A General Logical Approach to Learning from Time Series

Guido Sciavicco guido.sciavicco@unife.it

Applied Computational Logic and Artificial Intelligence Laboratory,
Department of Mathematics and Computer Science,
University of Ferrara, Italy

Today's menu

For the past 20 years or so, machine learning has been progressively more pervasive, and penetrated virtually every field of computer science. Reasoning with temporal data is no exception, and several methods and techniques have been developed for extracting temporal or temporal-like information from temporal data, often, if not always in the form of time series. In this talk we shall focus on

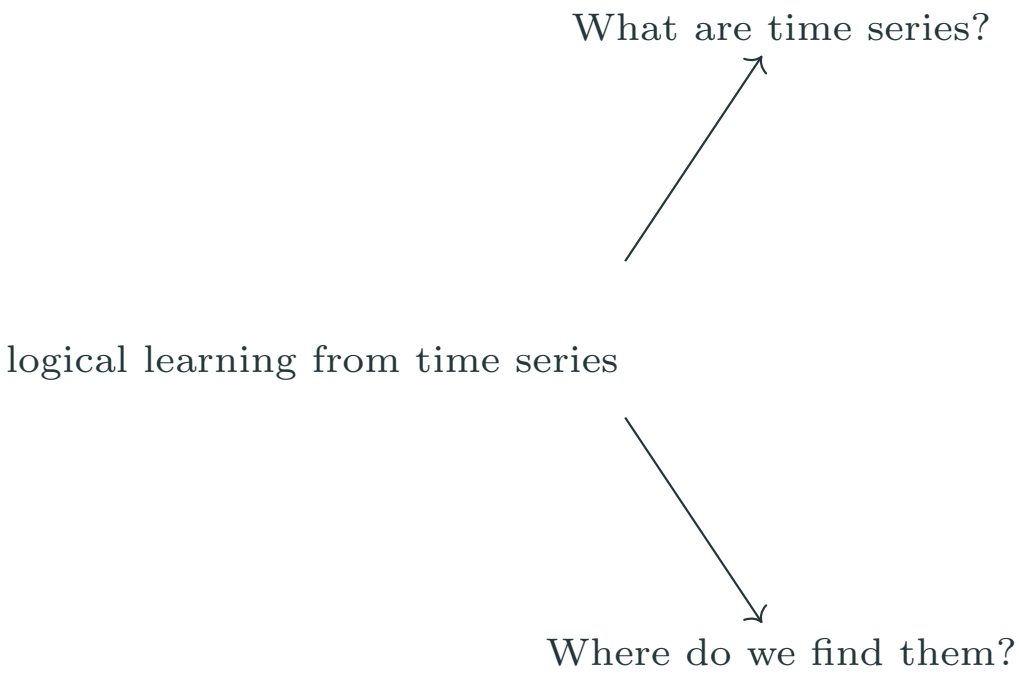
logical learning from time series

that is the sub-field of machine learning that focuses on extracting logical, symbolic information from temporal data.

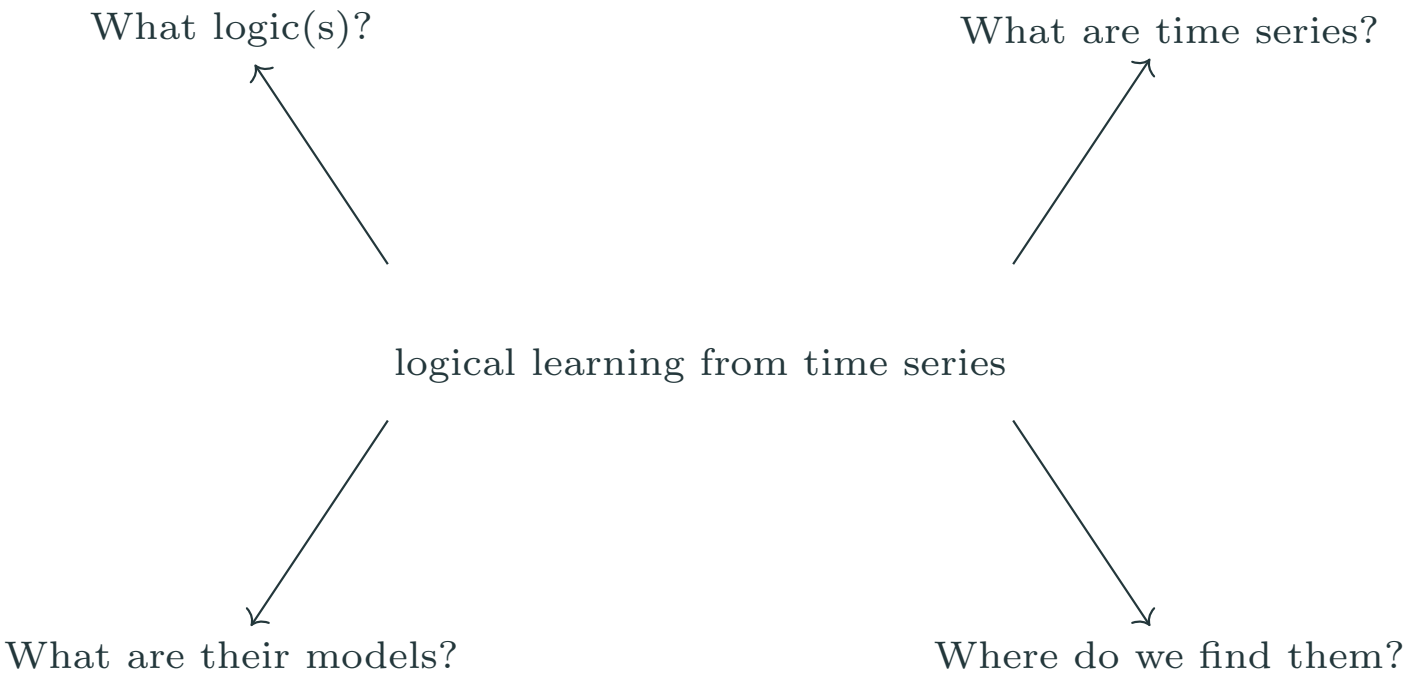
Today's menu

logical learning from time series

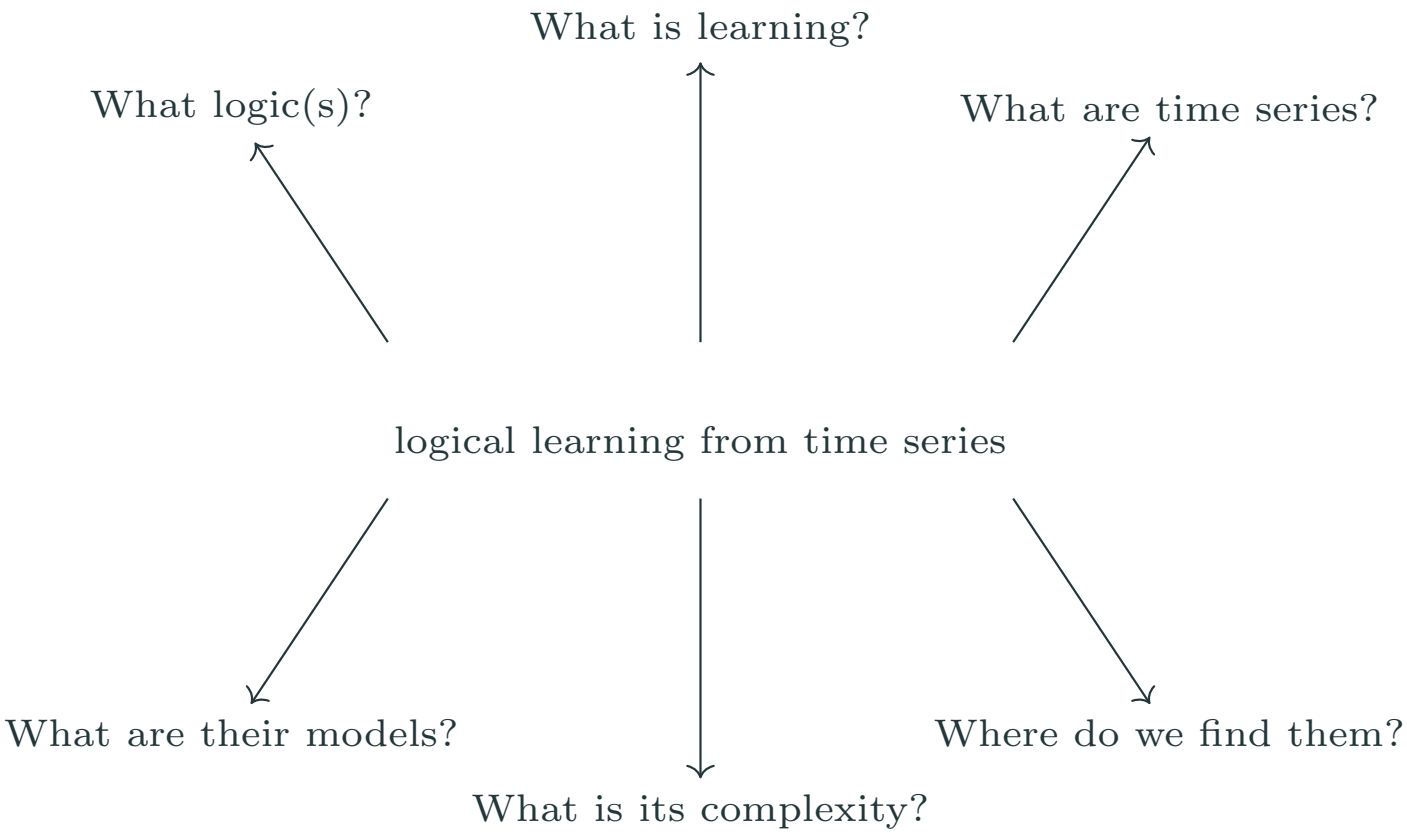
Today's menu


What are time series?
logical learning from time series
Where do we find them?

