Machine Learning

Section 2: Plausible Reasoning and Bayes Rule

Stefan Harmeling

11. October 2021

Life's most important questions are, for the most part, nothing but probability problems.

Pierre-Simon Laplace

Sophie Feytaud (fl.1841), Pierre-Simon Laplace, public domain

Probability theory as an extension of logic

Deductive and plausible reasoning (1)

Jaynes, 2003, Sec. 1

A = it will start to rain by 6pm

B = the sky will become cloudy before 6pm

Deductive reasoning:

if A is true, then B is true
A is true

B is true

"modus ponens"

if A is true, then B is true
B is false

A is false

"modus tollens"

Deductive and plausible reasoning (2)

Jaynes, 2003, Sec. 1

A = it will start to rain by 6pm

B = the sky will become cloudy before 6pm

Plausible reasoning:

$$\frac{\text{if } A \text{ is true, then } B \text{ is true} \qquad B \text{ is true}}{A \text{ becomes more plausible}}$$

$$\frac{\text{if } A \text{ is true, then } B \text{ is true} \qquad A \text{ is false}}{B \text{ becomes less plausible}}$$

Deductive and plausible reasoning (3)

Jaynes, 2003, Sec. 1

C. Löser, Otterndorf Regenwolke April-2017 DSC 1499, CC BY 3.0 DE

A = it will start to rain by 6pm

B = the sky will become cloudy before 6pm

Overview:

- Assume: if A is true, then B is true.
  - A is true, implies that B is true.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A is false.

Deductive and plausible reasoning (4)

Jaynes, 2003, Sec. 1

A = it will start to rain by 6pm

B = the sky will become cloudy before 6pm

Plausible reasoning:

$$\frac{\text{if } A \text{ is true, then } B \text{ more plausible} \quad B \text{ is true}}{A \text{ becomes more plausible}}$$

$$\frac{\text{if } A \text{ is true, then } B \text{ more plausible} \quad A \text{ is false}}{B \text{ becomes less plausible}}$$

Deductive and plausible reasoning (5)

Jaynes, 2003, Sec. 1

C. Löser, Otterndorf Regenwolke April-2017 DSC 1499, CC BY 3.0 DE

A = it will start to rain by 6pm

B = the sky will become cloudy before 6pm

Plausible reasoning:

if A is true, then B more plausible
A is true

B becomes more plausible

if A is true, then B more plausible
B is false

A becomes less plausible

Deductive and plausible reasoning (6)

Jaynes, 2003, Sec. 1

Overview:

- Assume: if A is true, then B is true.
  - A is true, implies that B is true.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A is false.
- Assume: if A is true, then B becomes more plausible.
  - A is true, implies that B becomes more plausible.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A becomes less plausible.

How can we formalize plausibility?

Formalizing plausibility (1) — Cox's axioms

Jaynes, 2003, Sec. 1-2

Cox's axioms (formalizing common sense)

- plausibility of B assuming that A is true is a real number $p(B|A)$
- plausibility $p(B|A)$ complies with common sense
- plausibility $p(B|A)$ is consistent

Cox's theorem (WARNING: proof is not completely rigorous)

- product rule: $p(A, B|C) = p(A|B, C)p(B|C) = p(B|A, C)p(A|C)$
- sum rule: $p(A|C) + p(\neg A|C) = 1$

Notes

- the product rule implies Bayes' rule

Formalizing plausibility (2) — Kolmogorov's axioms

Pearl, 1988; `http://en.wikipedia.org/wiki/Probability_axioms`

Kolmogorov's axioms (plausibility as a measure)

- $0 \leq p(A) \leq 1$
- $p(\Omega) = 1$
- $p(A \text{ or } B) = p(A) + p(B)$ if A and B are mutually exclusive

Theorem

- $p(A) = p(A, B) + p(A, \neg B)$
- $p(A) + p(\neg A) = 1$

Definition of conditional probability

- for $p(A) > 0$ define $p(B|A) = p(A, B)/p(A)$

Theorem

- Bayes' rule: $p(B|A) = p(A|B)p(B)/p(A)$

Formalizing plausibility (3) — Cox vs. Kolmogorov

Richard Threlkeld Cox

Jack Engeman, ‘‘Richard T. Cox’’, public domain

- plausibilities should comply with common sense
- defines $p(B|A)$
- Bayes’ rule is derived, from basic desiderata, thus very well justified

Andrey Nikolajevich Kolmogorov

Konrad Jacobs, ‘‘Kolmogorov’’, cropped, CC BY-SA 2.0 DE

- plausibilities should be a measure on sample space
- defines $p(A)$, defines $p(B|A)$
- Bayes’ rule is “defined”, i.e. follows directly from definition of $p(B|A)$
- Kolmogorov’s approach is more rigorous

Note:

- Debate: is Bayes’ rule an axiom?
- Plausibilities are just probabilities.

