Machine, Data and Learning

Druhan Rajiv Shah

Spring - 2023

Contents

DBS community	ML Community
Indexing	Vision
Query languages	Speech
Query optimizing	NLP
Transactions	Feature extraction
XML, Semantic web	Genetic/Swarm algirithms
OO/ORDBMS	Fuzzy logic/Neural networks

Course Structure

Textbooks

- Python ML - Yuxi Liu
- AI, A modern approach - Russell and Norvig

Grading scheme (tentative)

Midsem: 30%

Endsem: 35-40%

Quiz 2: 10%

Assignments (2-4): 20-25%

Topics of discussion

- Overview of AI/ML
- Logic
- Data and generalization
- Overfitting, Underfitting, Bias-variance tradeoff
- Cross-validation, Regularization
- Feature selection, preprocessing, feature engineering
- Search, informed search, game search
- —– Half sem —–
- Data Science, association rules
- Clustering, classification
- Probability and Utility Theory
- Reinforcement Learning
- Derision theory, Markov derision theory, Linear programming

1 Introduction

Communities Two main communities in the field

1.1 AI

Thinking styles Note that action here avoids philosophical issues like "is the system conscious" and such

Think like humans (e.g. Cog. Sciences)

Think raionally (e.g. Logic)

Act like humans (e.g. Turing test)

Act rationally (e.g. Logic + Domain)

Definition Most people prefer the definition of AI as acting rationally. (Intelligence/rationality is not to be confused with Conscience)

Turing Test System that acts like a human (description)

Chinese Room argument involves a person in a room who knows English but not Hindi. He receives notes in Hindi. He has a systematic English rulebook to write in Hindi given English. He only follows the rulebook.

Finally, Searle claims that there is no intelligence here, because the only systems here are following a strict set of rules and nothing more.

(Okay, philosophy done :P)

2 Propositional Logic and Reasoning

Mainly about obtaining new facts by applying a fixed number of operations to existing facts.

Expert systems Back then, before the advent of neural networks, there used to be Expert Systems, which were basically just a bunch of `if-else` statements. These systems relied heavily on Propositional Logic, and required the programmers to know the field absolutely, unlike the NN models (hence the name, as it required Experts). On the other hand, such models are advantageous over NN models because we can tell exactly the reasoning behind all the decisions made by the system.

AI winter Back in the time of Expert systems being the crux of AI, people were losing trust in AI for being able to solve problems, and even NN models. However, with the development of neural networks, Logic is now in a sort of 'winter'.

2.1 Knowledge

Knowledge Representation Knowledge representation is basically expressing information in computer-readable form, usually in the form of a formal language called logic. The total set of information known is called the knowledge base.

Reasoning Reasoning is the process of drawing inferences from the knowledge base. This could be in the form of answering queries, discovering new facts that follow form known information or making decisions.

Bunch of definitions

- Logic is a formal language fo representing information (as mentioned before).
- Syntax describes how to make sentences (basically how to represent the information in that language).
- Semantics is the way the sentences relate to reality. This meaning per se is not *intrinsic* to the sentence itself.
- Proof Theory is the set of rules for drawing inferences (math field Proof Theory is also the same :P).

Logical arguments Arguments can be logically valid, while having invalid conclusions because they are false despite the logic being completely correct. This simply means that the input statements are incorrect. The logic is never wrong (especially when it's a formal system).

Ok Google, translate this Conversion from English to Propositional Logic is pretty straightforward (again, à la DS/DSM). Or is it? `>:)`

Okay, to be fair, it is quite straightforward, but one needs to keep a track of 'not statements'. But otherwise, 'if', 'iff' and their variants are pretty much self-explanatory.

Why though? Natural languages all can have ambiguity (à la ITL2).

Formal languages on the other hand, promote rigour and thereby reduce the possibility of ambiguity and human error. They help reduce implicit or unstated assumptions by abstracting out the information and thus removing familiarity with the subject, which would avoud bias. They thus help achieve generality by allowing us to apply the rules obtained to things that they wouldn't usually apply to.

Logic vs semantics Advantages of logic include the ability to rely on truth tables. The point of truth tables is that we can verify the results obtained using the Proof Theory by verifying them based on our sematic knowledge of the situation (à la DSM Lab).

Note that, unlike in most contexts we've used truth tables, in general they require input statements, which **can in fact be complex**.

More terminology

- A sentence is 'valid' if it is True under all possible assignments to its propositional variables.
- A sentence is 'satisfiable' if it is true under some assignment of its propositional variables.
- Correspondingly, sentences can be 'unsatisfiable' or 'invalid'.
- Valid sentences are also called 'tautologies'.

Material Implication Conditional statements which involve a cause-effect relationship are called Material Implications. These aren't usually conveyed through the Logic, because there can be correlations without involving causations. For example, there is the statement $P \to \neg P$ is logically a tautology, but the semantics of the statement are wack.

Entailment A set S entails another set P ($S \Rightarrow P$) whenever if all the formulae in S are true, P is true. This is more of a semantic notion, than a logical one.

Recall rules like *Modus ponens*. That would lead to something like $S \Rightarrow P$ when $S = \{P, P \to Q\}$ and $P = \{Q\}$

Formal Proofs Proofs intend to capture the notion of proof that is commonly applied in other fields like mathaematics. A proof of a formula from a set of premises is a sequence of steps in which any step of the proof is either an axiom or a premise, or a formula that is already previously deduced using some rule of inference. The last step of the proof should deduce the formula we wish to prove. We say that S 'follows' from the set of premises 'P'.

Soundness and Completeness

A logic is 'sound' if it preserves truth (that is, if a set of premises is all true, then all conclusions that can be drawn are also true).

It is 'complete' if it is capable of proving all valid consequences if the premises.

It is 'decidable' if there is a mechanical procedure or computer program to prove any possible consequence.

By this definition, mathematics, as defined by the Peano axioms and the corresponding operations, is always sound, but incomplete (Hi, Kurt Gödel).