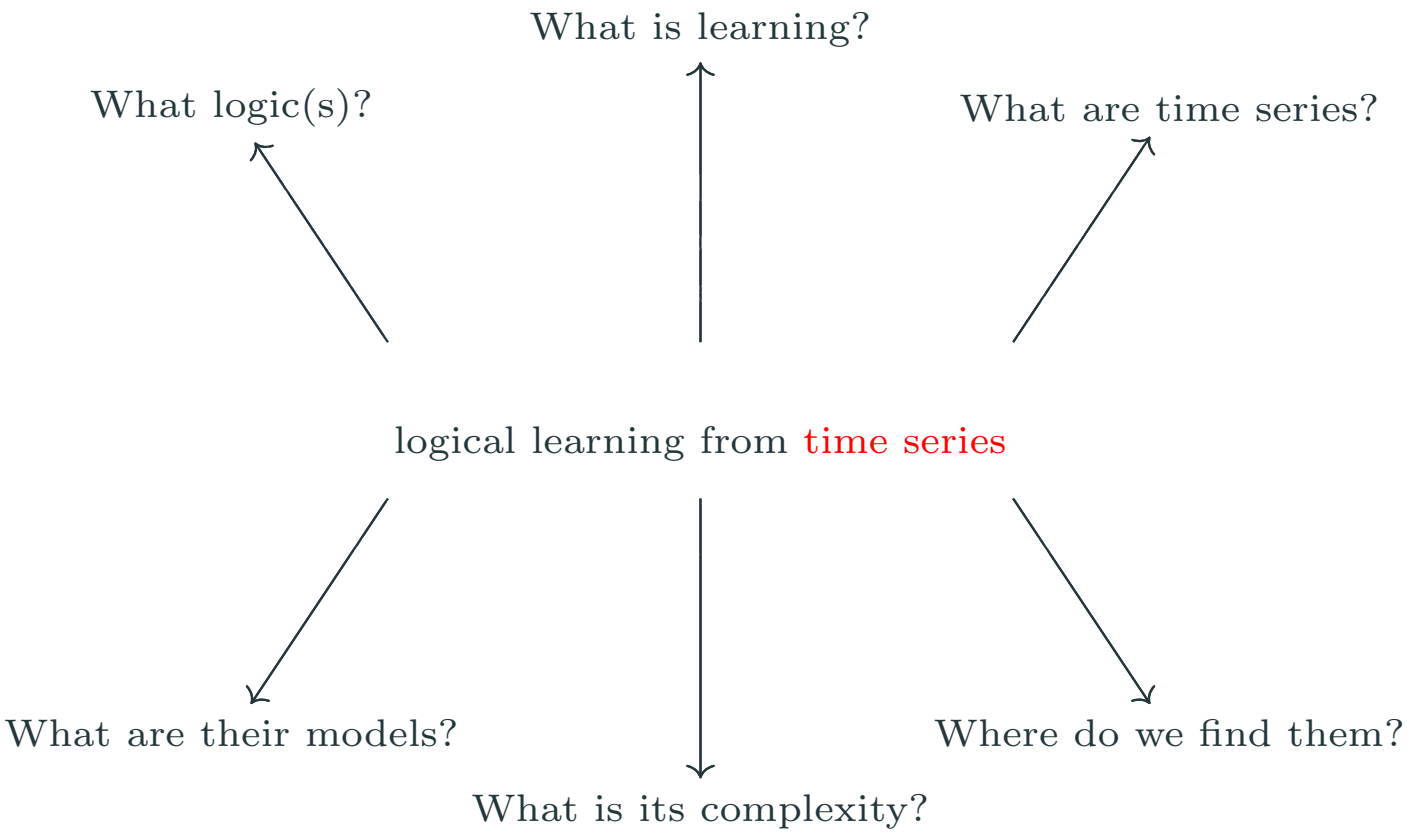
Today's menu


What logic(s)?
What are time series?
logical learning from time series
What are their models?
Where do we find them?

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Today's menu

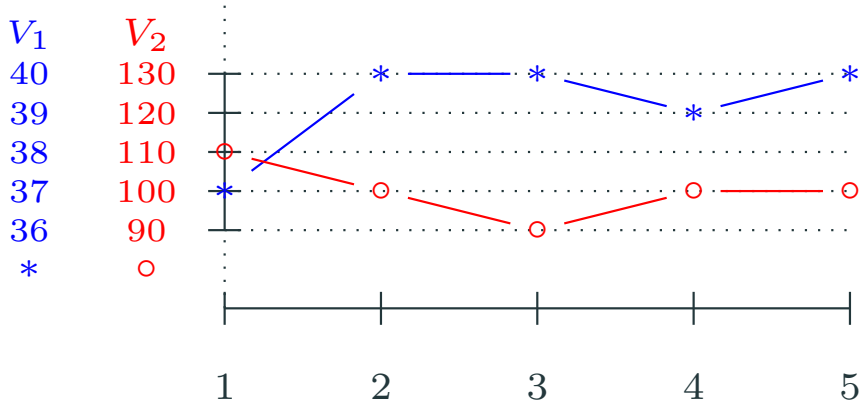

What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Appetizer: Strange Time Series and Where to Find Them

what is a time series?

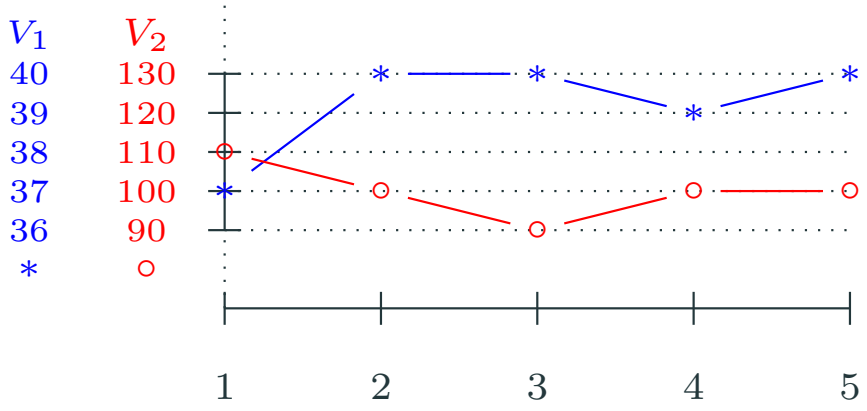
Appetizer: Strange Time Series and Where to Find Them

what is a time series?



Appetizer: Strange Time Series and Where to Find Them

what is a time series?



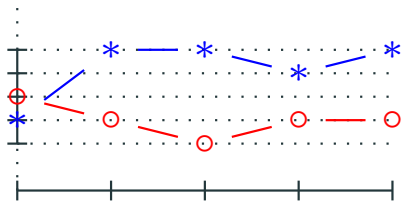
Let $\mathbb{D} = \langle \{1, 2, \ldots, N\}, <, = \rangle$ be a finite linear order (the domain) of size N, and let $V : \mathbb{D} \to \mathbb{R}$ be a temporal variable or time series; also, let $\mathcal{V}$ be a vocabulary of temporal variables. A multivariate time series T is a collection $T = \{V_1, \ldots, V_n \mid V_i \in \mathcal{V}, 1 \leq i \leq n\}$ of time series. In the above example, and $N = 5$ and $n = 2$.

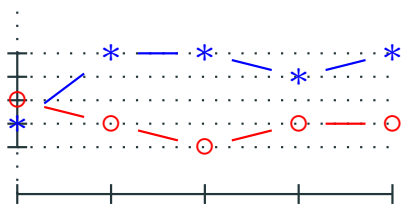
Appetizer: Strange Time Series and Where to Find Them

what is a temporal dataset?

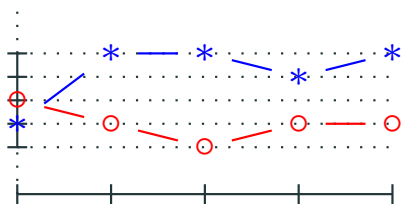
Appetizer: Strange Time Series and Where to Find Them

what is a temporal dataset?



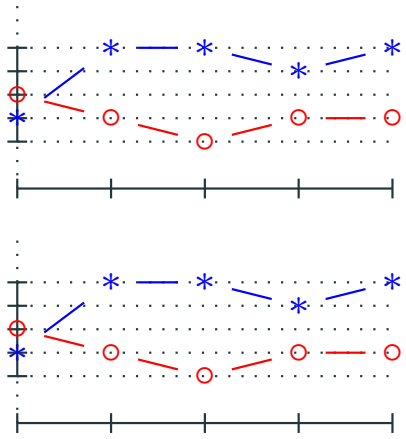


. . .

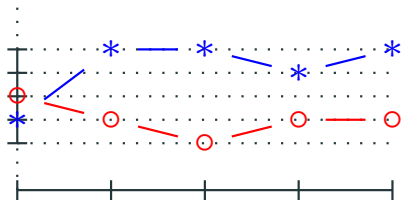


Appetizer: Strange Time Series and Where to Find Them

what is a temporal dataset?