The theory of probability is basically just common sense reduced to calculus (1814).
Pierre-Simon Laplace

Sophie Feytaud (fl.1841), Pierre-Simon Laplace, public domain

Bayes’ rule (1) — inverting probabilities

Unknown author, Thomas Bayes, public domain

Thomas Bayes, Bayes’ signature, public domain

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

Bayes’ rule inverts probabilistic relationships, it translates between $p(B|A)$ and $p(A|B)$.
adapted from Barber, 1.1, 2012

Bayes' rule (2) — updating beliefs

Plug-in data and some hypothesis:

$$p(H|\text{data}) = \frac{p(\text{data}|H)p(H)}{p(\text{data})}$$

Before you see the data: (right-hand-side of Bayes' rule)

- value your beliefs as probabilities
- without seeing data your prior belief in the hypothesis H is $p(H)$
- knowing that H is true, you believe that the data has a certain likelihood $p(\text{data}|H)$; knowing that it is wrong, its likelihood is $p(\text{data}|\neg H)$
- the evidence $p(\text{data})$ can be calculated from known quantities using the sum rule, $p(\text{data}) = p(\text{data}|H)p(H) + p(\text{data}|\neg H)p(\neg H)$

After you have seen the data: (left-hand-side of Bayes' rule)

- Bayes' rule calculates your posterior belief $p(H|\text{data})$ about the hypothesis after seeing the data

Bayes' rule tells you how to update your beliefs!

Bayes' rule (3) — monster vs. mouse

While you tried to sleep, you hear some noise under the bed...

$$
\begin{aligned}
n &= \text{some noise under your bed} \\
M &= \text{a monster under your bed} \\
m &= \text{a mouse under your bed} \\
e &= \text{something else (e.g. only air) under your bed}
\end{aligned}
$$

Question:

What is the type of n, M, m, e?

Answer:

Boolean, i.e. possible values are true/false.

Bayes' rule (3) — monster vs. mouse

While you tried to sleep, you hear some noise under the bed. . .

$$\begin{aligned} n &= \text{some noise under your bed} \\ M &= \text{a monster under your bed} \\ m &= \text{a mouse under your bed} \\ e &= \text{something else (e.g. only air) under your bed} \end{aligned}$$

Note:

The variables M, m and e are mutually exclusive,
i.e. only one of them can be true.

Use your world knowledge to assign probabilities:

- for $p(M)$, $p(m)$, $p(e)$, with $p(M) + p(m) + p(e) = 1$.
- for $p(n|M)$, $p(n|m)$, $p(n|e)$, with values between zero and one.

Bayes' rule (3) — monster vs. mouse

While you tried to sleep, you hear some noise under the bed...

$$\begin{aligned} n &= \text{some noise under your bed} \\ M &= \text{a monster under your bed} \\ m &= \text{a mouse under your bed} \\ e &= \text{something else (e.g. only air) under your bed} \end{aligned}$$

Joint probabilities:

$$\begin{aligned} p(n, M) &= p(n|M)p(M) && \text{noisy monster} \\ p(n, m) &= p(n|m)p(m) && \text{noisy mouse} \\ p(n, e) &= p(n|e)p(e) && \text{noisy something else} \end{aligned}$$

Given the noise, was it a monster?

$$p(M|n) = \frac{p(n|M)p(M)}{p(n)} = \frac{p(n, M)}{p(n, M) + p(n, m) + p(n, e)}$$

Bayes' rule (4) — monster vs. mouse

Given the noise, was it a monster, a mouse or something else?

$$p(M|n) = \frac{p(n|M)p(M)}{p(n)} = \frac{p(n|M)p(M)}{p(n,M)+p(n,m)+p(n,e)}$$

$$p(m|n) = \frac{p(n|m)p(m)}{p(n)} = \frac{p(n|m)p(m)}{p(n,M)+p(n,m)+p(n,e)}$$

$$p(e|n) = \frac{p(n|e)p(e)}{p(n)} = \frac{p(n|e)p(e)}{p(n,M)+p(n,m)+p(n,e)}$$

Bayes' rule in words:

The probability of some hypothesis H after seeing the data is the ratio between "how well does H describe the data" compared to all possible explanations.

Probabilities of all possible explanations sum to one:

$$p(M)+p(m)+p(e)=1 \qquad \text{before hearing the noise}$$

$$p(M|n)+p(m|n)+p(e|n)=1 \qquad \text{after hearing the noise}$$

Deductive and plausible reasoning — revisited (1)

Jaynes, 2003, Sec. 1

C. Löser, Otterndorf Regenwolke April-2017 DSC 1499, CC BY 3.0 DE

A = it will start to rain by 6pm

B = the sky will become cloudy before 6pm

Deductive reasoning:

if A is true, then B is true
A is true

B is true

“modus ponens”

if A is true, then B is true
B is false

A is false

“modus tollens”

Deductive and plausible reasoning — revisited (2)

Jaynes, 2003, Sec. 1

Overview:

- Assume: if A is true, then B is true.
  - A is true, implies that B is true.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A is false.
- Assume: if A is true, then B becomes more plausible.
  - A is true, implies that B becomes more plausible.
  - A is false, implies that B becomes less plausible.
  - B is true, implies that A becomes more plausible.
  - B is false, implies that A becomes less plausible.

