Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

Семинар по специальности на английском языке
(Workshop in English)

тема

Separating the explanations from formal (mathematical) and computer models

занятие 1

2 февраля
2022 г.

Суть проблемы - что первично

1) «материя или сознание»
2) «аппаратное или программное обеспечение»

Известно, что в 99% случаев люди воспринимают реальность не такой, какая она есть, а интерпретируем ее на основе когнитивных моделей (образцов), построенных на основе прошлого опыта.

Мир вещей, процессов и ...информации

Материальный мир, состоит из вещей и процессов, которые разделяются на две категории:

- те, о свойствах которых люди имеют эксплицитные или имплицитные знания и
- те, о которых люди знают лишь то, что они могут, некоторым образом, влиять на окружающую их материальную реальность.

В своей деятельности люди могут использовать только те вещи и процессы, о которых они имеют определенные знания или представления, т.е. те, которым можно сопоставит мыслимые понятия.

То, о чём люди ничего не мыслят или не знают, находится целиком за границей возможностей целенаправленного использования.

Очевидно, что вещи по другую сторону границы относительно познанного или мыслимого по определению являются немыслимыми.

Как можно упорядочить представления о реальности ?

Это делается введением понятия информация, считая это понятие атрибутом того, что для нас является мыслимым.

чем разница и в чем сходство между

…. физическим объектом и соответствующим ему абстрактным символом, который мыслим человеком ?

С учетом вышеизложенного, что например:

- 2А+А=3А
- А+Б=Б+А
- 2*2=4
- 1+1=0
- F=m*a
- AX=0

Computer science (CS)

(компьютерные науки)

CS interpretation:

- Perception (восприятие) - obtaining data through communication or sensors channels

–> data - algorithm - data as an numbers

- understanding (понимание) -matching the utility function to the data processing (согласование функции полезности с обработкой данных)

-> data - algorithm - data as an concepts

- knowledge itself (**собственно познание**) - building a logical model of perceived (воспринимаемых) data and an algorithm for calculating a subset of the model state space on which utility function reaches maximum (алгоритм для вычисления подмножества пространства состояний модели, на котором функция полезности достигает максимума)

-> data – algorithm – (**алгоритм**) algorithm

The such aspects composition can be considered as way to solve main problem “data-algorithm-understanding” .

The essence of the task - “constructing an algorithm for extract a subset of the state space of models on which some utility (goal oriented) function achieves maximum .

We trust in formal explanation system .....

Why ?

Model flexibility: The interpretation method can work with flexible (different) type of formal models (random forests , deep neural networks…)

Explanation flexibility: There are several forms of explanation – simple linear formula, graphic, etc.

Representation flexibility: The explanation system should be able to use a different type of model representation as the model that previous being use for explained.

High level look at model-agnostic interpretability:

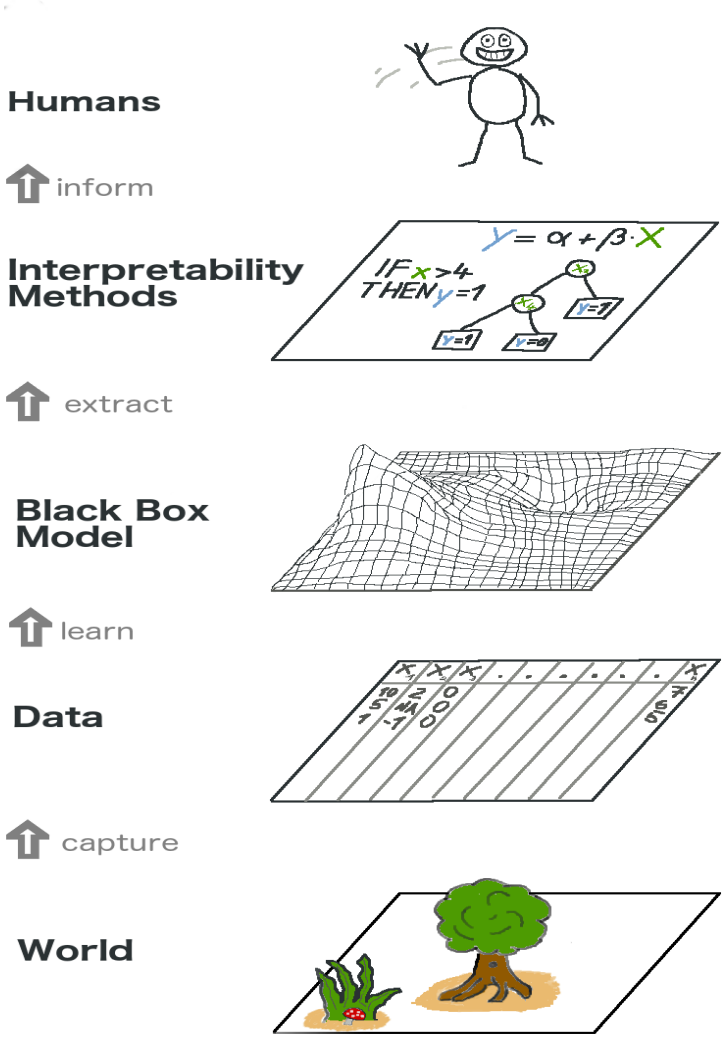
Humans - consumers of the explanations. Why we trust in …..