...



A temporal dataset is a collection
$\mathcal{T} = \{T_1, \ldots, T_m\}$ of multivariate time series.

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

Appetizer: Strange Time Series and Where to Find Them

where do we find them?

from industrial machines

sensors record all types of physical
values varying varying over time
data can be used for predictive maintenance
event anticipation, and so on

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

Appetizer: Strange Time Series and Where to Find Them

where do we find them?

from industrial machines

in physiology

voice can be converted
via Fourier(-like) transform into
multi-variate time-varying signals
for several purposes, such as age and gender identification

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

recordings from exams such as EEG
can be converted via Fourier(-like) transform
into multi-variate time-varying signals, to
extract all kinds of information from brain processes

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

wearable systems can produce
time-varying quantities from accelerometers,
dynamometries, and similar tools,
to predict and detect movements and situations

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

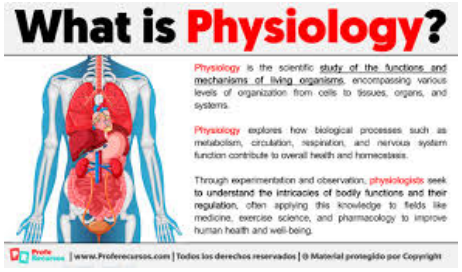
eye-tracking systems
produce time-varying signals,
that can be used to analyze gaze patterns
for example in neuroaesthetics

Appetizer: Strange Time Series and Where to Find Them

from industrial machines

where do we find them?

in physiology

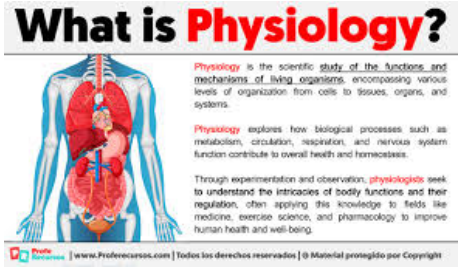


Appetizer: Strange Time Series and Where to Find Them

where do we find them?

from industrial machines

in physiology



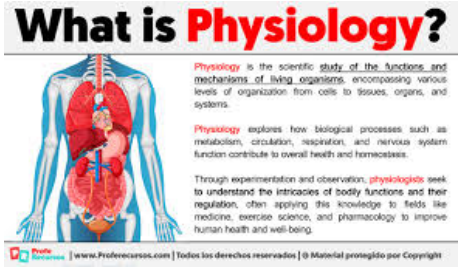
in physiopathology

Appetizer: Strange Time Series and Where to Find Them

where do we find them?

from industrial machines

in physiology



sensors and tests on hospitalized
and non-hospitalized patients produce,
medical time-varying data of interest for
diagnosis and monitoring

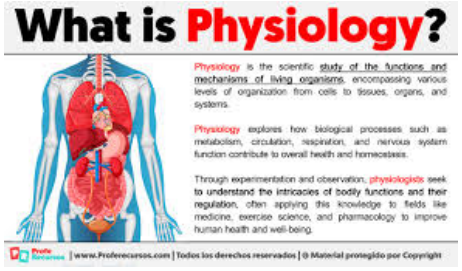
in physiopathology

Appetizer: Strange Time Series and Where to Find Them

where do we find them?

from industrial machines

in physiology



in physiopathology

from text

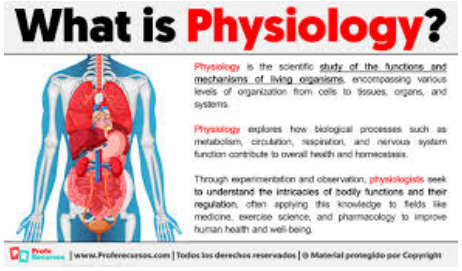
It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness...

Appetizer: Strange Time Series and Where to Find Them

where do we find them?

from industrial machines

in physiology



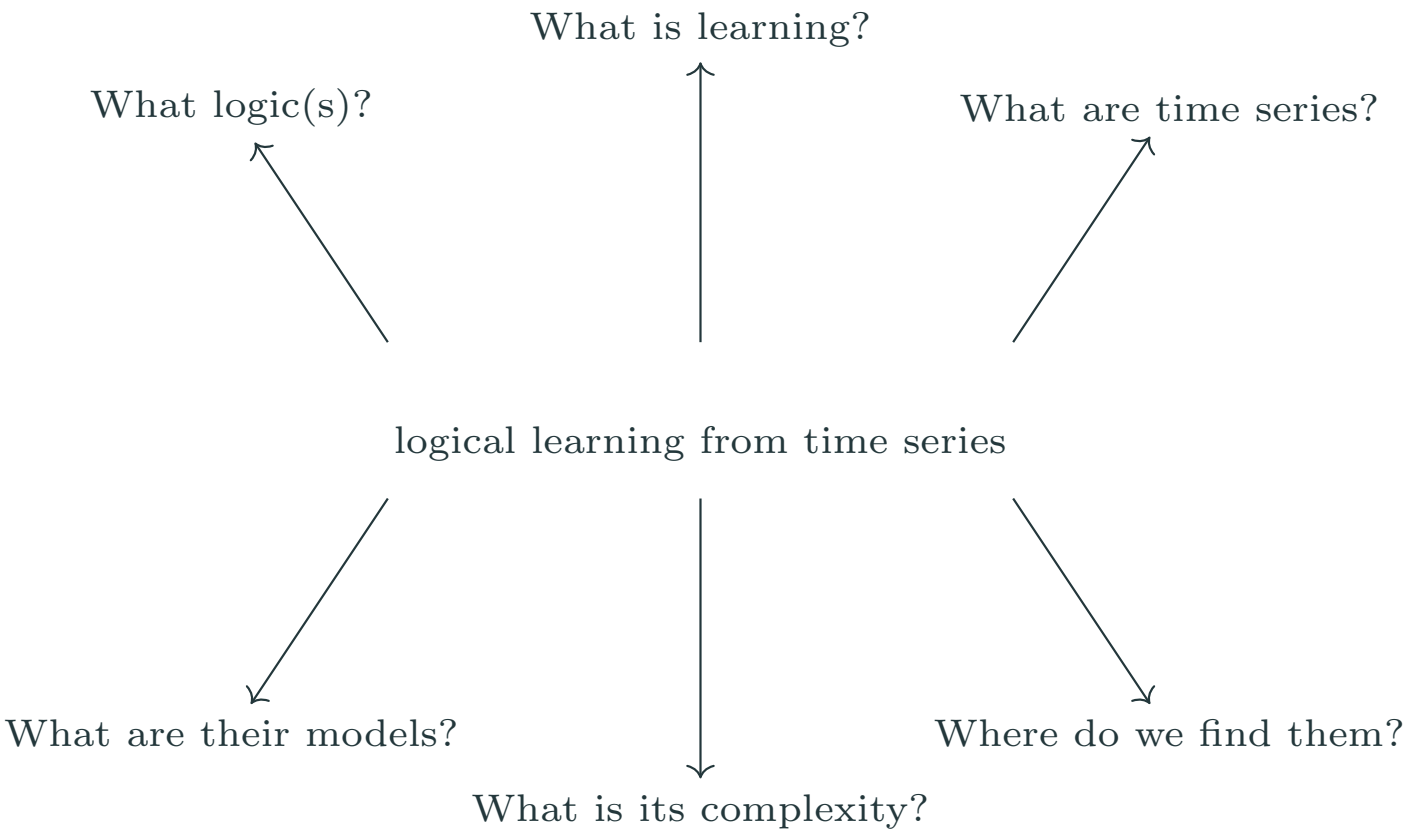
text can be interpreted
as a time-varying signal in several,
different ways, with several
different purposes

