

Foundations of

Data Science and data-driven engineering

$$P(Y|X)$$

ADA511 Data Science and data-driven engineering

Course description for academic year 2023/2024

...

Compulsory learning activities

One obligatory assignment. In order to take the examination, the assignment must be approved.

Assessment

1. Group project work with final report, counts for 50% of the final grade
2. Oral examination based on the final report, counts for 50% of the final grade.

Both parts must be passed to obtain a final grade.

Grade scale A-F, where F is fail.



This can be the starting point for your thesis project 😊

Background check: Courses

DAT158

Machine Learning Engineering and
Advanced Algorithms

INF264

Introduction to Machine Learning

DAT255

Deep learning engineering

INF265

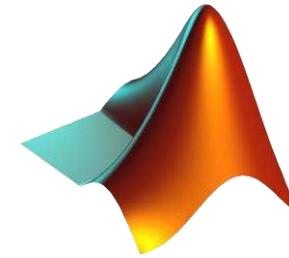
Deep Learning

?

Background check: ML Frameworks



TensorFlow



dmlc
XGBoost



K Keras

Machine learning topics

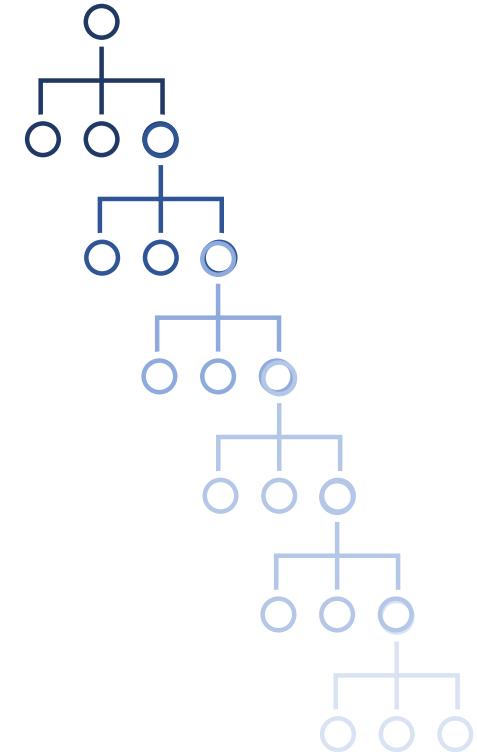
(1) Decision trees / random forests

Will focus on:

- Decision theory aspects of creating trees
- How the workings of tree models affect predictions
- Evaluating information contained in a tree model

Will not focus on:

- Optimisation and hyperparameter tuning



Machine learning topics

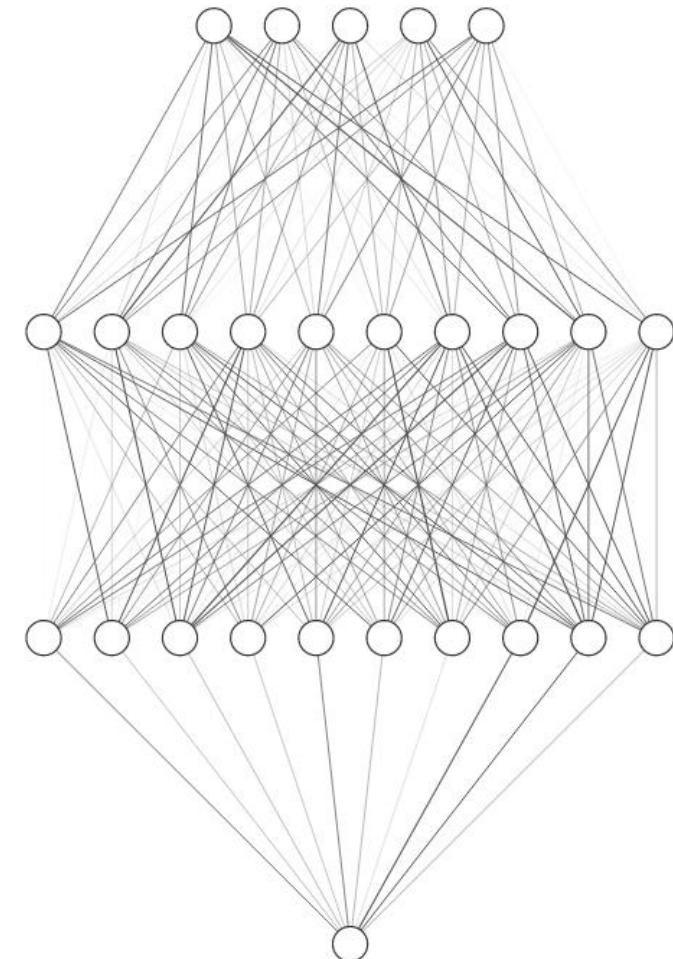
(2) Neural networks

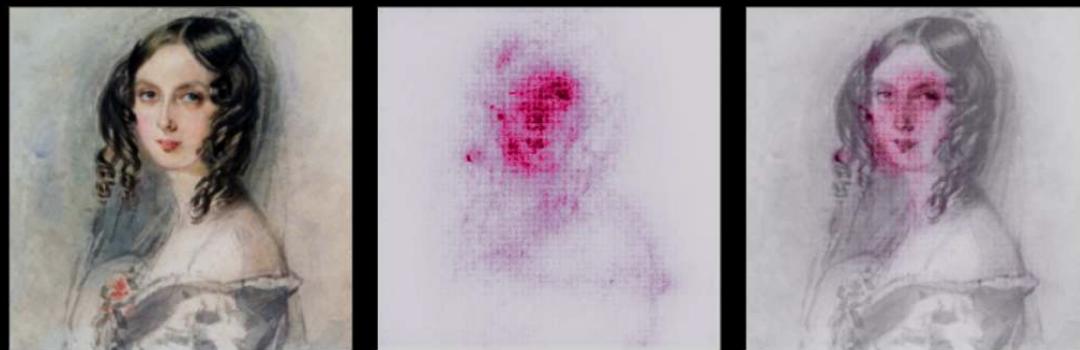
Will focus on:

- Interpretation of predictions
- Extensions to solve “ill-posed” (= real-world) problems

Will not focus on:

- The most recent architectures
- Optimisation and hyperparameter tuning



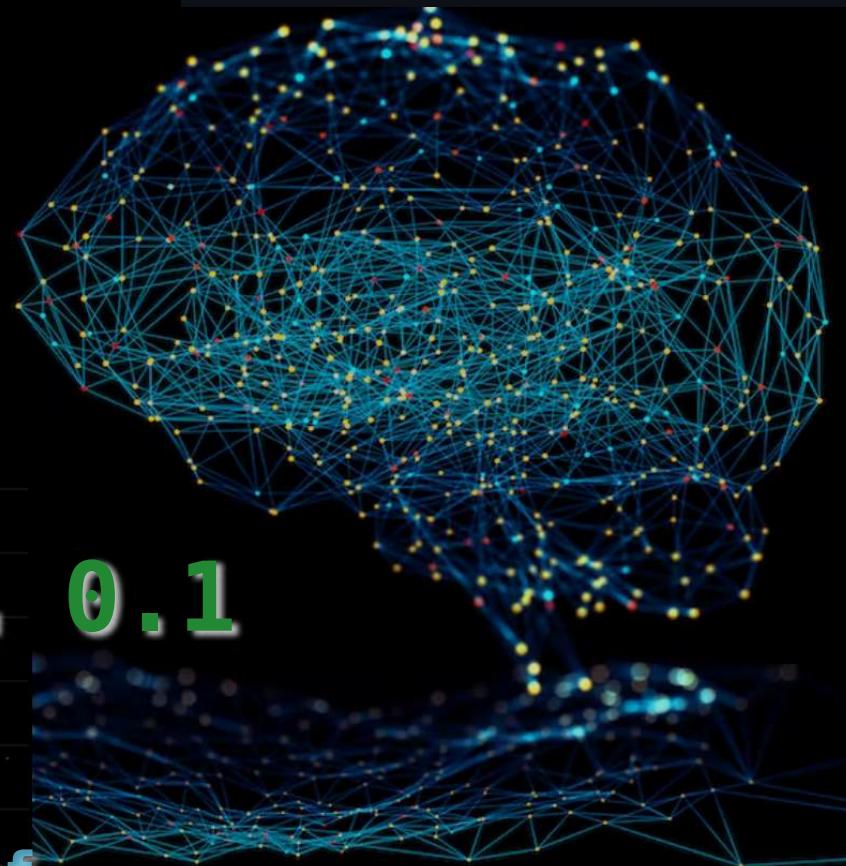
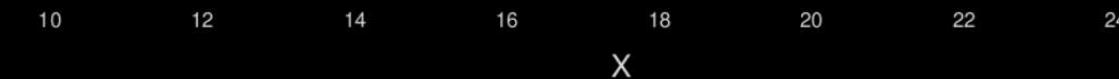