Interpretability Methods layer, which helps human deal with the opacity of machine learning models (how machine calculate explanations)

Black Box Model layer - algorithms using data from the real world to make predictions, find structures or invariants

The Data layer contains 'digital twins" anything from images, texts, tabular data and so on in order to make it process able for computers and also to store information.

The World layer contains everything that can be observed and is of interest.



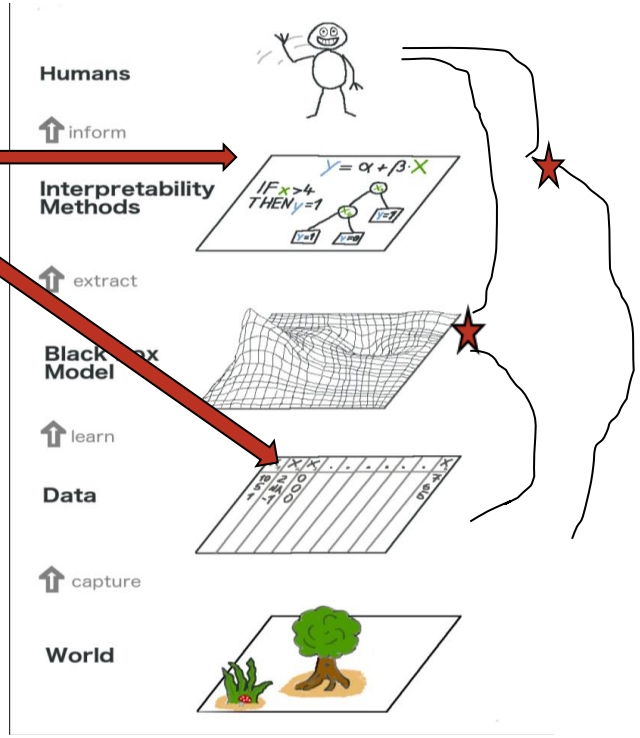
This multi-layered abstraction

We need to understand the differences in approaches between statisticians and machine learning approaches.

Statisticians deal with the Data layer, such as planning, estimation, predictions, skip the Black Box Model layer and go right to the Interpretability Methods layer.

Machine learning specialists also deal with the Data layer, train a black box machine learning mode and skip The Interpretability Methods layer, so Humans directly deal with the Black Box model predictions.

Interpretable machine learning merge (unite, join) the work of statisticians and machine learning specialists.



Фундаментальная проблема

- «Интеллектуальная» регуляризация решения «обратных задач» восприятия реальности

Одна и та же система имеет различные физические свойства в зависимости от имеющейся информации (в одном случае она способна совершить работу, в другом – нет)

- Мера информации оказывается согласованной с общефизическими понятиями энергии и энтропии
- Информация как объективное описание состояния системы наравне с ее физическими параметрами меняет ее свойства. Т.е. в зависимости от имеющейся информации о системе систему можно или нельзя использовать для совершения работы. (в одном случае система способна совершить работу, в другом – нет)

Фундаментальный вопрос современной науки: существует ли «физика» .... мышления

Так, чтение или письмо – есть тренировка для головного мозга, в особенности если при этом вы узнаёте или выражаете нечто новое.

- Изменение сознания в процессе мышления приводит к изменениям в физическом теле интеллектуального субъекта.
- «Машина» обретет способность мыслить», если приобретет свойства «процессора управляемого данными»

Цель семинара

Использование понятия «информации» в рамках как имплицитных, так и эксплицитных форм знаний о природных закономерностях, учитывая что 1) имплицитное знание не имеет символьной формы описания, поэтому не отделимо от субъекта – непосредственным носителем знаний, 2) эксплицитные знания могут быть выражены в форме математических законов физики.

Точное значения используемых понятий следующее:

- Имплицитный (от латинского слова implicitum) значит «невыраженный», «подразумеваемый», «неразвернутый» т.е.– «скрытый».

- Эксплицитный (от латинского слова explicitum) – значит «явно выраженный», «развернутый», т.е. – «явный».

- **Надо понять** - как нужно рассуждать об информации или об интеллекте, чтобы эти рассуждения имели научный смысл т.е. объяснялись с помощью введенных ранее понятий?.