in physiopathology

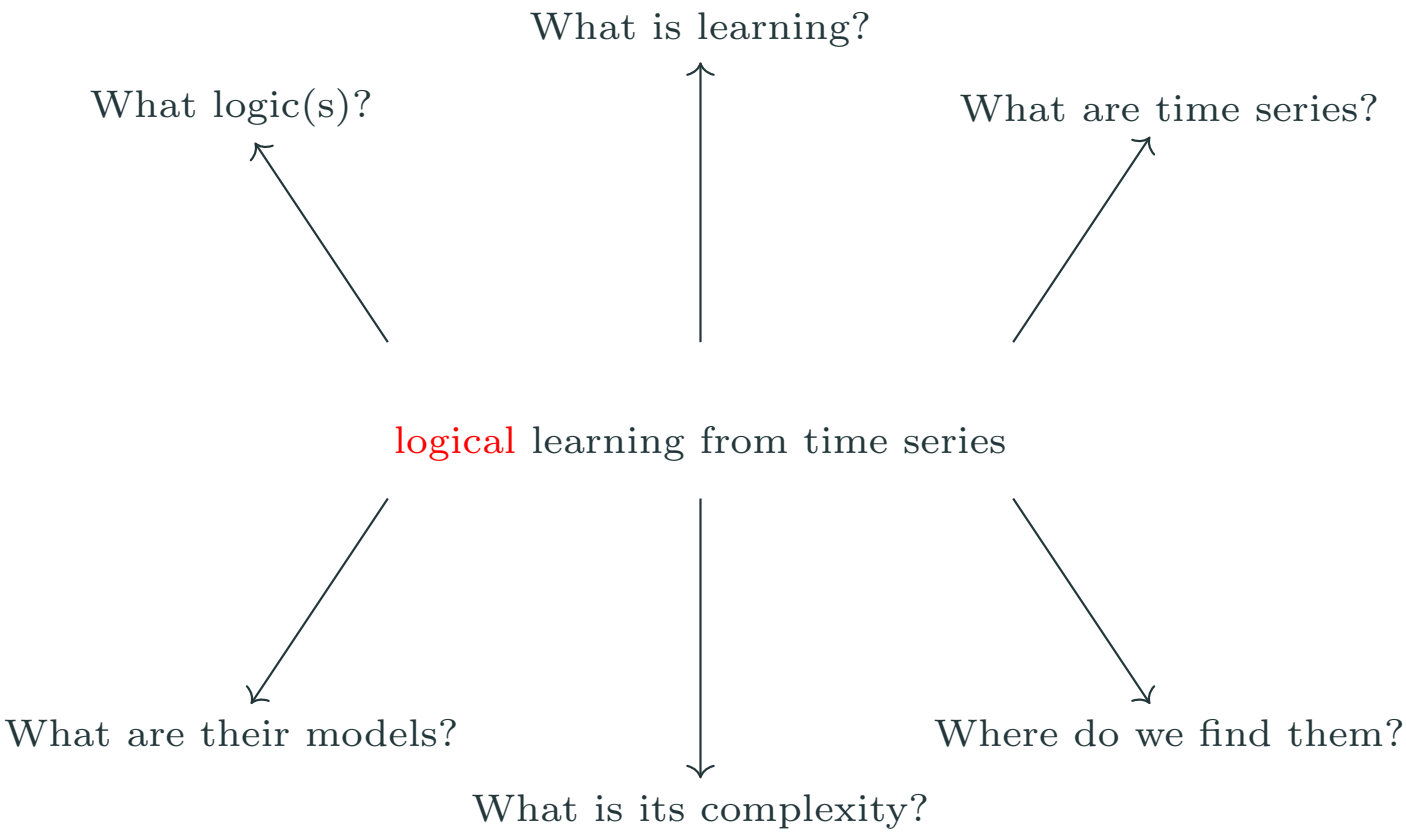
from text

It was the best of
times, it was the worst
of times, it was the age
of wisdom, it was the
age of foolishness...

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

First Course: Logics, one Approach to Rule Them All

we want to give meaning to

$$\mathcal{T} \models \varphi$$

where $\mathcal{T}$ is a temporal dataset and $\varphi \in L$, for some logical language L

to this end, we first need to establish the meaning of

$$T \models \varphi$$

where $T \in \mathcal{T}$ and $\varphi \in L$, for some logical language L

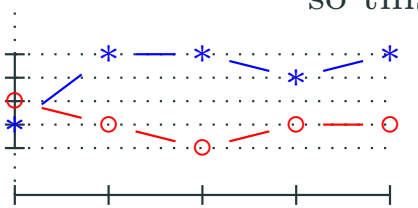
First Course: Logics, one Approach to Rule Them All

In propositional logic (PL) we simply consider a set
$\mathcal{F} = \{F_1, \ldots, F_k\}$ of feature extraction functions, applied to
every variable of the time series in order to
convert T into a vectorial description, which is a model for a propositional formula
in the language of the letters $\mathcal{P} = \{F_i(V_j) \bowtie r \mid F \in \mathcal{F}, V \in \mathcal{V}, r \in \mathbb{R}\}$

First Course: Logics, one Approach to Rule Them All

In propositional logic (PL) we simply consider a set
$\mathcal{F} = \{F_1, \ldots, F_k\}$ of feature extraction functions, applied to
every variable of the time series in order to
convert T into a vectorial description, which is a model for a propositional formula
in the language of the letters $\mathcal{P} = \{F_i(V_j) \bowtie r \mid F \in \mathcal{F}, V \in \mathcal{V}, r \in \mathbb{R}\}$

so this



$\models \varphi,\ \varphi \in PL$

First Course: Logics, one Approach to Rule Them All

In propositional logic (PL) we simply consider a set
$\mathcal{F} = \{F_1, \ldots, F_k\}$ of feature extraction functions, applied to
every variable of the time series in order to
convert T into a vectorial description, which is a model for a propositional formula
in the language of the letters $\mathcal{P} = \{F_i(V_j) \bowtie r \mid F \in \mathcal{F}, V \in \mathcal{V}, r \in \mathbb{R}\}$

becomes this

$$[F_1(V_1), F_2(V_1), \ldots, F_k(V_1), F_1(V_2), F_2(V_2), \ldots, F_k(V_2), \ldots] \models \varphi,\ \varphi \in PL$$

$\mathcal{F}$ typically contains maximum, minimum, mean,
and generic statistical description functions that are applied
to the entire time series, variable per variable

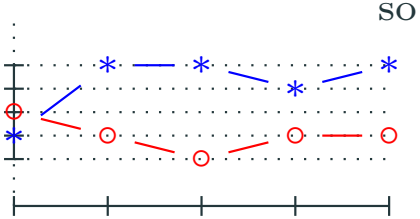
First Course: Logics, one Approach to Rule Them All

In linear temporal logic (e.g., LTL/STL) we consider the value
of each variable at each time point
through no feature extraction function, but preserving
their relative temporal order, *de facto* converting
T into a point-based temporal model that allows us to interpret
a point-based temporal formula in the language of the letters $\mathcal{P} = \{V_j \bowtie r \mid V \in \mathcal{V}, r \in \mathbb{R}\}$

First Course: Logics, one Approach to Rule Them All

In linear temporal logic (e.g., LTL/STL) we consider the value
of each variable at each time point
through no feature extraction function, but preserving
their relative temporal order, *de facto* converting
T into a point-based temporal model that allows us to interpret
a point-based temporal formula in the language of the letters $\mathcal{P} = \{V_j \bowtie r \mid V \in \mathcal{V}, r \in \mathbb{R}\}$

so this



$\models \varphi,\ \varphi \in LTL/STL$

First Course: Logics, one Approach to Rule Them All

In linear temporal logic (e.g., LTL/STL) we consider the value
of each variable at each time point
through no feature extraction function, but preserving
their relative temporal order, *de facto* converting
T into a point-based temporal model that allows us to interpret
a point-based temporal formula in the language of the letters $\mathcal{P} = \{V_j \bowtie r \mid V \in \mathcal{V}, r \in \mathbb{R}\}$

becomes this

$$\begin{array}{l} V_1 > r_1^1 \\ V_2 < r_2^1 \end{array} \quad \cdots \quad \begin{array}{l} V_1 \leq r_1^3 \\ V_2 \geq r_2^3 \end{array} \qquad \models \varphi,\ \varphi \in LTL/STL$$

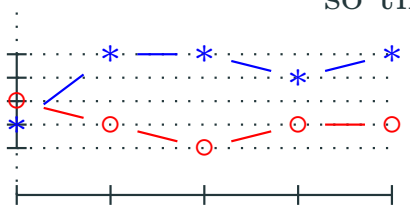
First Course: Logics, one Approach to Rule Them All

With interval temporal logic (e.g., HS) we can generalize both the previous approaches into a single one: since feature extraction functions can be applied to every interval of time series (which is a times series on its own) on every interval we can evaluate the truth of propositional letters from the set $\mathcal{P} = \{F(V_j) \bowtie r \mid V \in \mathcal{V}, r \in \mathbb{R}\}$, allowing us to interpret a multivariate time series as an interval temporal model

First Course: Logics, one Approach to Rule Them All

With interval temporal logic (e.g., HS) we can generalize both the previous approaches into a single one: since feature extraction functions can be applied to every interval of time series (which is a times series on its own) on every interval we can evaluate the truth of propositional letters from the set $\mathcal{P} = \{F(V_j) \bowtie r \mid V \in \mathcal{V}, r \in \mathbb{R}\}$, allowing us to interpret a multivariate time series as an interval temporal model

so this

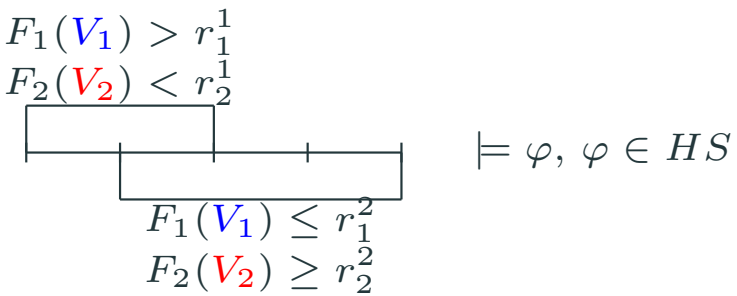


$\models \varphi,\ \varphi \in LTL$

First Course: Logics, one Approach to Rule Them All

With interval temporal logic (e.g., HS) we can generalize both the previous approaches into a single one: since feature extraction functions can be applied to every interval of time series (which is a times series on its own) on every interval we can evaluate the truth of propositional letters from the set $\mathcal{P} = \{F(V_j) \bowtie r \mid V \in \mathcal{V}, r \in \mathbb{R}\}$, allowing us to interpret a multivariate time series as an interval temporal model

becomes this



$\models \varphi, \varphi \in HS$

First Course: Logics, one Approach to Rule Them All

in the end, we understand that

$$\mathcal{T} \models \varphi$$

where $\varphi \in L$, for some logical language L

First Course: Logics, one Approach to Rule Them All

is a function of

$$T_1 \models \varphi, T_2 \models \varphi, \ldots, T_n \models \varphi$$

First Course: Logics, one Approach to Rule Them All

is a function of

$$T_1 \models \varphi, T_2 \models \varphi, \ldots, T_n \models \varphi$$

but what function?