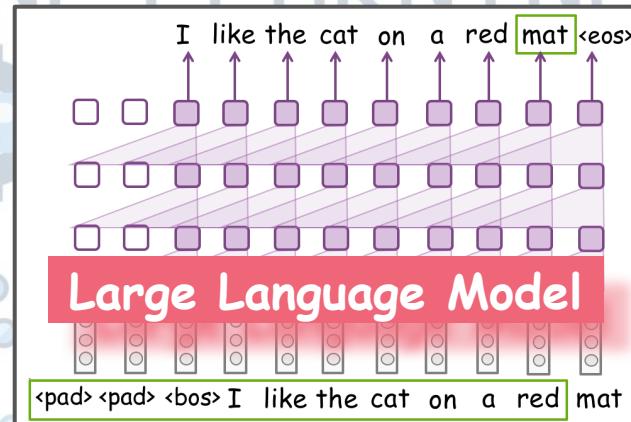
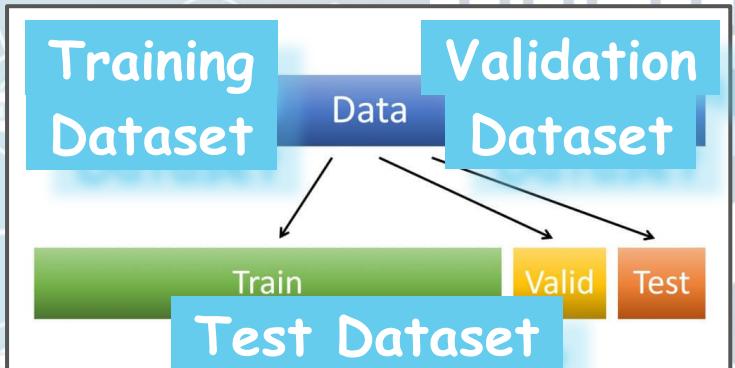
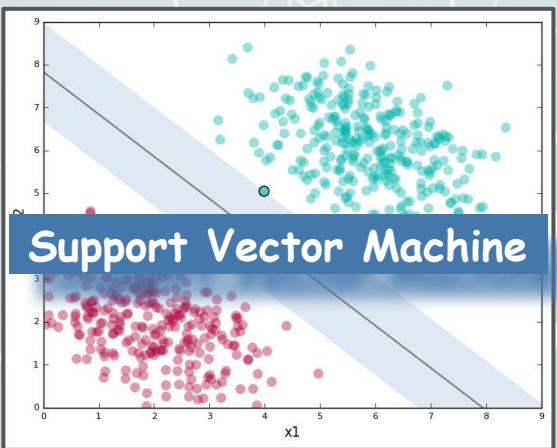
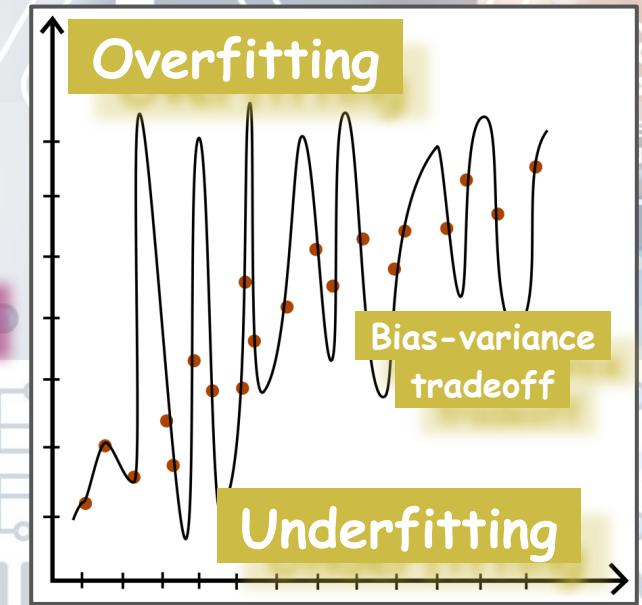
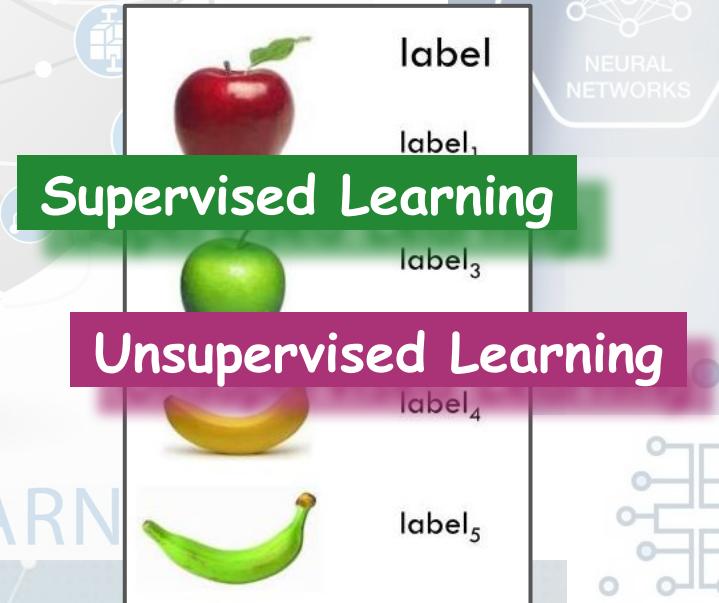
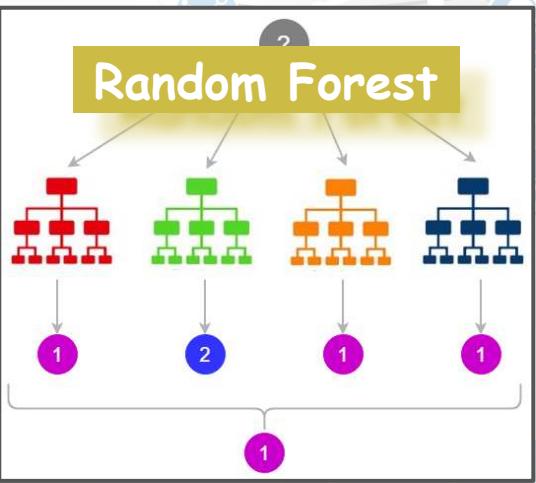
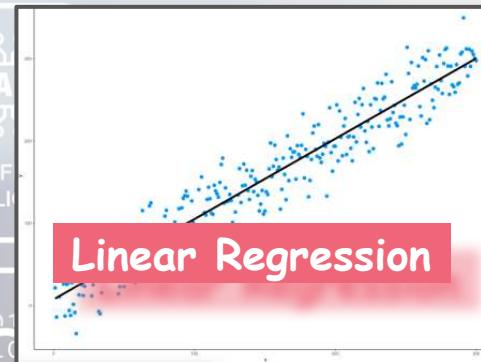
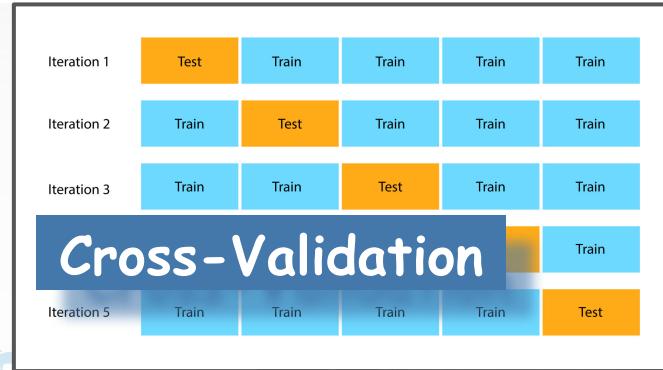
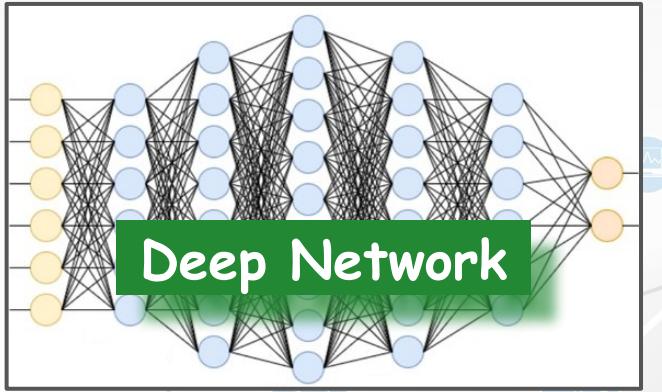
ADA 511

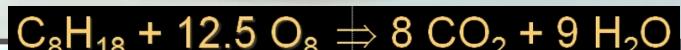
Foundations of

Data Science and data-driven engineering

$P(Y|X)$







Isooctane

Oxygen

Carbon dioxide

Water



SEDAN



HATCHBACK



CUV



SUV



MICRO



CABRIOLET



SUPERCAR



COUPE



VAN



MINIVAN



CAMPERVAN



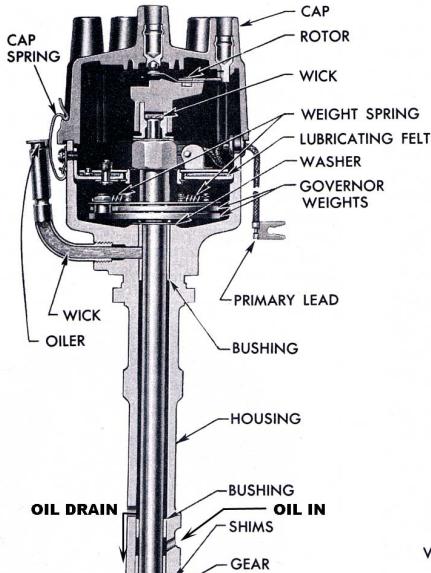
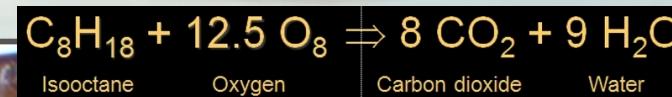
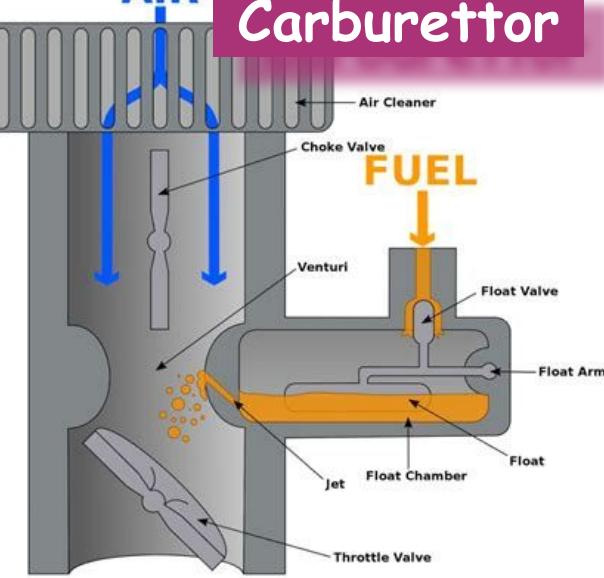
TRUCK



BIGTRUCK

AIR

Carburettor



Distributor

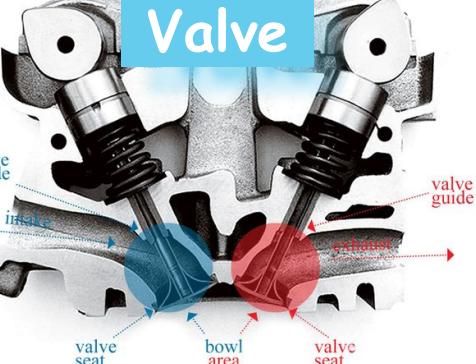


Spark plug

Fuel



Octanes



Silencer (muffler)