Deductive and plausible reasoning — revisited (3)

Semanitics:

Plausibility is measured as probability.

The value $p(B|A)$ is the probability of B assuming that A is true.

The value $p(B)$ is the probability of B assuming nothing.

$p(A) = 1$ is the statement that A is true assuming nothing.

$p(A) > p(B)$ is the statement that A is more probable than B.

Basic laws of probability:

$p(A, B) = p(A\|B)p(B) = p(B\|A)p(A)$	product rule
$p(A, B\|C) = p(A\|B, C)p(B\|C) = p(B\|A, C)p(A\|C)$	product rule
$p(A) + p(\neg A) = 1$	sum rule
$p(A\|C) + p(\neg A\|C) = 1$	sum rule
$p(A, B) + p(A, \neg B) = p(A)$	sum rule

Deductive and plausible reasoning — revisited (4)

Deductive reasoning:

if A is true, then B is true
A is true

B is true

if A is true, then B is true
B is false

A is false

Probabilistic reasoning:

- assume $p(B|A) = 1$
- show $p(B|A) = 1$
- proof: by assumption

- assume $p(B|A) = 1$
- show $p(\neg A|\neg B) = 1$
- proof: apply Bayes' rule

Deductive and plausible reasoning — revisited (5)

Plausible reasoning:

$$\frac{\text{if } A \text{ is true, then } B \text{ is true} \qquad B \text{ is true}}{A \text{ becomes more plausible}}$$

$$\frac{\text{if } A \text{ is true, then } B \text{ is true} \qquad A \text{ is false}}{B \text{ becomes less plausible}}$$

Probabilistic reasoning:

- assume $p(B|A) = 1$
- show $p(A|B) \geq p(A)$
- proof: apply Bayes' rule

- assume $p(B|A) = 1$
- show $p(B|\neg A) \leq p(B)$
- proof: apply Bayes' rule

Deductive and plausible reasoning — revisited (6)

Plausible reasoning:

if A is true, then B more plausible
B is true

A becomes more plausible

if A is true, then B more plausible
A is false

B becomes less plausible

Probabilistic reasoning:

- assume $p(B|A) \geq p(B)$
- show $p(A|B) \geq p(A)$
- proof: apply Bayes' rule

- assume $p(B|A) \geq p(B)$
- show $p(B|\neg A) \leq p(B)$
- proof: apply Bayes' rule

Probabilistic reasoning — overview

Lemma 2.1

$p(B|A) = 1$ *implies*

- $p(B|A) = 1$ *“modus ponens”*
- $p(B|\neg A) \leq p(B)$
- $p(A|B) \geq p(A)$
- $p(\neg A|\neg B) = 1$ *“modus tollens”, alternative* $p(A|\neg B) = 0$

Lemma 2.2

$p(B|A) \geq p(B)$ *implies*

- $p(B|A) \geq p(B)$
- $p(B|\neg A) \leq p(B)$
- $p(A|B) \geq p(A)$
- $p(\neg A|\neg B) \geq p(\neg A)$

The rules of probability, combined with Bayes' rule make for a complete reasoning system, one which includes traditional deductive logic as a special case (see Jaynes, 2004).

adapted from Barber, 1.2, 2012

The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind (1850).

James Clerk Maxwell

George J. Stodart, Engraving of James Clerk Maxwell, public domain

Interpretations of probability

- http://plato.stanford.edu/archives/sum2012/entries/probability-interpret/
- http://en.wikipedia.org/wiki/Probability_interpretations

Approaches to reasoning under uncertainty in AI

Pearl, 1988, 1.1.3

Three schools (according to Pearl):

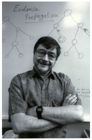
1. Logicists: non-numerical techniques, e.g. non-monotonic logic, circumscription

John McCarthy

2. Neo-calculists: numerical techniques, but not probabilities, instead new calculi, e.g. Dempster-Shafer calculus, fuzzy logic, certainty factors

Lotfi Zadeh

3. Neo-probabilists: probability theory together with clever computations



Judea Pearl

Summary

From Logic to Probability theory

- Deductive vs. plausible reasoning
- Cox's and Kolmogorov's approaches
- Rules of probability
- Bayes rule

Next section:

What exactly do we mean by $p(A)$ or $p(A, B)$, etc?