Machine Learning

Section 3: From Logic to Probabilities

Stefan Harmeling

13. October 2021

Today:

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

Let's define...

Syntax

- What are allowed strings?
- e.g. $A \wedge B$, A, $A \to B$, $A \vee \neg B \vee C$

Semantics

- What do these strings mean?
- i.e. when is $A \wedge B$ true, when is it false?

Propositional logic (1) — syntax

Definition 3.1 (alphabet)

The alphabet $\mathcal{A} = \mathcal{V} \cup \{\neg, \wedge, \vee, \rightarrow,), (\}$ *consists of various symbols:*

- *a finite set* $\mathcal{V}$ *of symbols* $X, Y, \ldots$; *aka non-logical symbols, propositional variables*
- *junctors* $\neg$, $\wedge$, $\vee$, $\rightarrow$; *aka logical symbols, connectives, names of truth functions*
- *parenthesis*), (

Definition 3.2 (formula)

A formula A is a finite-length string of symbols build along the following inductive definition:

- *all symbols in* $\mathcal{V}$ *are formulas*
- *if A and B are formulas then* $\neg A$, $A \wedge B$, $A \vee B$, $A \rightarrow B$ *and* (A) *are also formulas*

$\mathcal{F}$ *is the set of all formulas.* $\mathcal{F}$ *is a subset of all strings, i.e.* $\mathcal{F} \subset \mathcal{A}^*$.

In the theory of formal grammars, $\mathcal{F}$ is a context-free language, i.e. definition 2 is a context-free grammar (type II of Chomsky's hierarchy).

Example

Let's define formulas based on two propositional variables.

- $\mathcal{V} = \{X, Y\}$, the set of propositional variables
- $\mathcal{F} = \{X, Y, \neg X, \neg Y, X \vee Y, X \wedge Y, X \vee X, X \vee \neg X, \neg X \vee X, X \wedge Y \wedge Y, \ldots\}$
- Some strings are not formulas, e.g. $XX\wedge \notin \mathcal{F}$, or with quotation marks $"XX\wedge" \notin \mathcal{F}$

Propositional logic (2) — semantics

Definition 3.3 (boolean assignment)

A boolean assignment ω *assigns every propositional variable in* $\mathcal{V}$ *a truth value, i.e.*

$$\omega : \mathcal{V} \to \{0, 1\}$$

where 0 *represents false and* 1 *represents true.*

Definition 3.4 (entailment)

A boolean assignment ω *induces a truth function* $q : \mathcal{F} \to \{0, 1\}$ *that maps all formulas onto truth values as follows:*

- $q(X) := \omega(X)$ *for propositional variables* $X \in \mathcal{V}$
- $q(\neg A) := 1 - q(A)$ *and* $q(A \wedge B) := q(A)q(B)$ *and* $q((A)) = q(A)$
- $q(A \vee B) := q(\neg(\neg A \wedge \neg B))$ *and* $q(A \to B) := q(\neg A \vee B)$

If $q(A) = 1$ *we say that* ω *entails* A *and write* $\omega \models A$*.*

Example

Let's define formulas based on two propositional variables.

- $\mathcal{V} = \{X, Y\}$, the set of propositional variables
- $\mathcal{F} = \{X, Y, \neg X, \neg Y, X \vee Y, X \wedge Y, X \vee X, X \vee \neg X, \neg X \vee X, X \wedge Y \wedge Y, \ldots\}$
- example of boolean assignment: $\omega(X) = 1, \omega(Y) = 0$
- ω induces the following truth function q:

$$
\begin{aligned}
q(X) &= \omega(X) = 1 \\
q(Y) &= \omega(Y) = 0 \\
q(\neg X) &= 1 - q(X) = 0 \\
q(\neg Y) &= 1 - q(Y) = 1 \\
q(X \wedge Y) &= q(X)q(Y) = 0 \\
q(X \vee Y) &= q(\neg(\neg X \wedge \neg Y)) = 1 - (1 - q(X))(1 - q(Y)) = 1 \\
q(X \vee X) &= 1 - (1 - q(X))(1 - q(X)) = 1 \\
q(X \vee \neg X) &= 1 - (1 - q(X))(1 - (1 - q(X))) = 1 \\
q(\neg X \wedge X) &= (1 - q(X))q(X) = 0 \\
q(X \wedge Y \wedge Y) &= q(X)q(Y)q(Y) = 0 \ldots
\end{aligned}
$$

From propositional logic to probabilities (1)

Definition 3.5 (sample space)

The set Ω of all boolean assignments is called sample space. Note, that for n propositional variables it has 2^n elements.