Engine oil

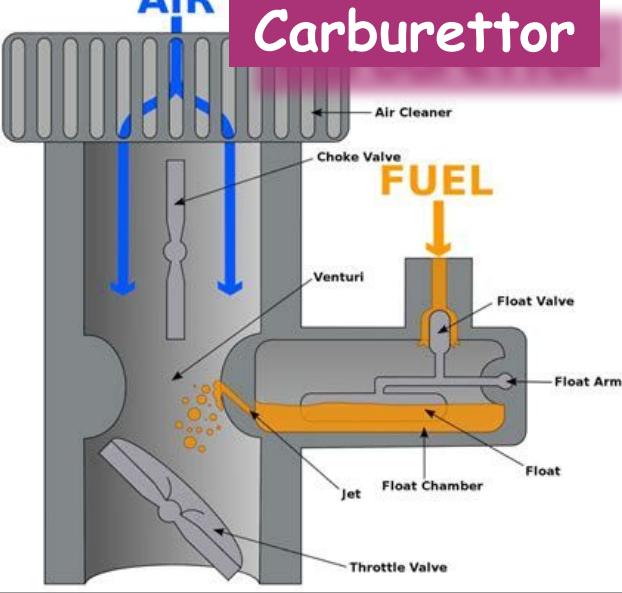
CAMPERVAN

TRUCK

Green emissions

AIR

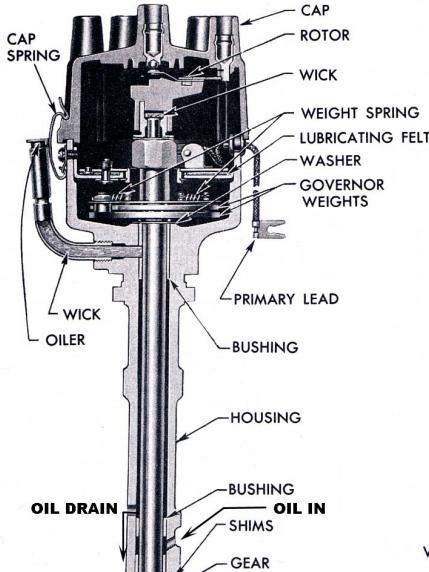
Carburettor



Fuel



Octanes



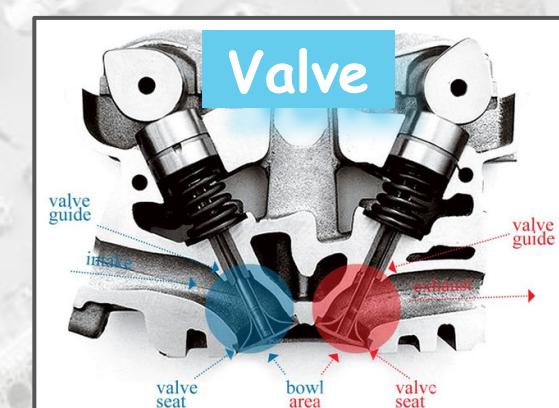
Distributor



Spark plug



Engine oil



Valve

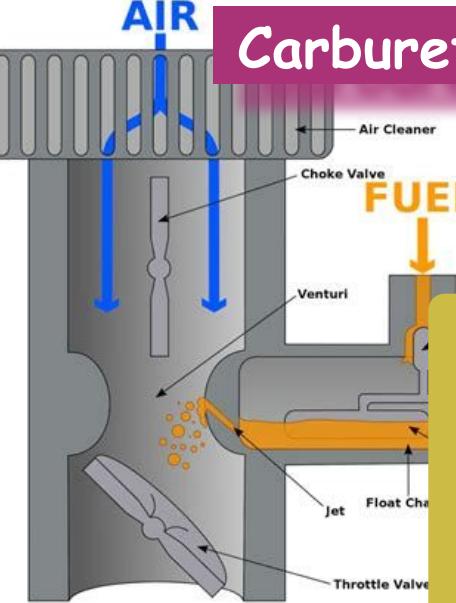
Silencer (muffler)



Green emissions

AIR

Carburettor



Fuel

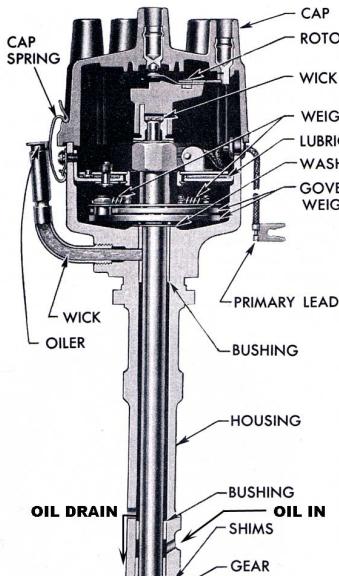
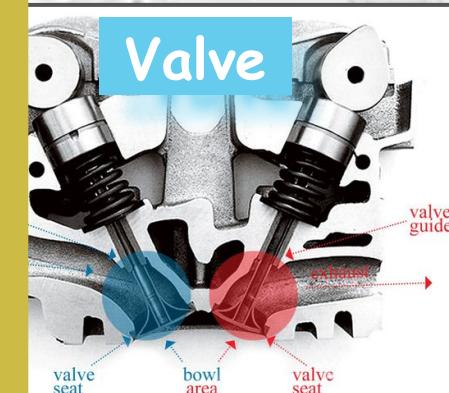


Octanes

car mechanic



Valve



Distributor



Spark plug



Engine oil

Silencer (muffler)



Green emissions

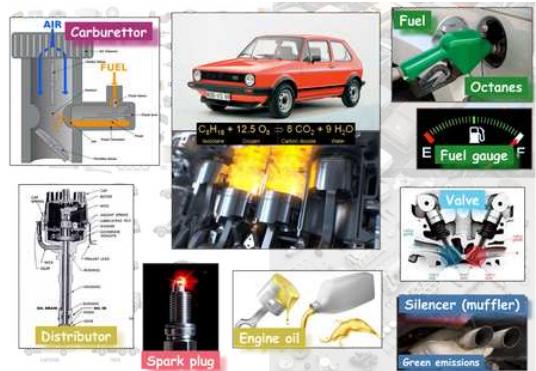
🔧 car mechanic



cartwright



car mechanic



cartwright



car mechanic



EV mechanic



cartwright



car mechanic



EV mechanic



 cartwright



 car mechanic



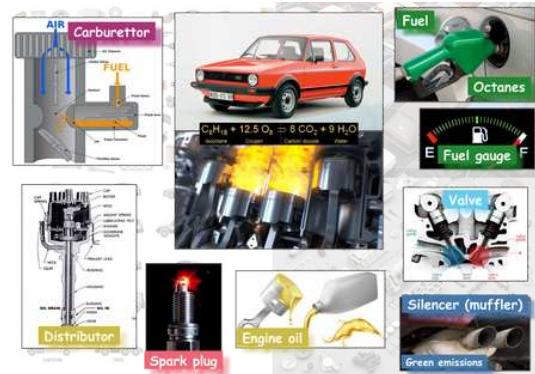
 EV mechanic



 cartwright



 car mechanic



 EV mechanic



automotive engineer



car mechanic



automotive engineer



car mechanic



automotive engineer



 knowledge

car mechanic



Q *specific*

automotive engineer



knowledge

i i general

car mechanic



🔍 *specific*

🔨 *practical*

automotive engineer



🎓 **knowledge**

💡 *general*

Ἁ *principled*

car mechanic



Q *specific*

🔧 *practical*

☰ *procedural*

automotive engineer



🎓 **knowledge**

蔀 *general*

Ἁ *principled*

⌚ *exploratory*

car mechanic



🔍 *specific*

🔨 *practical*

🛠️ *procedural*

⚓ *technology-dependent*

...

automotive engineer



🎓 *knowledge*

💡 *general*

Ἁ *principled*

🧭 *exploratory*

🐦 *technology-independent*

...

car mechanic



🔍 *specific*

🔨 *practical*

🛠 *procedural*

⚓ *technology-dependent*

...

automotive engineer



🎓 **knowledge**

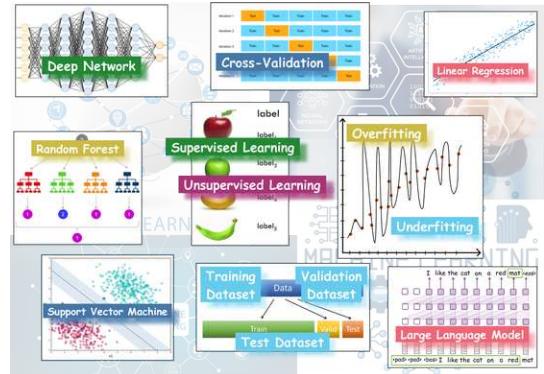
💡 *general*

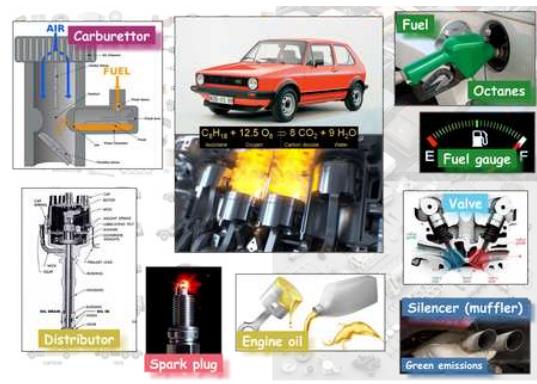
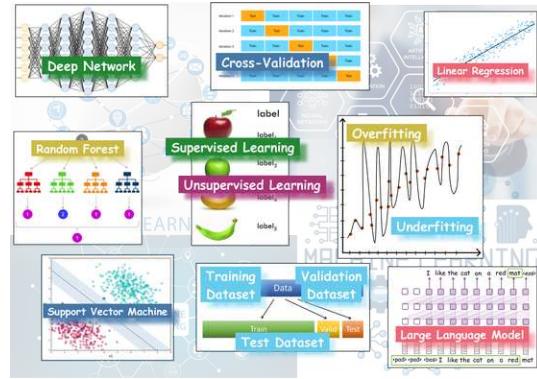
Ἁ *principled*