First Course: Logics, one Approach to Rule Them All

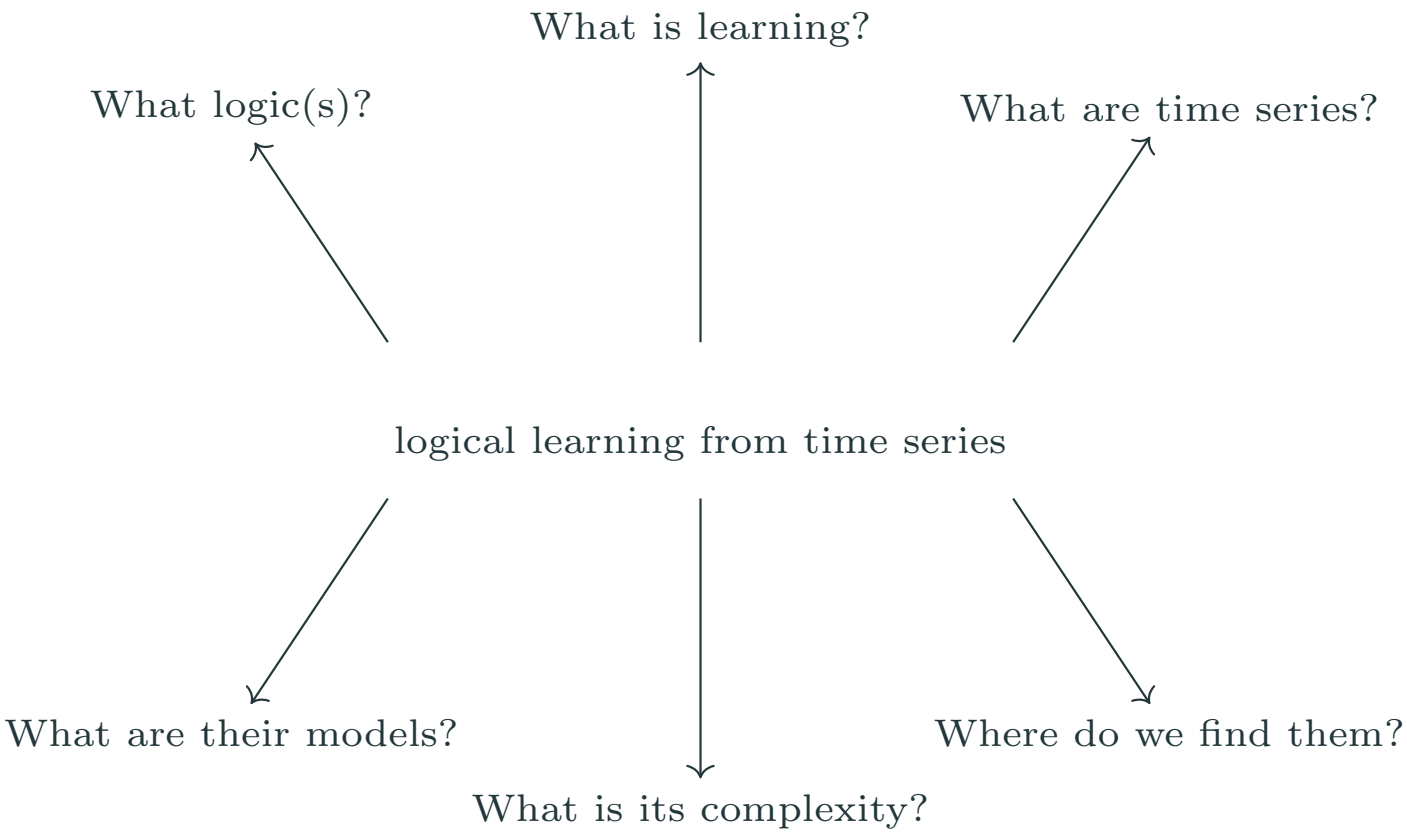
is a function of

$$T_1 \models \varphi, T_2 \models \varphi, \ldots, T_n \models \varphi$$

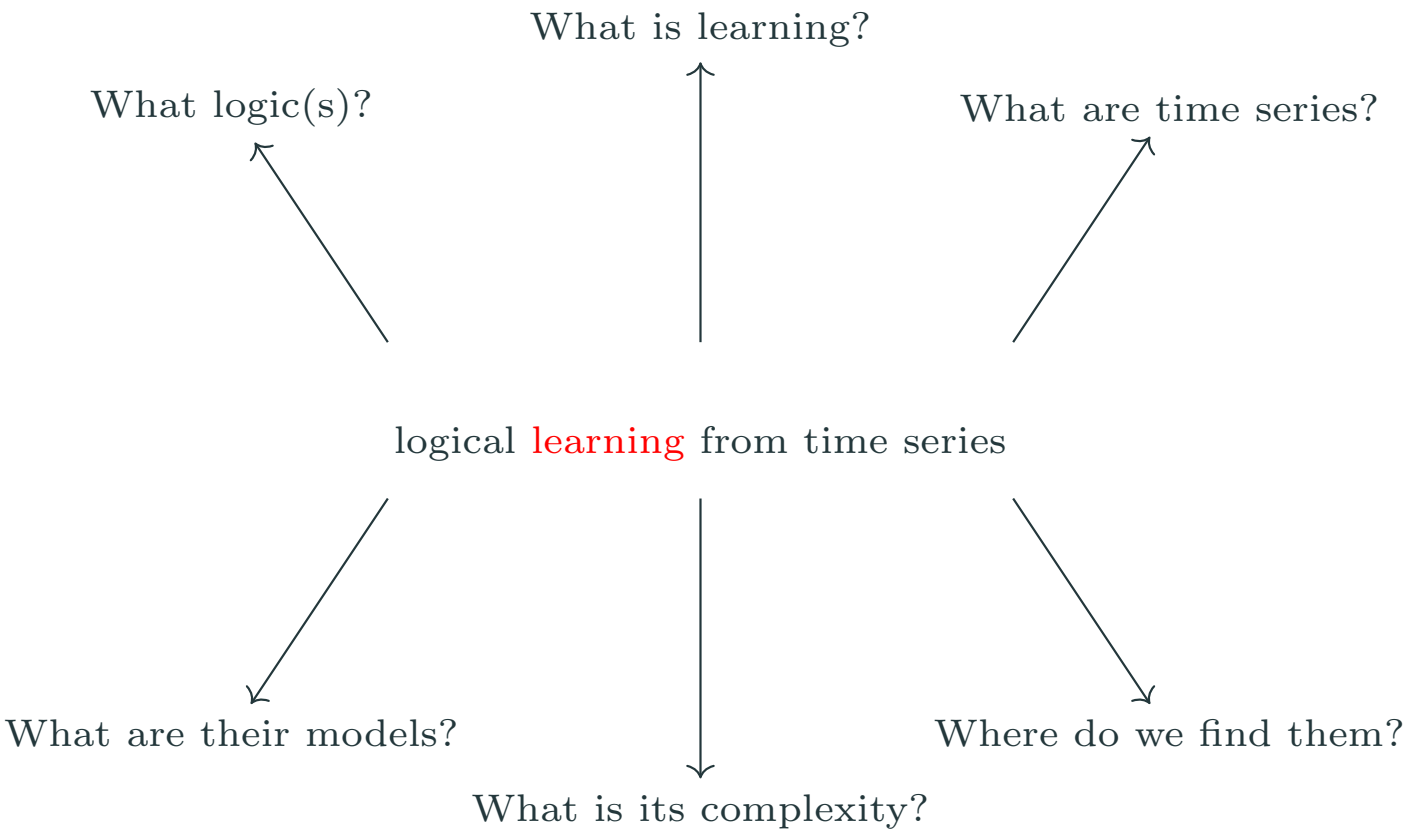
but what function?

Since a multivariate time series can be seen as a model of a logic formula that belongs to a suitable logic; extracting a formula from a set of time series is a learning problem that gives us the constraints that such formula should meet, and therefore the complexity of the problem itself

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Second Course: Learning, Actually

what is learning (from multivariate time series)?

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

learning

with 2 or more classes (supervised)

Second Course: Learning, Actually

what is learning (from multivariate time series)?

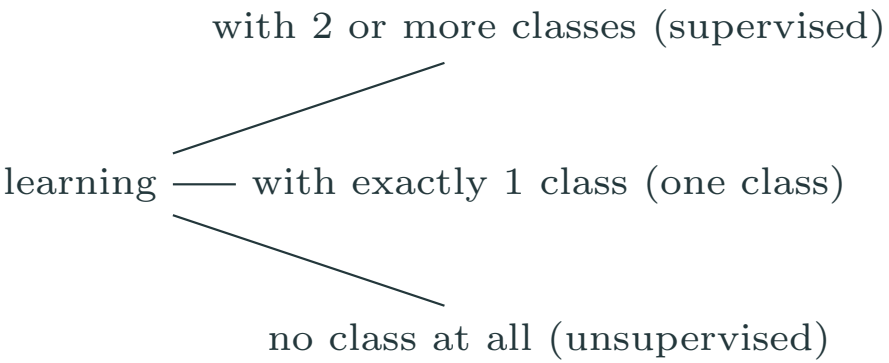
symbolic (i.e., logic) learning
comes in three popular flavours

learning —— with 2 or more classes (supervised)
learning —— with exactly 1 class (one class)

Second Course: Learning, Actually

what is learning (from multivariate time series)?