Definition 3.6 (probability mass function)

The probability mass function $f : \Omega \to [0,1]$ assigns each boolean assignment a probability, such that

- $0 \leq f(\omega) \leq 1$ *for all* $\omega \in \Omega$
- $\sum_{\omega \in \Omega} f(\omega) = 1$

Definition 3.7 (event)

An event $E \subset \Omega$ is a subset of the sample space Ω. Each formula A naturally induces an event E_A:

$$E_A := \{\omega \in \Omega \text{ such that } \omega \models A\} \subset \Omega$$

Different formulas can induce the same event. Note that $\Omega = E_{A \vee \neg A}$.

Example

Let's define formulas based on two propositional variables.

- $\mathcal{V} = \{X, Y\}$, the set of propositional variables
- $\mathcal{F} = \{X, Y, \neg X, \neg Y, X \vee Y, X \wedge Y, X \vee X, X \vee \neg X, \neg X \vee X, X \wedge Y \wedge Y, \ldots\}$
- all boolean assignments $\omega_{00}, \omega_{01}, \omega_{10}, \omega_{11}$

$$\begin{array}{llll} \omega_{00}(X) = 0 & \omega_{01}(X) = 0 & \omega_{10}(X) = 1 & \omega_{11}(X) = 1 \\ \omega_{00}(Y) = 0 & \omega_{01}(Y) = 1 & \omega_{10}(Y) = 0 & \omega_{11}(Y) = 1 \end{array}$$

- a possible probability mass function:

$$f(\omega_{00}) = 0.1 \qquad f(\omega_{01}) = 0.4 \qquad f(\omega_{10}) = 0.3 \qquad f(\omega_{11}) = 0.2$$

- note: $f(\omega_{00}) + f(\omega_{01}) + f(\omega_{10}) + f(\omega_{11}) = 1$
- a possible event:

$$E = \{\omega_{01}, \omega_{10}\}$$

- the event that is induced by the formula $A = X \vee Y$

$$E_A = \{\omega_{01}, \omega_{10}, \omega_{11}\}$$

From propositional logic to probabilities (2)

Definition 3.8 (probability)

The probability $p(E)$ of some event E is the probability mass of E, i.e.

$$p(E) := \sum_{\omega \in E} f(\omega)$$

The probability $p(A)$ of some formula A is defined as the probability $p(E_A)$ of the induced event E_A, i.e.

$$p(A) := p(E_A) = \sum_{\omega \in E_A} f(\omega) = \sum_{\omega \models A} f(\omega)$$

Definition 3.9 (joint probability)

The joint probability $p(A, B)$ of several formulas A and B is the probability of their conjuction, i.e. $p(A, B) := p(A \wedge B)$, similarly for more than two. The joint probability of several events is the probability of their intersection (remember events are subsets of Ω).

Example

Let's define formulas based on two propositional variables.

- $\mathcal{V} = \{X, Y\}$, the set of propositional variables
- $\mathcal{F} = \{X, Y, \neg X, \neg Y, X \vee Y, X \wedge Y, X \vee X, X \vee \neg X, \neg X \vee X, X \wedge Y \wedge Y, \ldots\}$
- a possible probability mass function:

$$f(\omega_{00}) = 0.1 \qquad f(\omega_{01}) = 0.4 \qquad f(\omega_{10}) = 0.3 \qquad f(\omega_{11}) = 0.2$$

- a possible event:

$$E = \{\omega_{01}, \omega_{10}\}$$

- the event that is induced by the formula $A = X \vee Y$

$$E_A = \{\omega_{01}, \omega_{10}, \omega_{11}\}$$

- probabilities

$$p(E) = f(\omega_{01}) + f(\omega_{10}) = 0.7$$
$$p(E_A) = f(\omega_{01}) + f(\omega_{10}) + f(\omega_{11}) = 0.9$$

From propositional logic to probabilities (3)

Theorem 3.10 (Kolmogorov's axioms and more)

1. $0 \leq p(A) \leq 1$
2. $p(\Omega) = 1$
3. $p(A \vee B) = p(A) + p(B)$ *if* $E_A \cap E_B = \varnothing$
4. $p(A \vee B) = p(A) + p(B) - p(A, B)$
5. $p(A) = p(A, B) + p(A, \neg B)$
6. $p(A) + p(\neg A) = 1$
7. $p(A \vee B) = p(\neg(\neg A \wedge \neg B))$, $p(A \to B) = p(\neg A \vee B)$

The first three are Kolmogorov's axioms which imply 4., 5., 6.