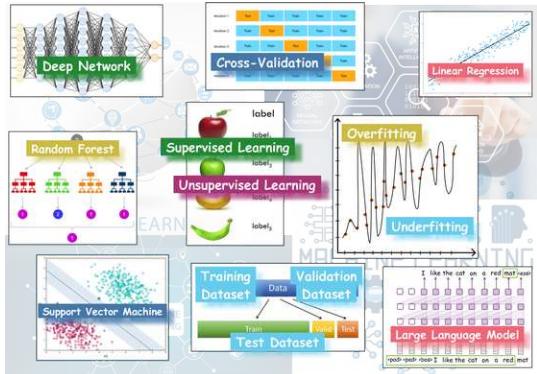
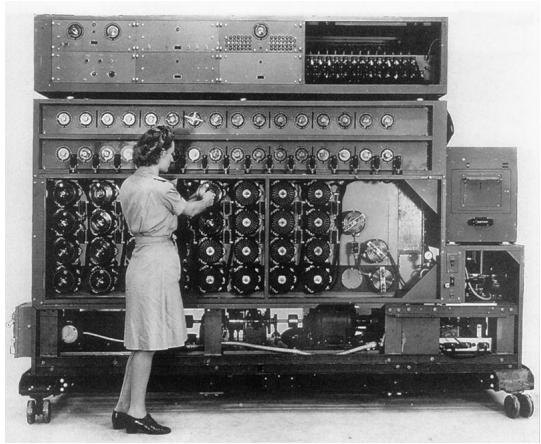
🧭 *exploratory*

🐦 *technology-independent*

...

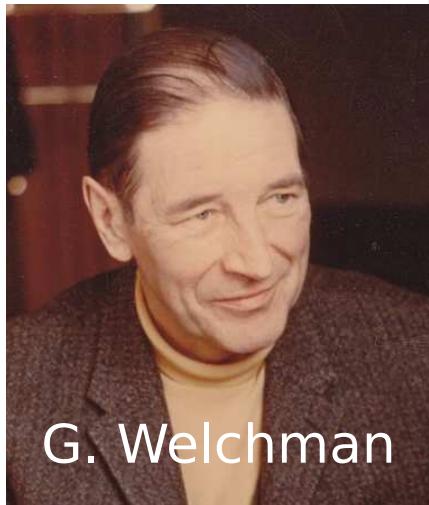








Alan Turing



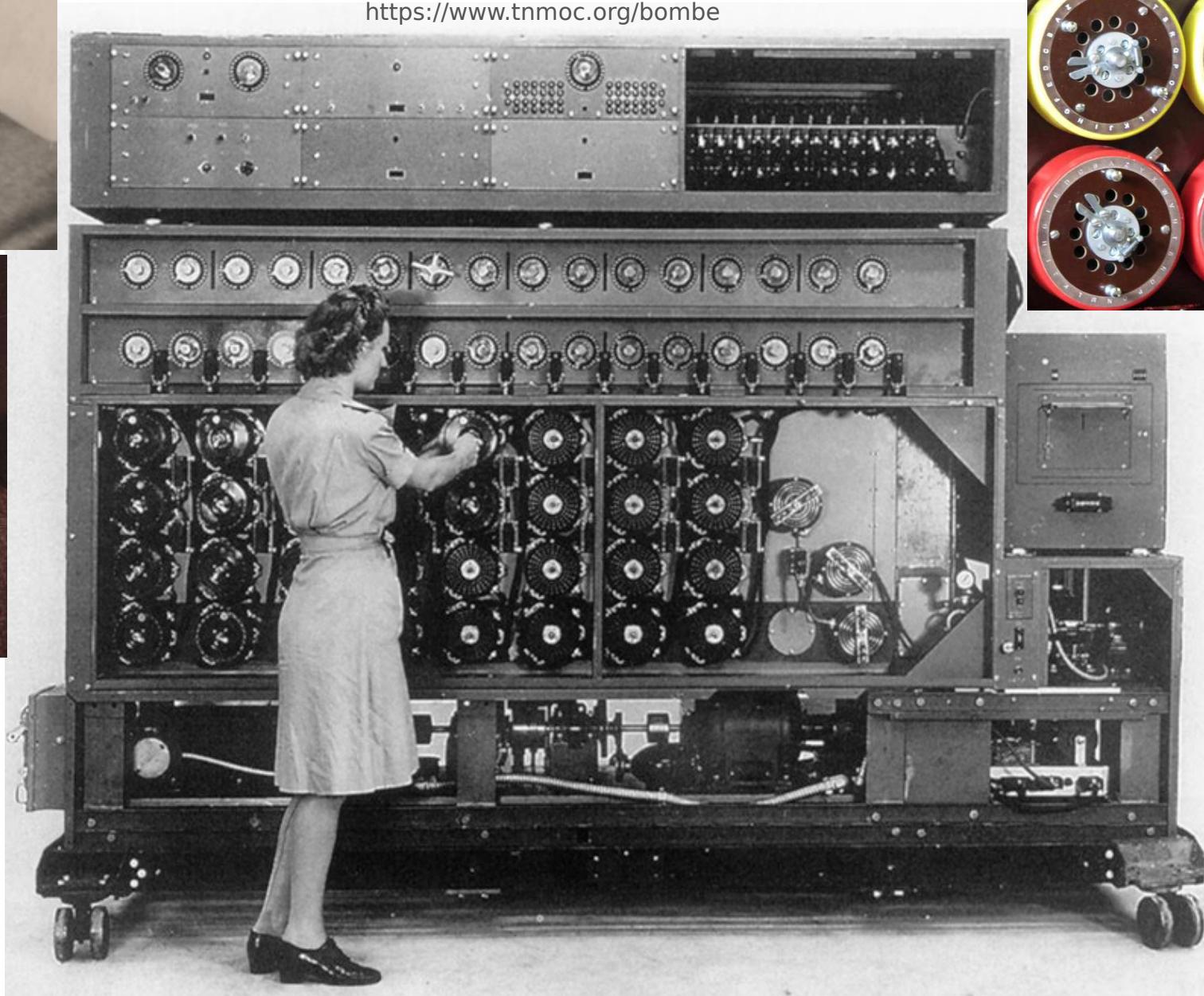
G. Welchman

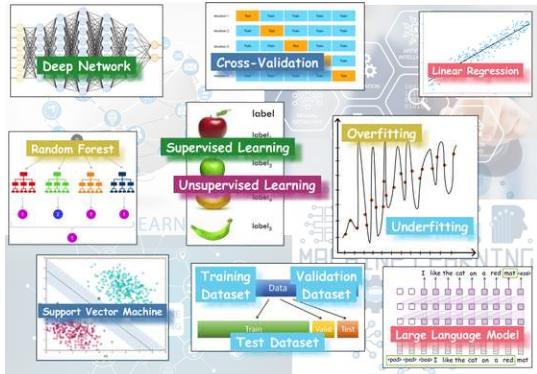
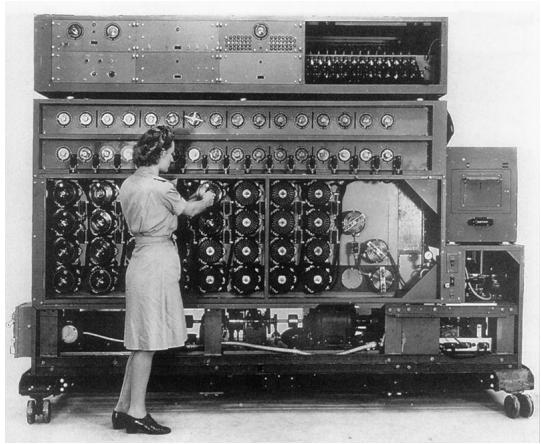
The Bombe a.k.a. Agnes electromechanical computer

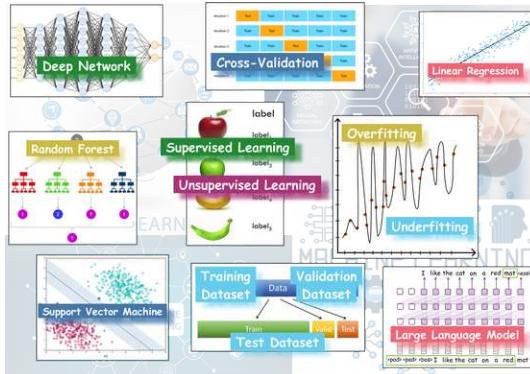
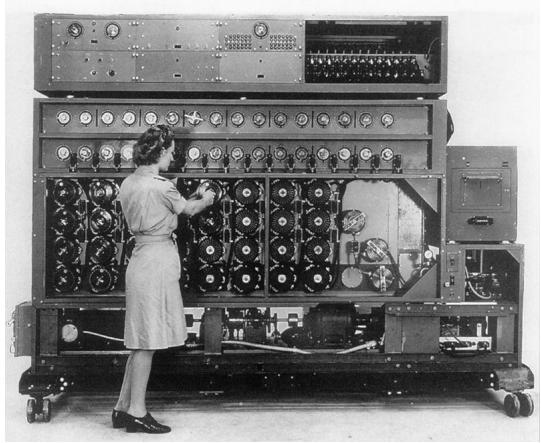
Decoded encrypted Nazi communications during World War II
(efficiency rate: 98.8%)

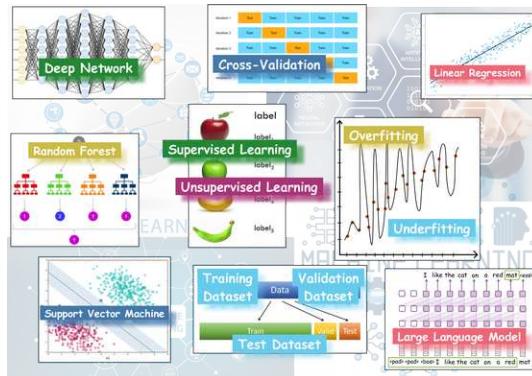
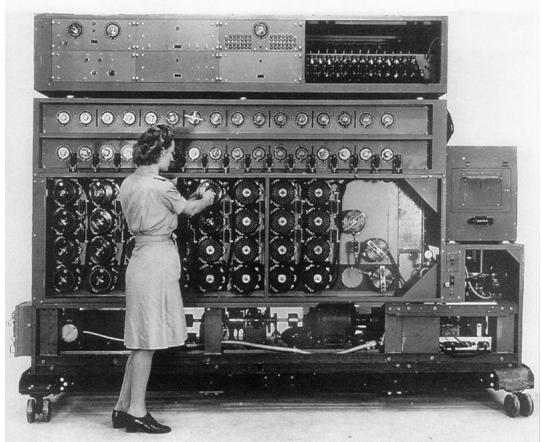
<https://www.britannica.com/topic/Bombe>