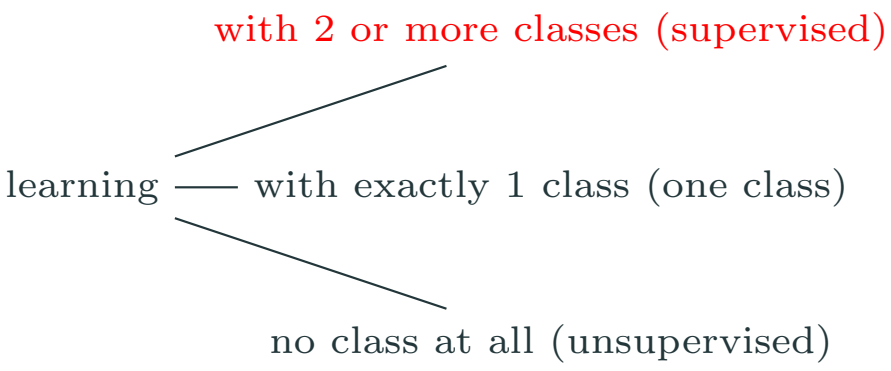
symbolic (i.e., logic) learning
comes in three popular flavours



Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

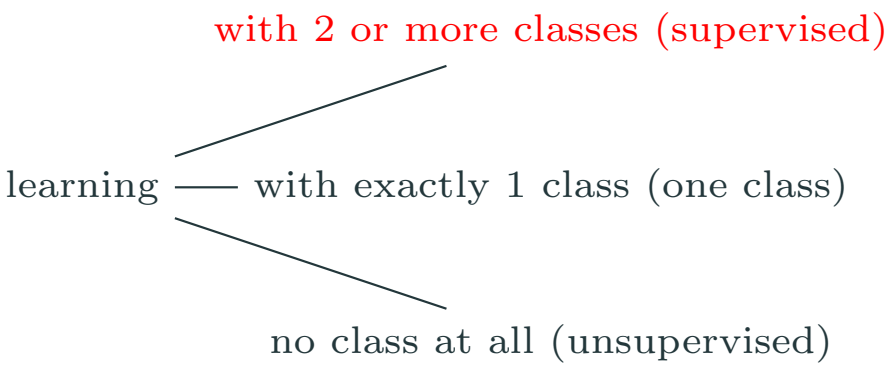


supervised classification, to
learn the differences between
classes that is, to learn
one or more concepts

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

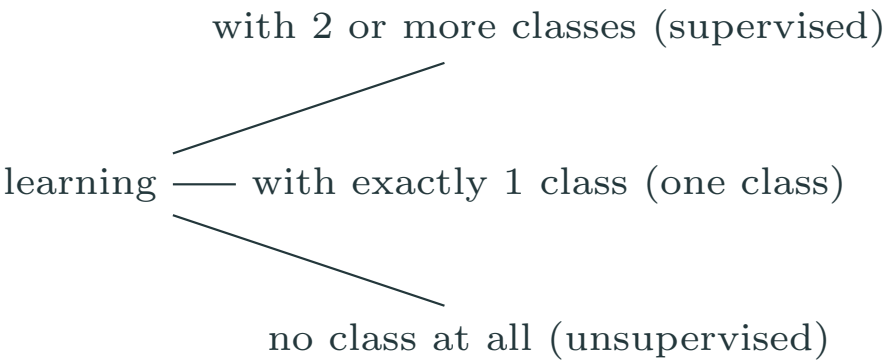


φ should be minimal in size
(to avoid overfitting) and
accurate (in training,
test, cross-validation. . .)

Second Course: Learning, Actually

what is learning (from multivariate time series)?

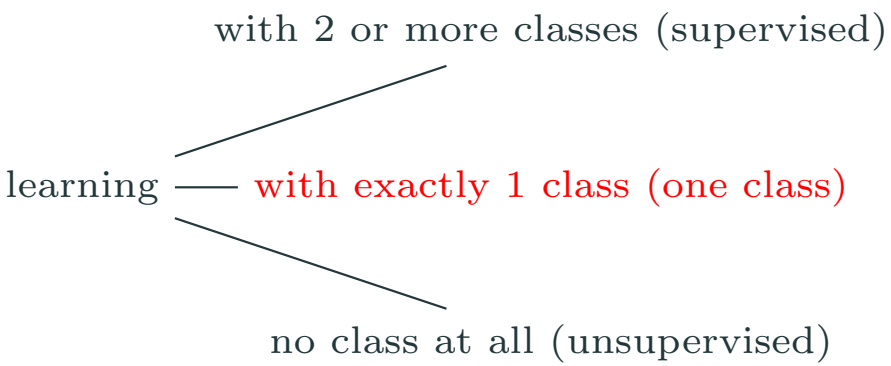
symbolic (i.e., logic) learning
comes in three popular flavours



Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

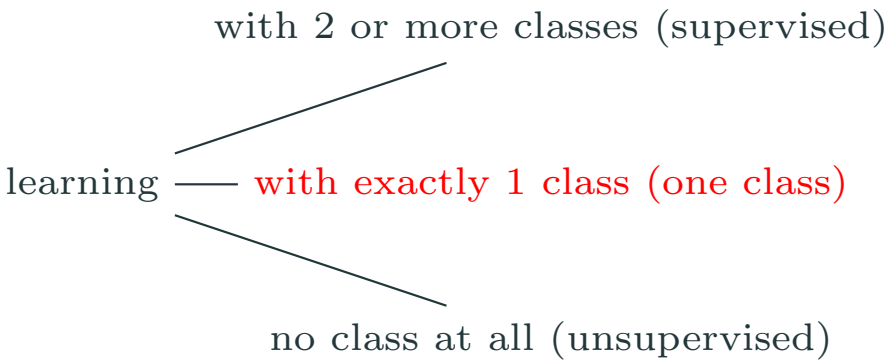


single classification, to
describe an entire dataset
and extract the essential idea

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

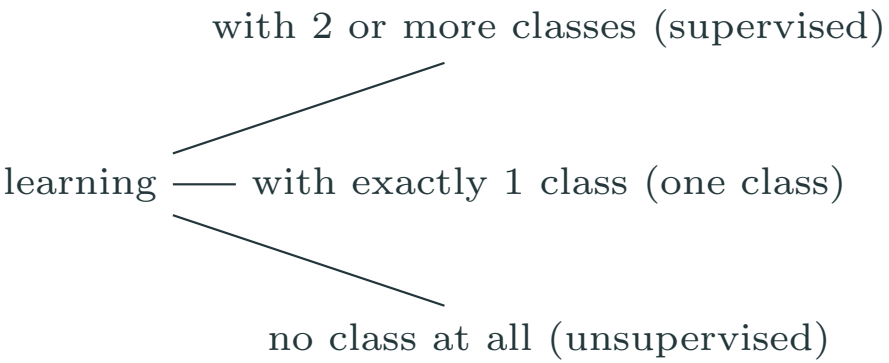


φ should be minimal in size,
satisfied by all time series
in the dataset, and subject
to some *ad hoc* semantical and
syntactical constraint to rule out
trivial formulas

Second Course: Learning, Actually

what is learning (from multivariate time series)?

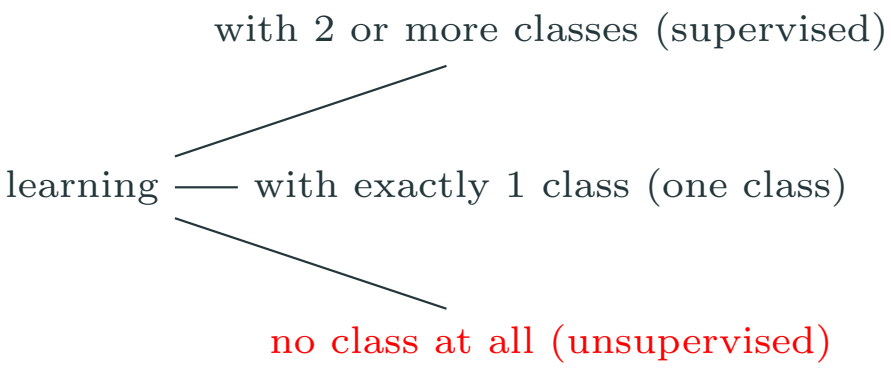
symbolic (i.e., logic) learning
comes in three popular flavours



Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

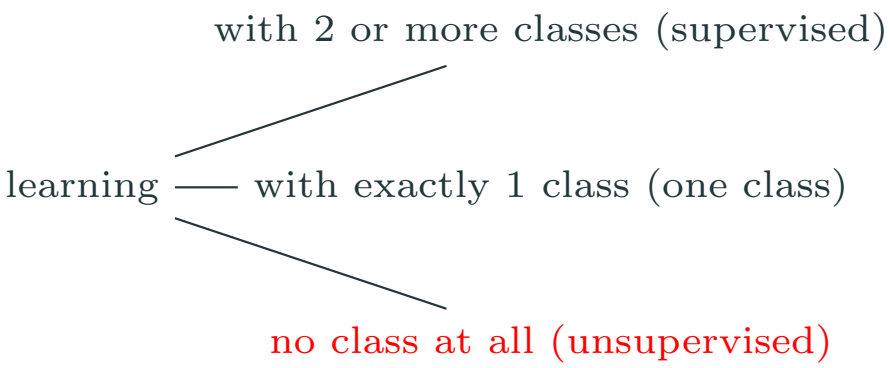


clustering, to
extract common patterns
that relate variables in some
non-trivial way