Proof of 3:

$$p(A \vee B) = \sum_{\omega \in E_{A \vee B}} f(\omega) = \sum_{\omega \in E_A} f(\omega) + \sum_{\omega \in E_B} f(\omega) = p(A) + p(B)$$

The second equality holds because of $E_A \cap E_B = \varnothing$.

From propositional logic to probabilities (4)

Definition 3.11 (conditional probability)

For some formula or event B with non-zero probability, i.e. $p(B) > 0$, the conditional probability $p(A|B)$ is the ratio of the joint probability $p(A, B)$ and $p(B)$, i.e.

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

What about defining $p(A|B) = 1$ for $p(B) = 0$ (which looks like "ex falso quod libet")? This is a bad idea, since it would imply $p(A|B) + p(\neg A|B) = 2 \neq 1$.

Theorem 3.12 (Kolmogorov's axioms, Bayes' rule and more)

- *For fixed B, the conditional probability $p(A|B)$ fulfills Kolmogorov's axioms.*
- *Bayes' theorem: $p(B|A) = p(A|B)p(B)/p(A)$*

Note that Bayes' theorem is often falsely called "Bayes' rule". I also call it Bayes' rule ;).

From propositional logic to probabilities (5)

Lemma 3.13 (implication vs. conditional probability)

Are $p(A \to B)$ and $p(B|A)$ the same thing? Assume $p(A) > 0$, otherwise $p(B|A)$ is not defined.

1. $p(B|A) = \frac{p(A)-(1-p(A \to B))}{p(A)}$
2. $p(A) = 1$ *implies* $p(B|A) = p(A \to B)$
3. $p(A \to B) \geq p(B|A) \geq p(B, A)$
4. $p(A \to B) \geq 1 - p(A)$
5. $p(A \to B) = 1$ *if and only if* $p(B|A) = 1$

Proof:

$$p(B|A) = \frac{p(A, B)}{p(A)} = \frac{p(A) - p(A, \neg B)}{p(A)} = \frac{p(A) - (1 - p(\neg A \vee B))}{p(A)}$$
$$= \frac{p(A) - (1 - p(A \to B))}{p(A)}$$

More curious insights

Question: Does the comma or the bar has more binding power?

$$p(A, B|C) = p((A, B)|C) \qquad ???$$
$$p(A, B|C) = P(A, (B|C)) \qquad ???$$

Correct is $p(A, B|C) = p((A, B) \mid C) = p(A \wedge B \mid C)$, because is unclear how to combine A and $B|C$ to a single formula, however, read on...

Question: Can we condition on a condition?

$$p(A|B|C) = ???$$

Yes, while no one uses that notation, a possible interpretation (that makes sense) is:

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

We first resolved the outer conditioning on C, and then the inner one on B. So, one could say that the vertical bar is left-binding.

This is a distribution for A conditioned several times. All conditions can be combined into a list $p(A \mid B, C)$ or also a formula $p(A \mid B \wedge C)$.

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

Events as inputs

- events E are subsets of Ω (the set of all events)
- $p(E)$ is the probability that event E happens

Formulas as inputs

- a formula $A \vee B$ induces an event $E_{A \vee B}$
- $p(A \vee B)$ is the probability that formula $A \vee B$ is true (defined via the event sets)

We can use anything that is either true or false as input for p.

Summary

Probability notation

- p is a function of events. Also a function of formulas, since they induce events.
- $p(\cdot)$, plug in anything that is either true or false.

There are only two important rules:

$$
\begin{aligned}
p(A, B|C) &= p(A|B, C)\ p(B|C) && \text{product rule} \\
p(B|C) &= p(A, B|C) + p(\neg A, B|C) && \text{sum rule}
\end{aligned}
$$

. . . with some variations:

$$
\begin{aligned}
p(A, B) &= p(A|B)\ p(B) && \text{product rule} \\
p(B) &= p(A, B) + p(\neg A, B) && \text{sum rule} \\
1 &= p(A) + p(\neg A) \\
p(A, B) &= p(B|A)\ p(A) \\
p(B|A) &= p(A|B)\ p(B)/p(A) && \text{Bayes rule} \\
&\vdots
\end{aligned}
$$