<https://www.tnmoc.org/bombe>









cartwright

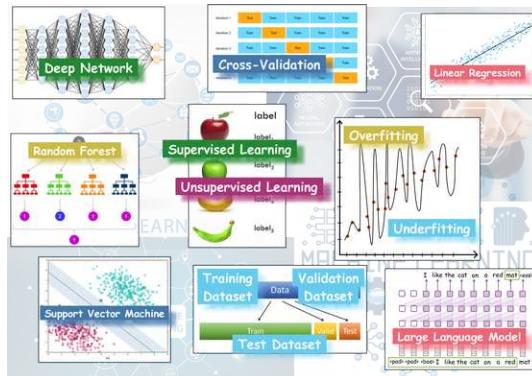
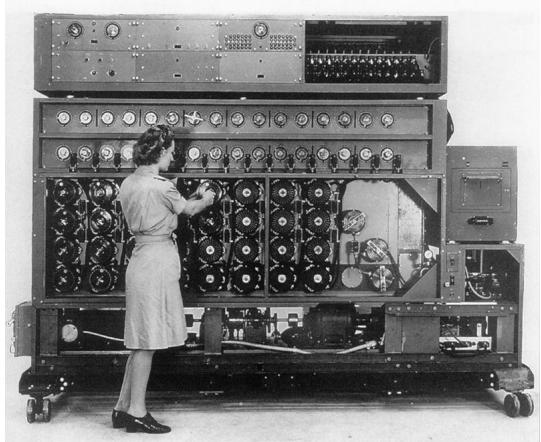


car mechanic



EV mechanic





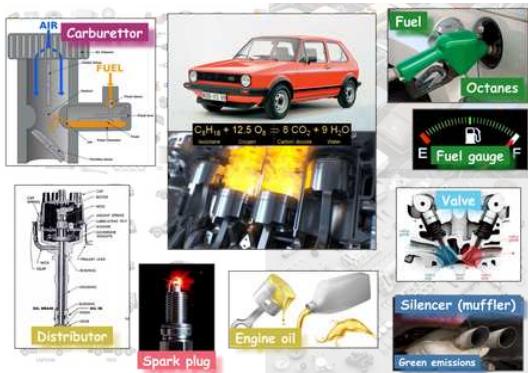
“data mechanic”



cartwright

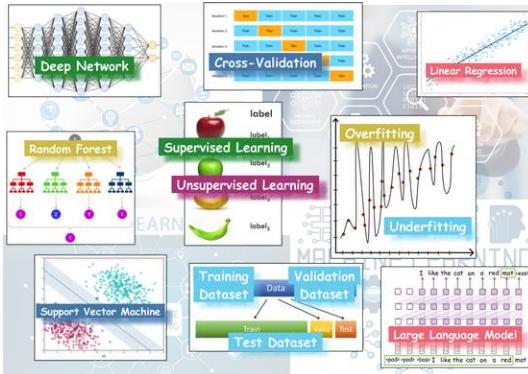
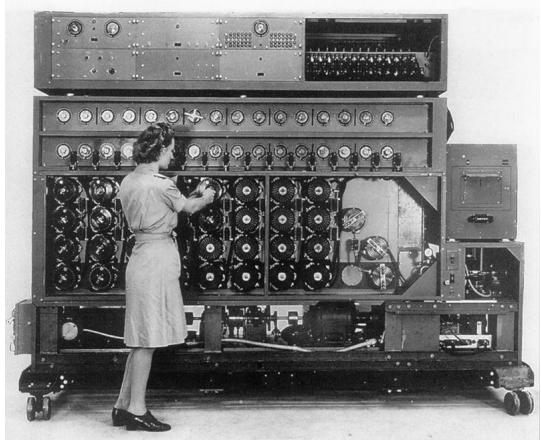


car mechanic



EV mechanic





“data mechanic”



cartwright

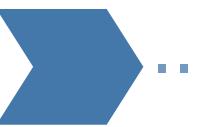
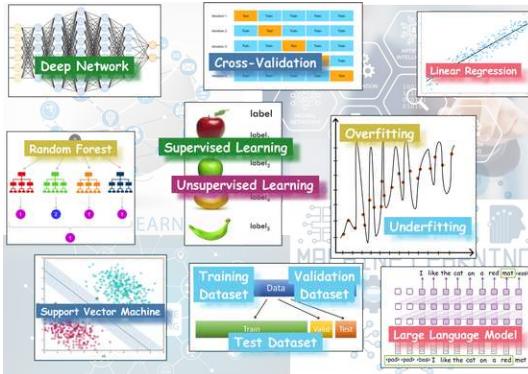
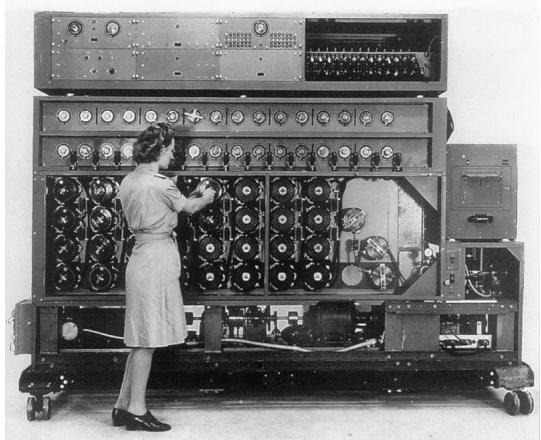


car mechanic



EV mechanic





“data mechanic”

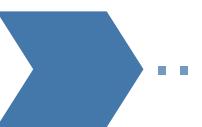
cartwright

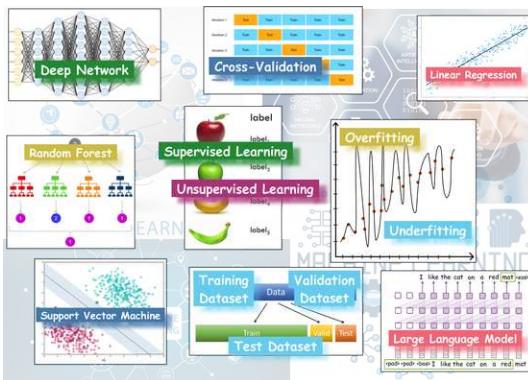
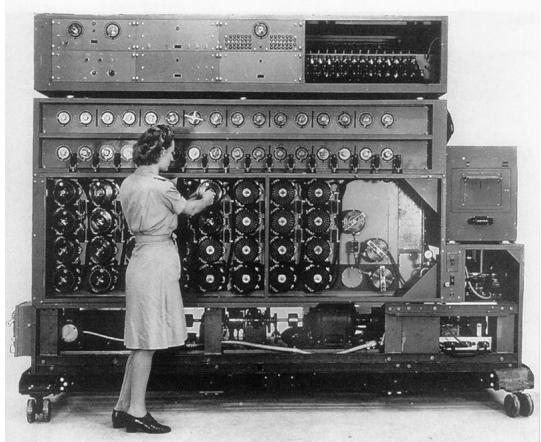


car mechanic



EV mechanic

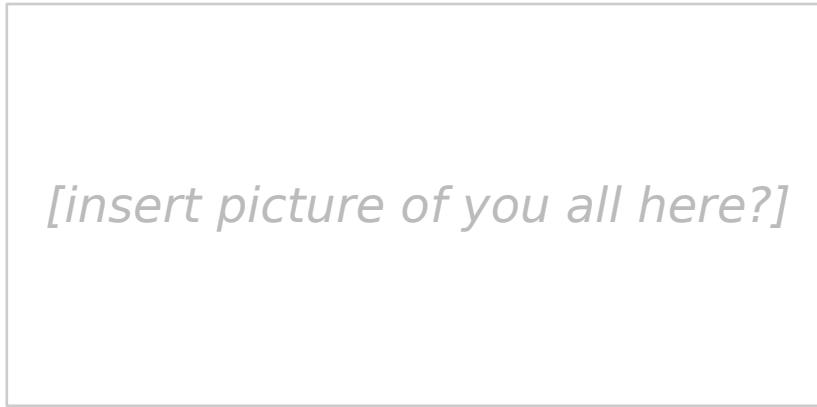




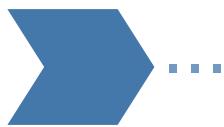
“data mechanic”



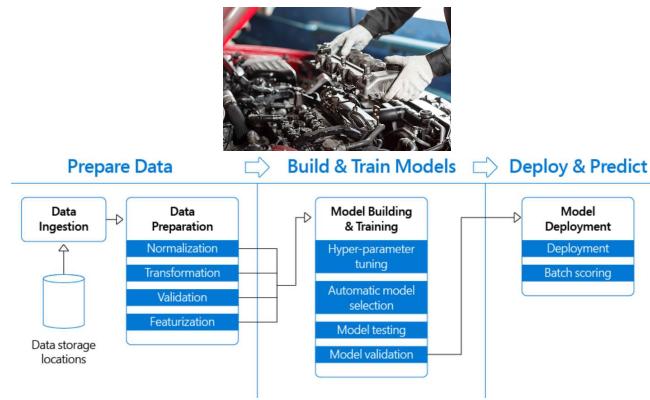
“data engineer”



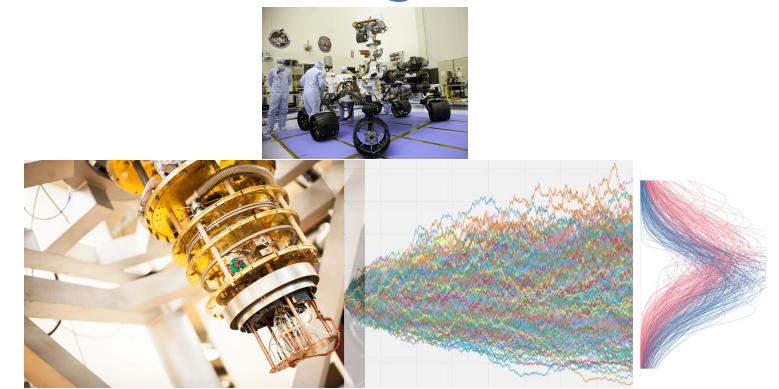
[insert picture of you all here?]