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

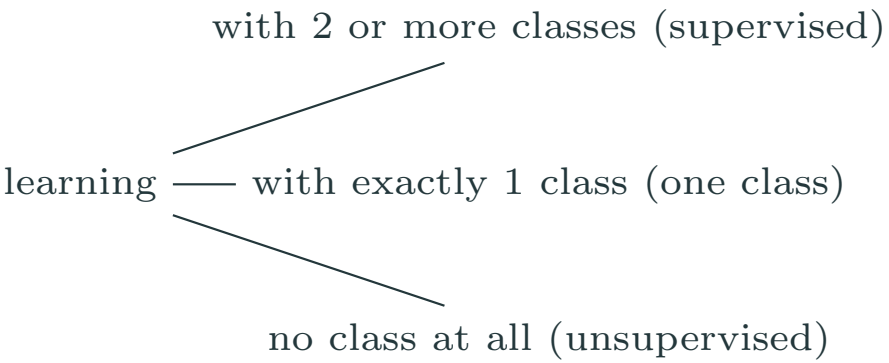


φ should be maximal in size,
satisfied by the maximal number
of time series and subject
to specific syntactic constraints

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

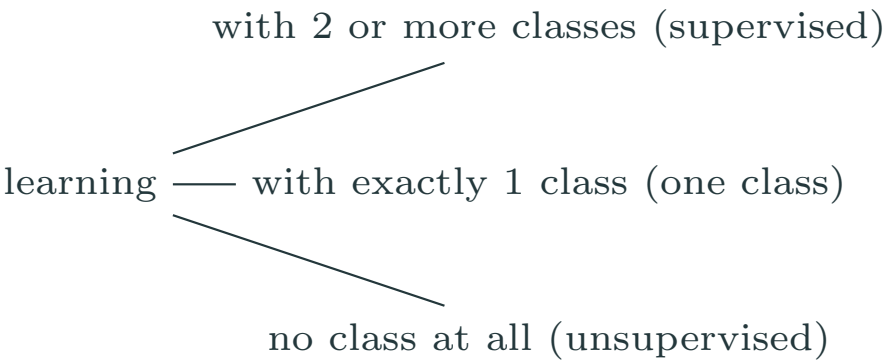


Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours

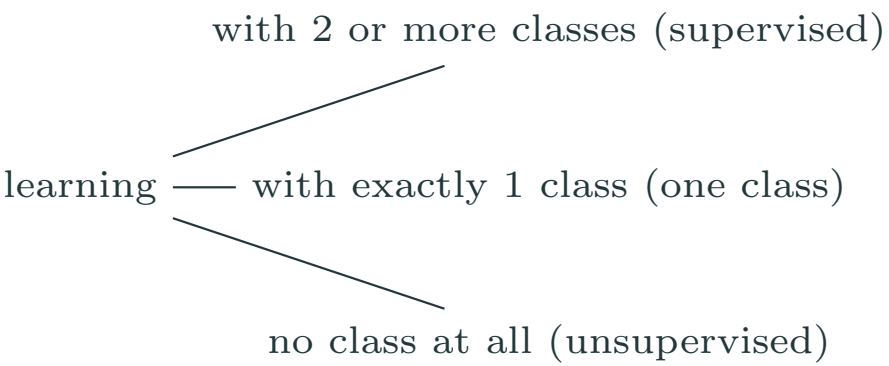
φ can be extracted
in several forms



Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours



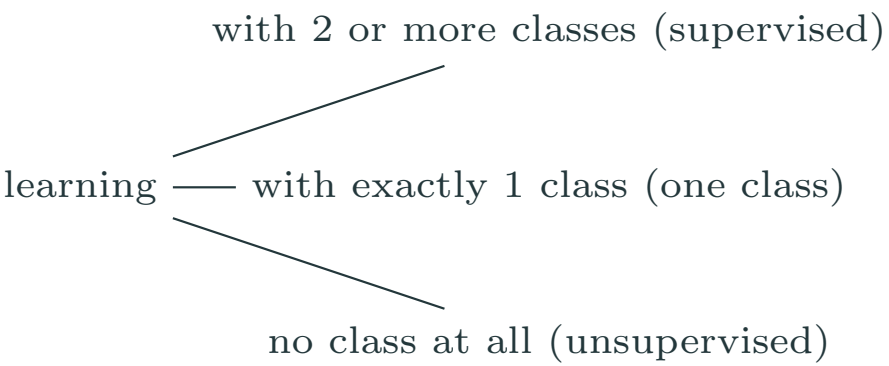
φ can be extracted
in several forms

as a decision tree
which can be immediately
turned into a logic
formula

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours



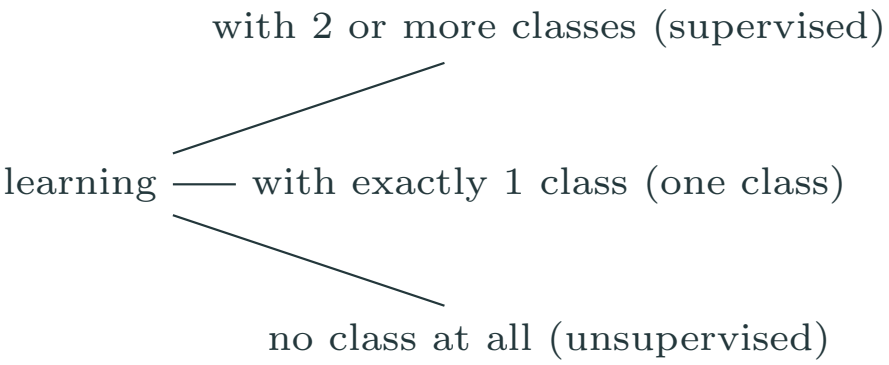
φ can be extracted
in several forms

as a decision list
which can be also
immediately turned
into a logic formula

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours



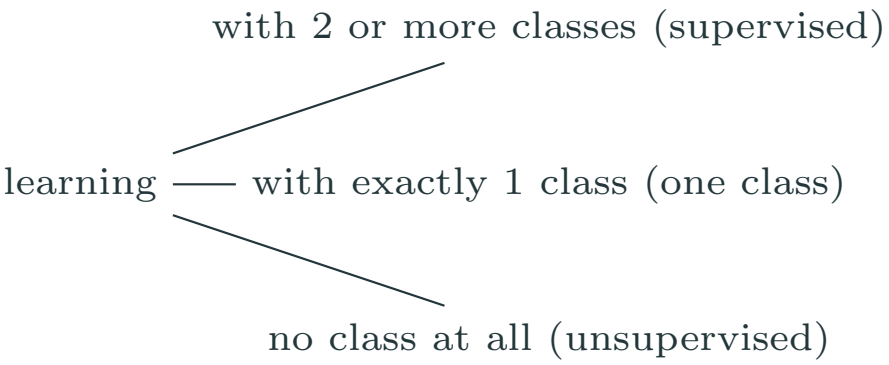
φ can be extracted
in several forms

as a logical formula, directly
with or without the
constraint of belonging
to a specific grammar

Second Course: Learning, Actually

what is learning (from multivariate time series)?

symbolic (i.e., logic) learning
comes in three popular flavours



φ can be extracted
in several forms

but we can equivalently
say that the temporal
learning problem consists
of extracting a logic formula
from a temporal dataset

Second Course: Learning, Actually

what is learning (from multivariate time series)?

the logical temporal learning problem is the problem, given a temporal dataset $\mathcal{T}$ of extracting a (temporal) logic formula φ such that $optimize(\mathcal{T}, \varphi) \geq z$, for some optimization function $optimize()$

which combinations are known?

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL			
LTL/STL			
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete		
LTL/STL			
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete	?, probably trivial in NP	
LTL/STL			
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete	?, probably trivial in NP	classic ass. rules NP-hard
LTL/STL			
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete	?, probably trivial in NP	classic ass. rules NP-hard
LTL/STL	partially known NP-complete		
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete	?, probably trivial in NP	classic ass. rules NP-hard
LTL/STL	partially known NP-complete	partially known NP-complete	
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete	?, probably trivial in NP	classic ass. rules NP-hard
LTL/STL	partially known NP-complete	partially known NP-complete	mostly unexplored NP-hard
HS			

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
	classic DTs,DLs	?, probably trivial	classic ass. rules
PL	NP-complete	in NP	NP-hard
	partially known	partially known	mostly unexplored
LTL/STL	NP-complete	NP-complete	NP-hard
	partially known		
HS	NP-hard		

Second Course: Learning, Actually

which combinations are known?

	supervised	one class	unsupervised
PL	classic DTs,DLs NP-complete	?, probably trivial in NP	classic ass. rules NP-hard
LTL/STL	partially known NP-complete	partially known NP-complete	mostly unexplored NP-hard
HS	partially known NP-hard	? NP-hard	

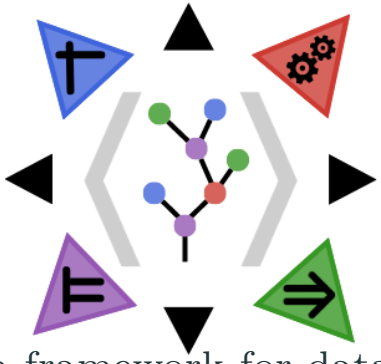
Second Course: Learning, Actually

	supervised	one class	unsupervised
	classic DTs,DLs	?, probably trivial	classic ass. rules
PL	NP-complete	in NP	NP-hard
	partially known	partially known	mostly unexplored
LTL/STL	NP-complete	NP-complete	NP-hard
	partially known	?	partially known
HS	NP-hard	NP-hard	NP-hard

