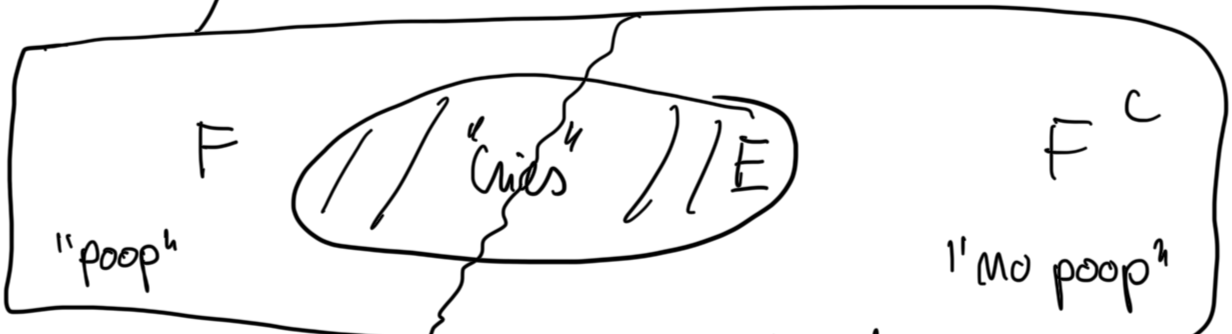
So $P(E|F) \neq P(F|E)$

Generalized chain Rule

$$\Pr(E_1 \text{ and } E_2 \text{ and } \dots E_m)$$
$$= \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1, E_2) \cdots \Pr(E_m | E_1 \cdot E_2 \cdot \dots \cdot E_{m-1})$$

LAW OF TOTAL PROBABILITY

Say E and F are events in S



mutually exclusive

$$P(E) = P(E \cap F) + P(E \cap F^c)$$
$$= P(E|F) \cdot P(F) + P(E|F^c) \cdot P(F^c)$$

We were missing the probability of crying given "no poop"

If we say it is e.g. $1/8$

$P(E) = 0{,}25 + 1/8 \cdot 0.5 = 0{,}\ldots$

Evolution of Bacteria

$P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)$ Law of Total Probability

You have bacteria in your gut which is causing a disease.
10% have a mutation which makes them resistant to anti-biotics
You take half a course of anti-biotics...

Probability a bacteria survives given it has the mutation: 20%
Probability a bacteria survives given it doesn't have the mutation: 1%
What is the probability that a randomly chosen bacteria survives?

$P(F) = 10\%$

F = mutation

E = Survives

$P(E|F) = 20\%$

$P(E|F^c) = 1\%$

BTW

$P(E^c|F) = 0.8$

$P(E^c|F^c) = 0.99$

are also known

$$P(E) = P(E|F)\cdot P(F) + P(E|F^c)\cdot P(F^c)$$
$$= 0.2 \cdot 0.1 + 0.01 \cdot 0.9$$
$$= 0.02 + 0.009$$
$$= 0.029 = 2{,}9\%$$

"<u>The Idea that wouldn't die</u>" – Bayes' Theorem

Allows us to get from $P(E|F)$ to $P(F|E)$!

Theorem: For events E and F, where $P(E)>0$, $P(F)>0$

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

Apply law of total Prob!

Expanded form:

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

We just applied the law of total probs.

Example of Application

Detecting SPAM email:

- 60% of all email in 2016 is spam
- 20% of spam has the word "Dear"
- 1% of non-spam has the word "Dear"
- You get an email with "Dear". What is the probability that the email is spam

$P(S) = 0.6$ $\quad P(S^c) = 0.4$

$P(D|S) = 0.2$ $\quad P(D|S^c) = 0.01$

$$P(S|D) = ?$$

$$P(S|D) = \frac{\overset{0.2}{P(D|S)} \cdot \overset{0.6}{P(S)}}{P(D)}$$

$$P(S|D) = \frac{P(D|S) \cdot P(S)}{P(D|S) \cdot P(S) + P(D|S^c) \cdot P(S^c)}$$

$$= \frac{0.2 \cdot 0.6}{0.2 \cdot 0.6 + 0.01 \cdot 0.4} = \frac{0.12}{0.12 + 0.004}$$

$$= \frac{0.12}{0.124}$$

$$= 0.96$$

$$= 96\%$$

There is some terminology around Bayes th.

$$\underset{\text{(Posterior Belief)}}{P(F|E)} = \frac{\underset{\text{(likelihood)}}{P(E|F)}\ \underset{\text{(Prior belief)}}{P(F)}}{\underset{\text{(normalization constant)}}{P(E)}}$$

So we basically update a belief

Based on a prior belief and the likelihood of new information.

—————— // ——————

SARS Virus Testing

- A test is 98% effective at detecting SARS
- False positive is 1% $= P(\text{Tested pos.} \mid \text{NO SARS}) = P(E \mid F^c)$
- 0.5% of pop. has SARS $\rightarrow P(F) = 0.5\%$
- $P(F \mid E) = ?$

$\hookrightarrow P(\text{you have SARS} \mid \text{you tested positive})$

$$P(F \mid E) = \frac{P(E \mid F) \cdot P(F)}{P(E)} = 0.330 \approx 33\%$$

$$P(E \mid F) = P(\text{you tested positive} \mid \text{you have SARS})$$

$P(E \mid F) = 0.98$ $\quad$ $P(F) = 0.005$

$$P(E) = P(E \mid F) \cdot P(F) + P(E \mid F^C) \cdot P(F^c)$$

$$= 0.98 \cdot 0{,}005 + 0{,}01 \cdot (1 - 0{,}005)$$

$$P(F \mid E) = \frac{0{,}98 \cdot 0{,}005}{0{,}98 \cdot 0{,}005 + 0{,}01 \cdot 0{,}995}$$

$$= \frac{0{,}0049}{0{,}0049 + 0{,}00995}$$

$$= \frac{0{,}0049}{0{,}01485} = 0{,}329$$