“data mechanic”



“data engineer”



knowledge

🔍 *specific*

👀 *general*

🔨 *practical*

📐 *principled*

🛠️ *procedural*

🧭 *exploratory*

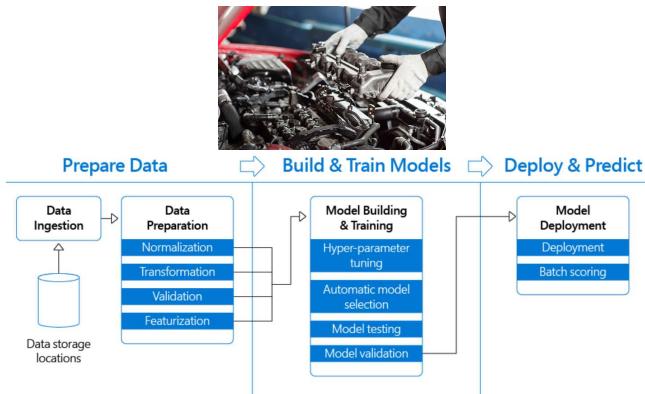
⚓ *technology-dependent*

🐦 *technology-independent*

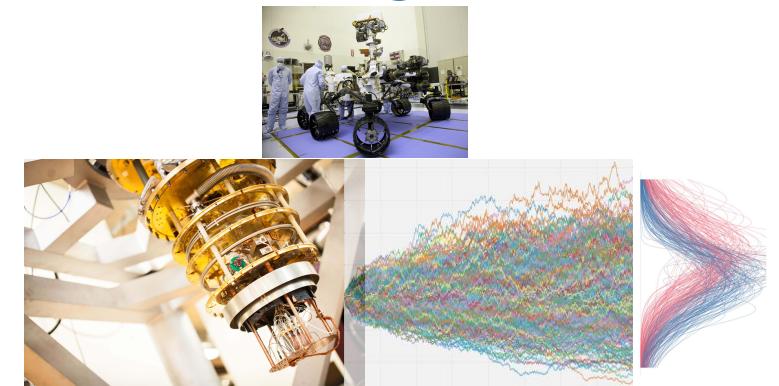
...

...

“data mechanic”



“data engineer”



knowledge

🔍 *specific*

👀 *general*

🔨 *practical*

📐 *principled*

🛠️ *procedural*

🧭 *exploratory*

⚓ *technology-dependent*

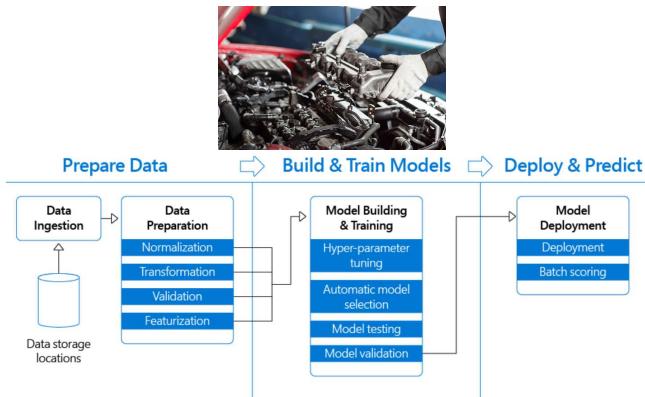
🐦 *technology-independent*

...

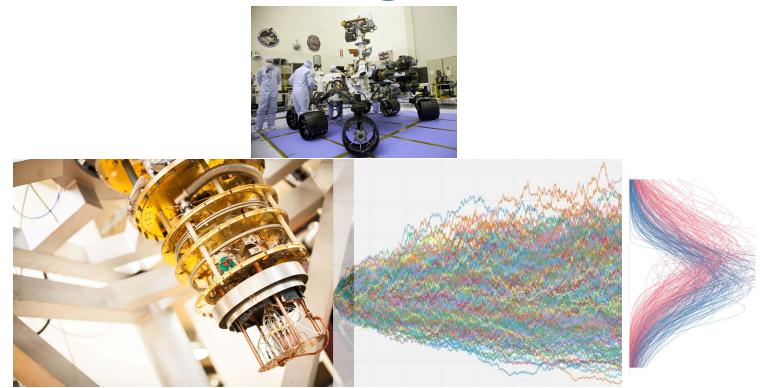
...

THIS IS THE BEST! 😍
don't listen to what he's saying

“data mechanic”



“data engineer”



knowledge

🔍 specific

👀 general

🔨 practical

📐 principled

🛠️ procedural

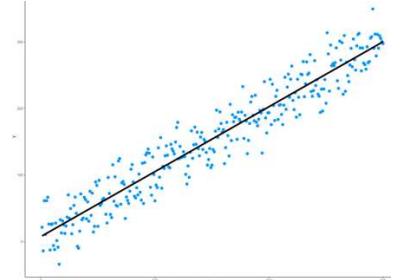
🧭 exploratory

⚓ technology-dependent

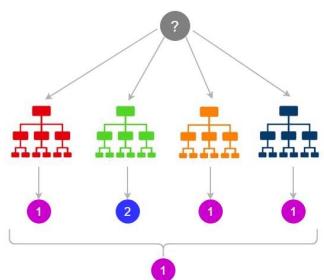
🐦 technology-independent

...

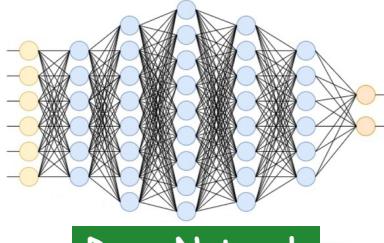
...



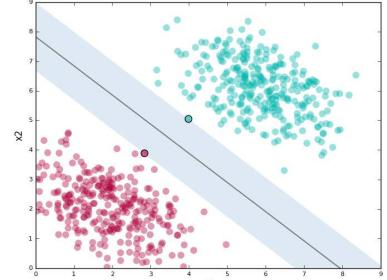
Linear Regression



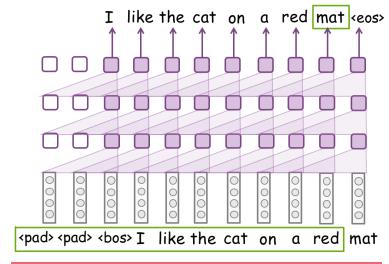
Random Forest



Deep Network



Support Vector Machine

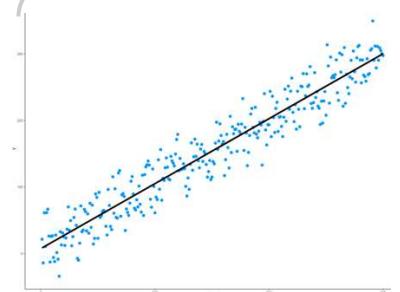
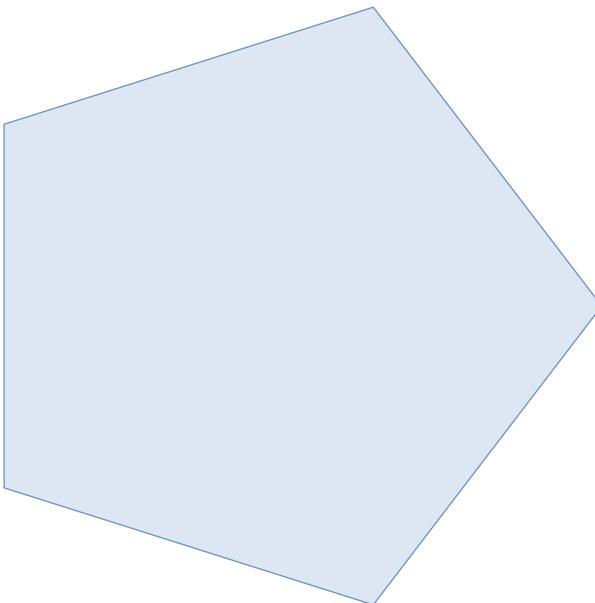


Large Language Model

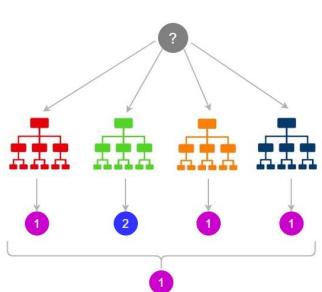


...