Second Course: Learning, Actually

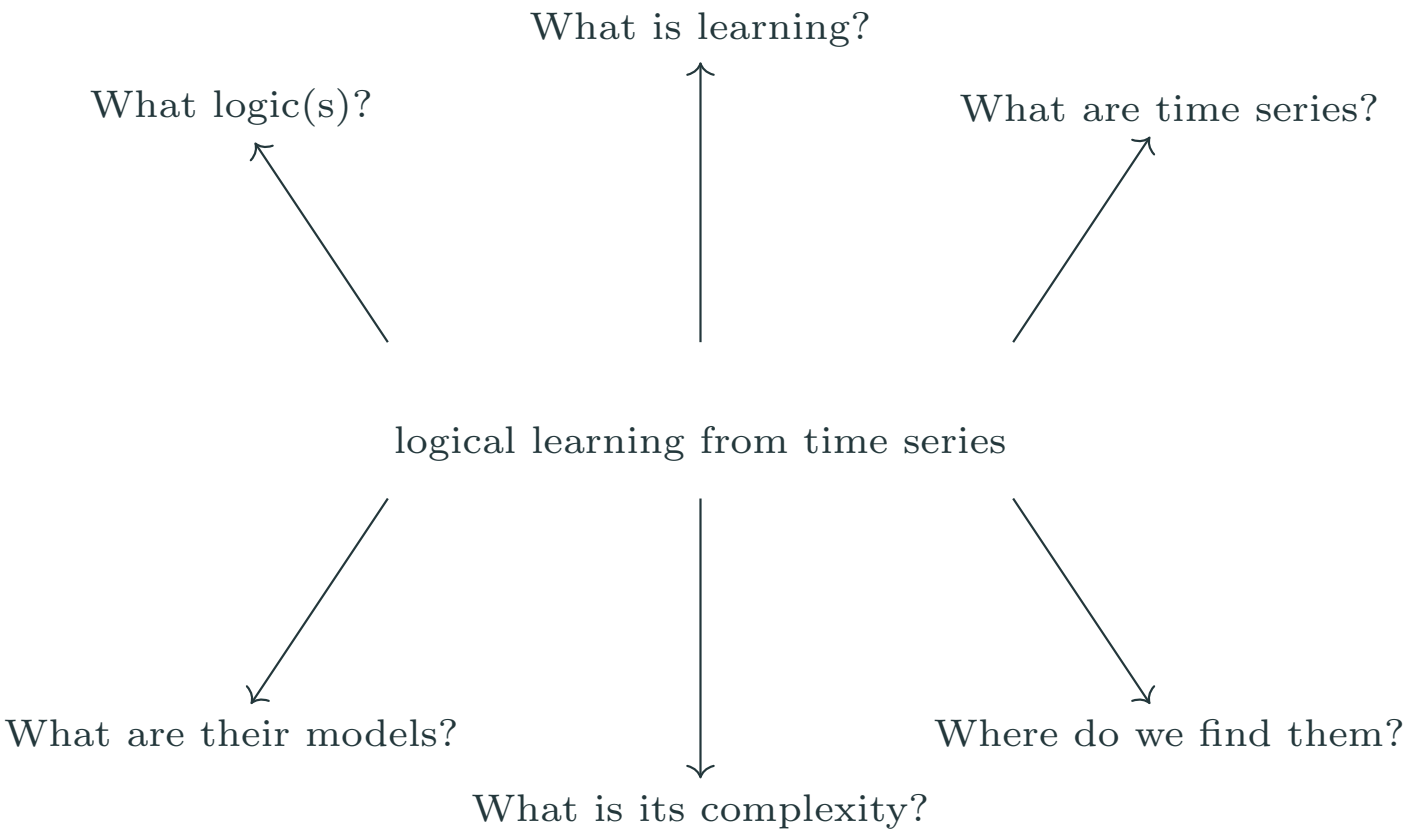
as a final comment, we should mention that formulas are
mainly learnt in one of two ways: deterministically, such as
in the case of DTs, DLs, and ensambles, and
non-deterministically, such as
in the case of pre-formed formulas whose parameters must be
set via evolutionary algorithms or
similar meta-heuristic approaches, and
from a wider perspective, it should be said
that there are no available integrated approach, nor
framework that allows one to compare the results
and study the different solutions.

Second Course: Learning, Actually

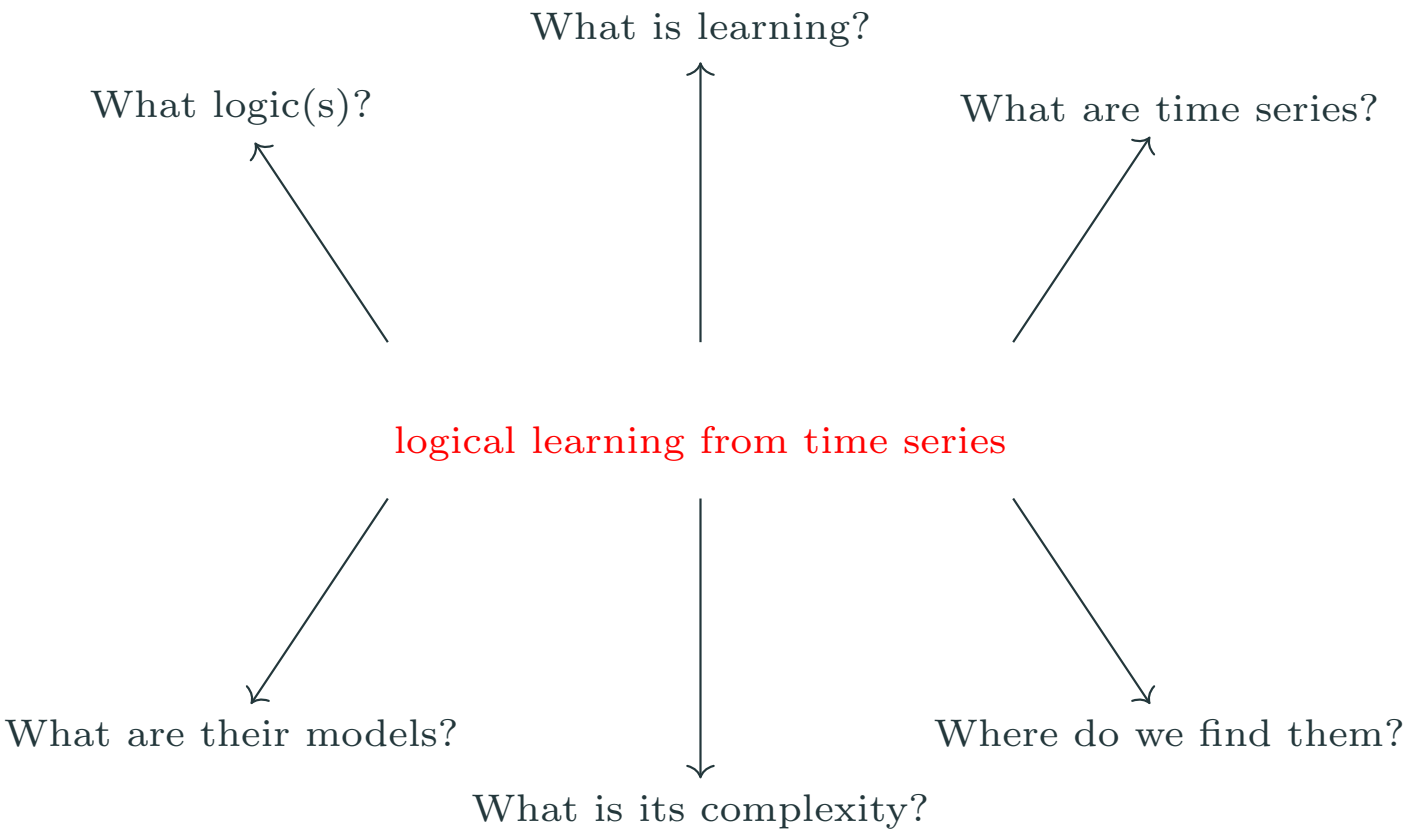


SOLE.jl is the Julia framework for data analysis, learning, reasoning, and model visualization and post-hoc fine tuning developed at the University of Ferrara that aims to establish itself as the standard open source tool for extracting logical knowledge from non-tabular data, including temporal data

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Today's menu


What is learning?
What logic(s)?
What are time series?
logical learning from time series
What are their models?
What is its complexity?
Where do we find them?

Dessert: Temporal Logic Learning or Nothing

as we have seen, time series, and temporal datasets, are literally ubiquitous;
this should give one the sense of the importance of the role that
learning from time series
plays in the applied artificial intelligence landscape

Dessert: Temporal Logic Learning or Nothing

the supposed generality of non-logical, pre-trained
function-based learning systems (a.k.a. neural networks, in their
countless variants) is often sold off as an alternative for every
learning situations, including temporal one
but not only they are usual designed as a alchemy of billions of parameters,
they also allow no inspection at all, certainly no logical reading,
and are very rarely adaptable to real, data-scarce, and explanation-hungry cases

Dessert: Temporal Logic Learning or Nothing

the supposed generality of non-logical, pre-trained
function-based learning systems (a.k.a. neural networks, in their
countless variants) is often sold off as an alternative for every
learning situations, including temporal one
but not only they are usual designed as a alchemy of billions of parameters,
they also allow no inspection at all, certainly no logical reading,
and are very rarely adaptable to real, data-scarce, and explanation-hungry cases