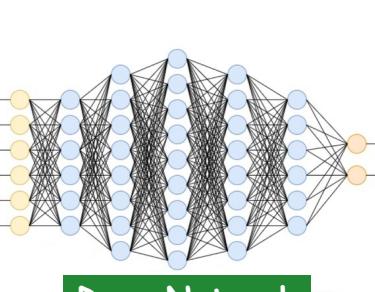
Optimal Predictor Machine



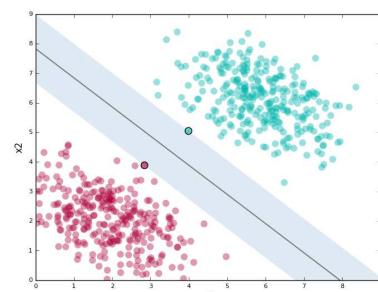
Linear Regression



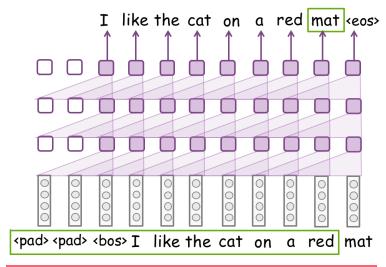
Random Forest



Deep Network



Support Vector Machine



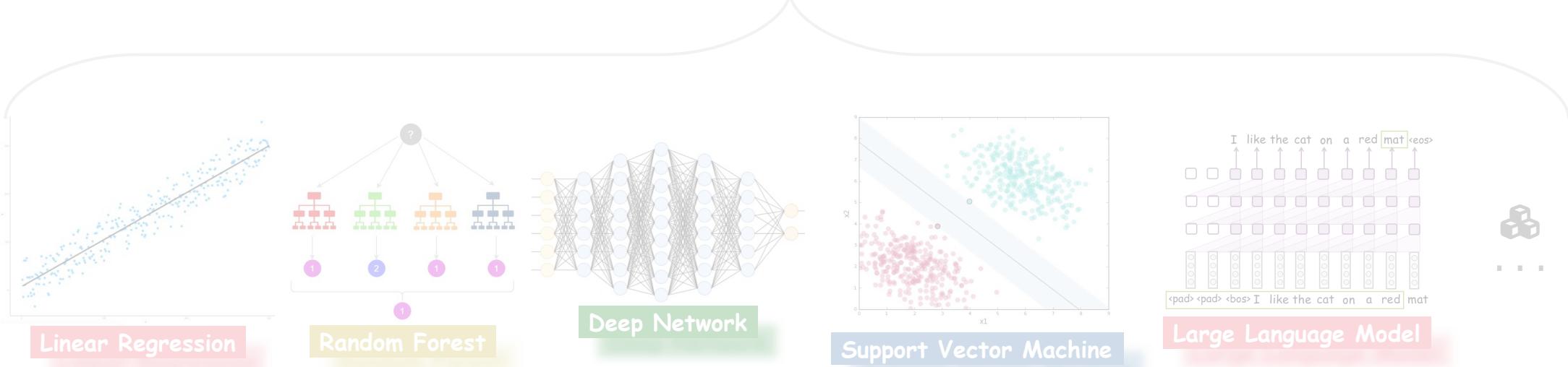
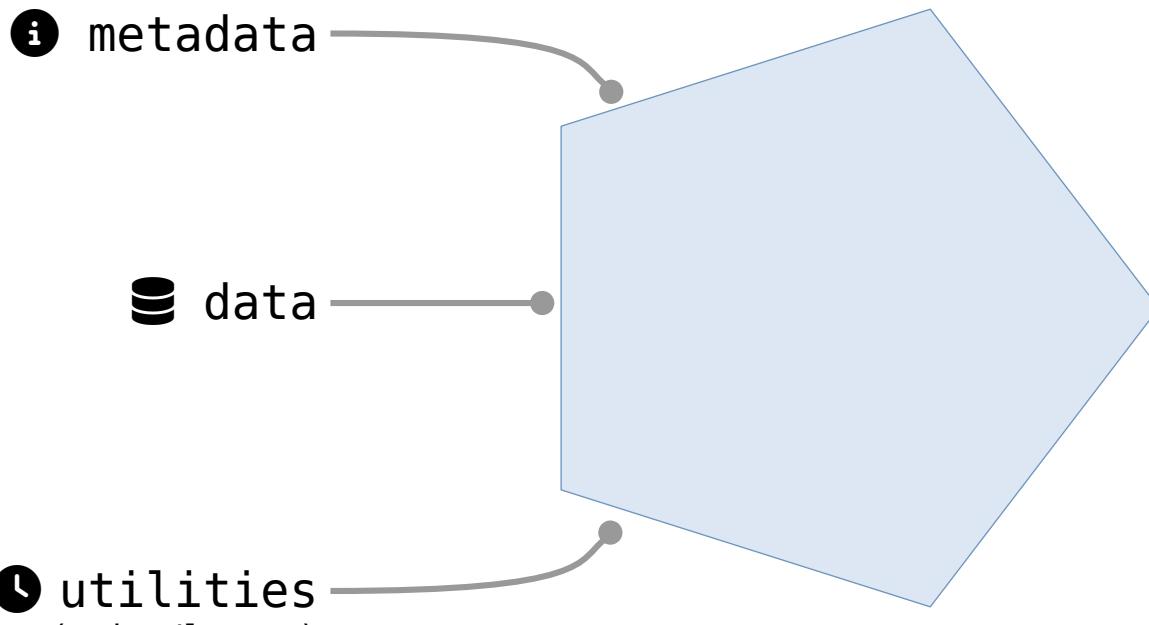
Large Language Model



...

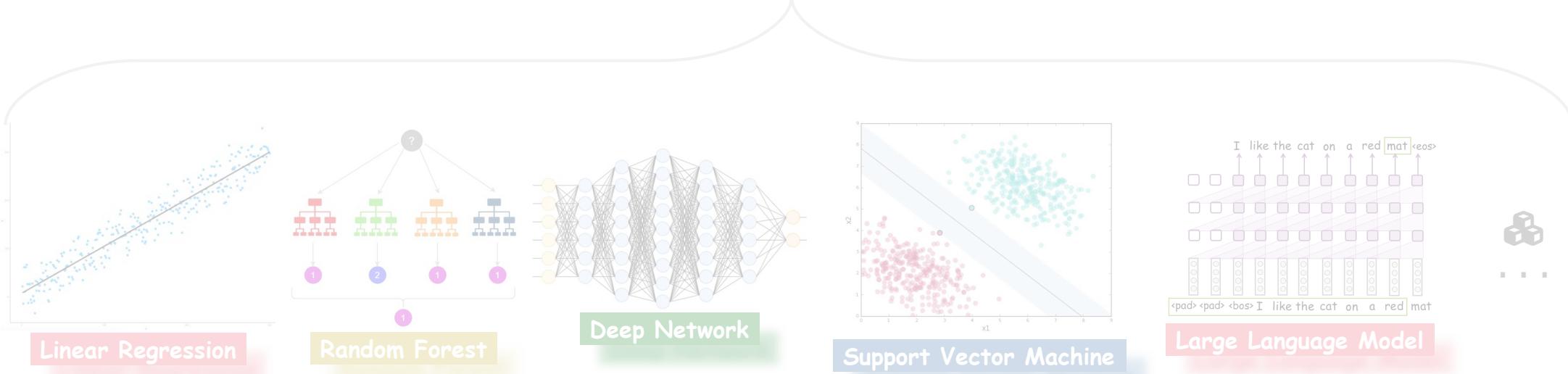
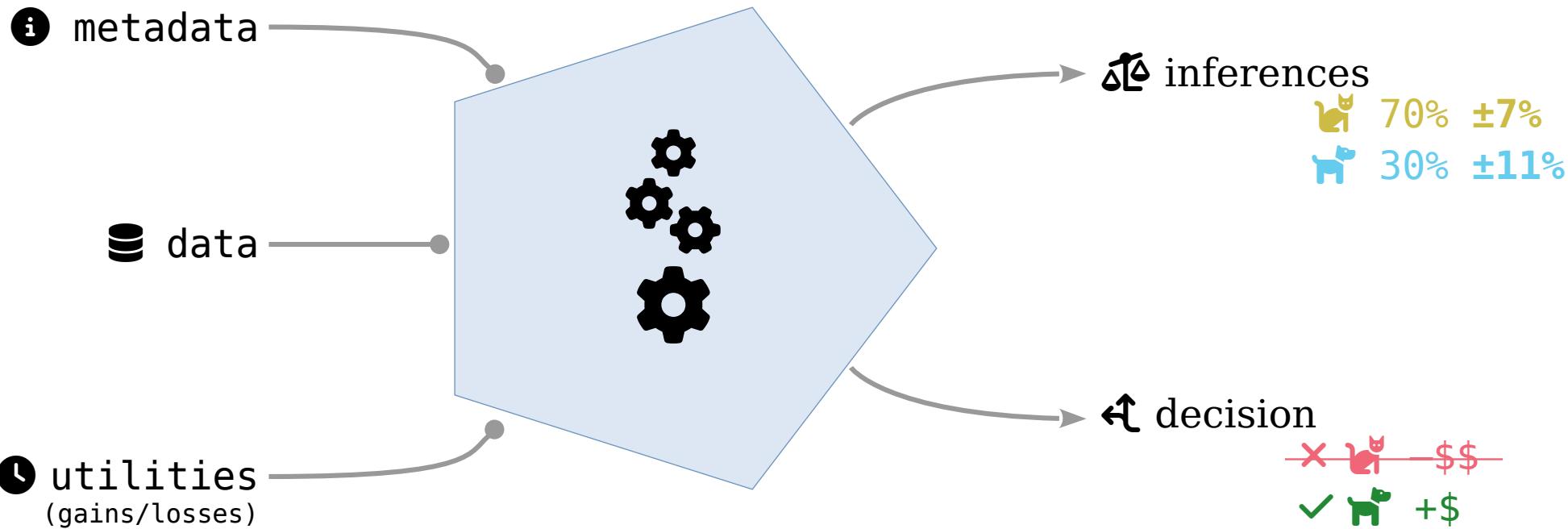
Optimal Predictor Machine

problem-dependent

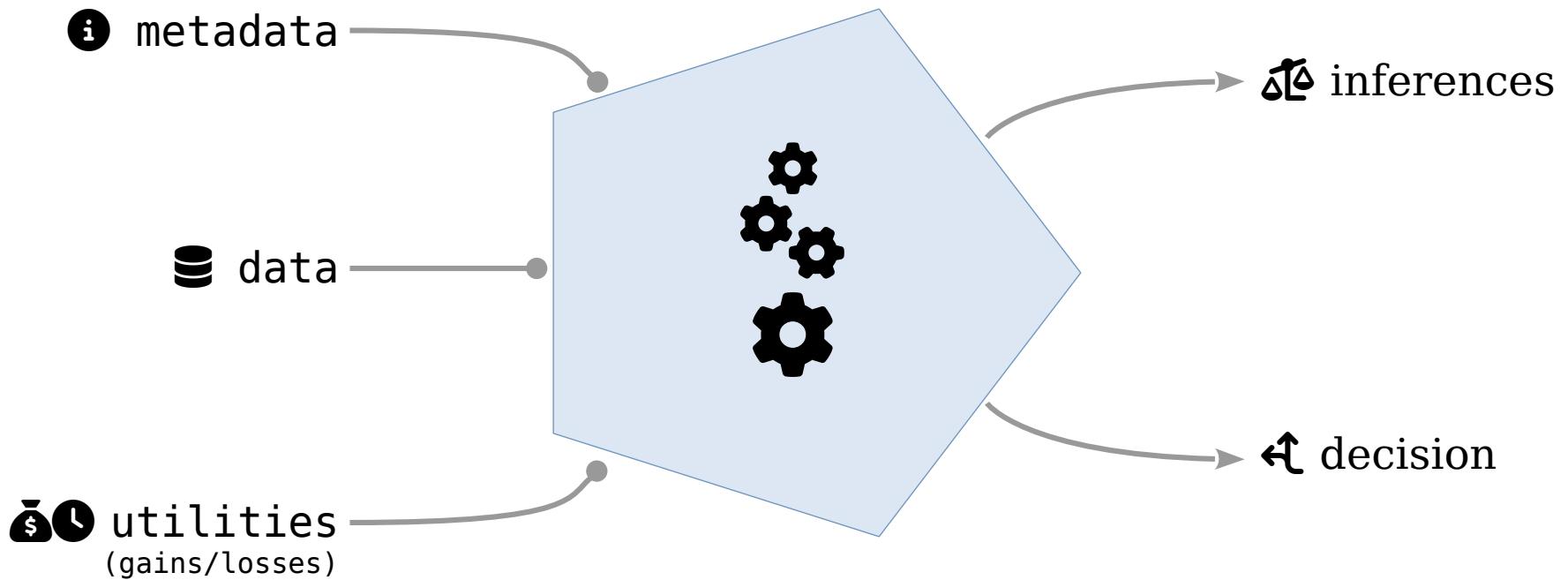


Optimal Predictor Machine

problem-dependent

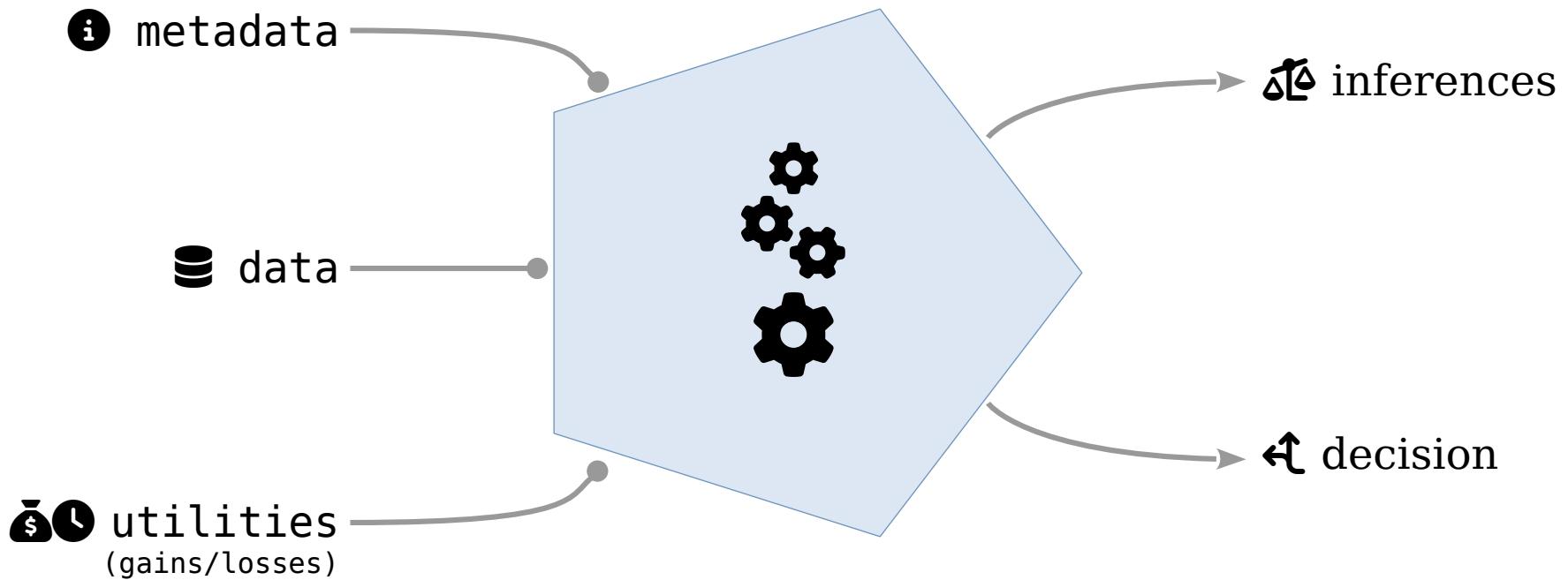


Optimal Predictor Machine



✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!

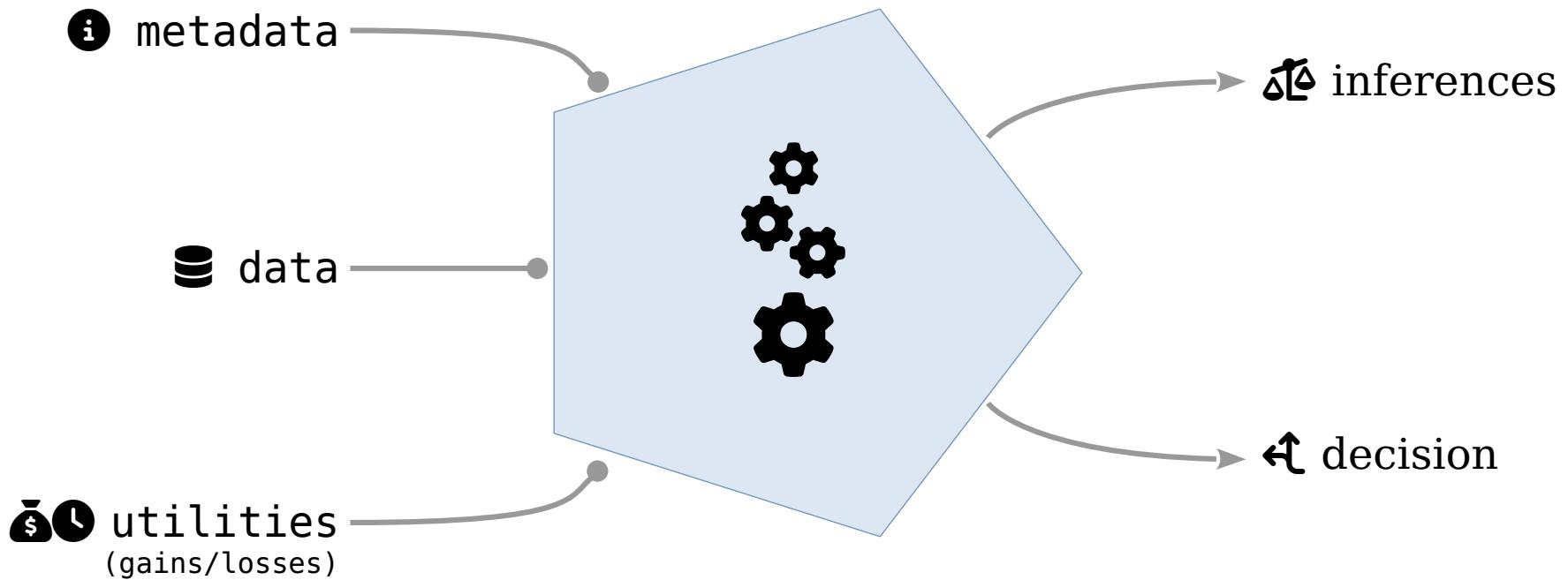
Optimal Predictor Machine



✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!



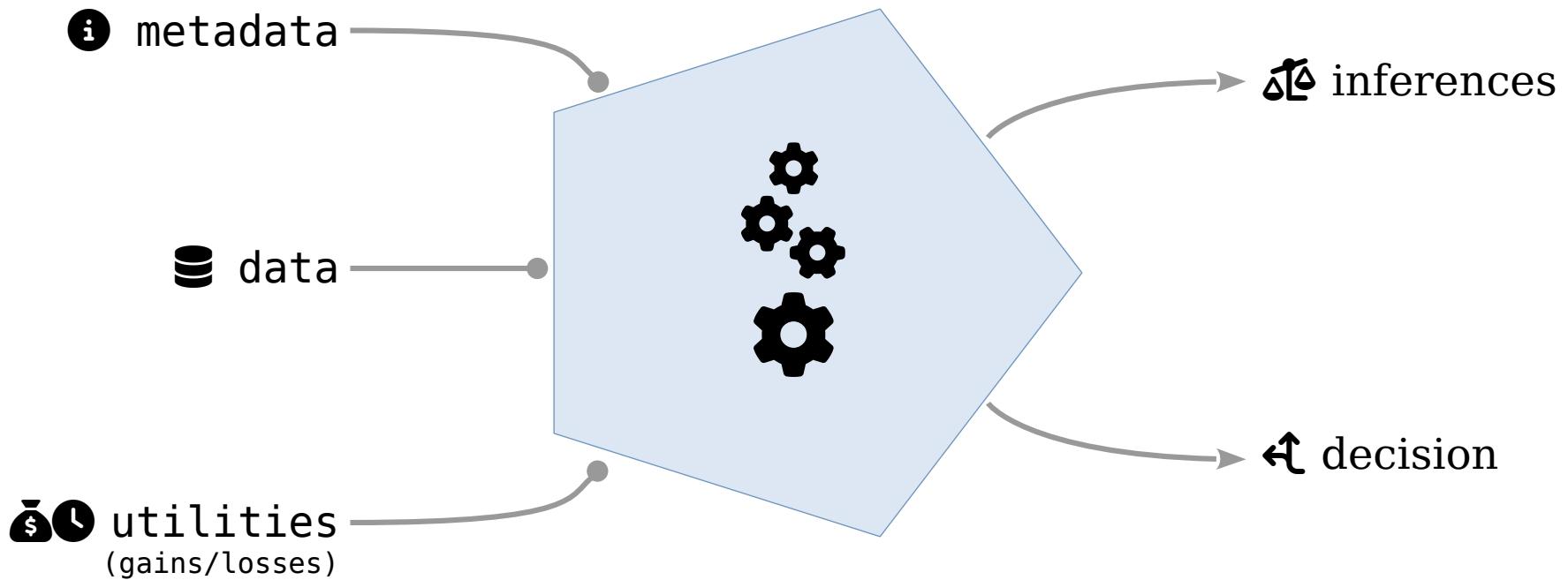
Optimal Predictor Machine



- ✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!
- ✓ mathematically guaranteed maximally-informative output



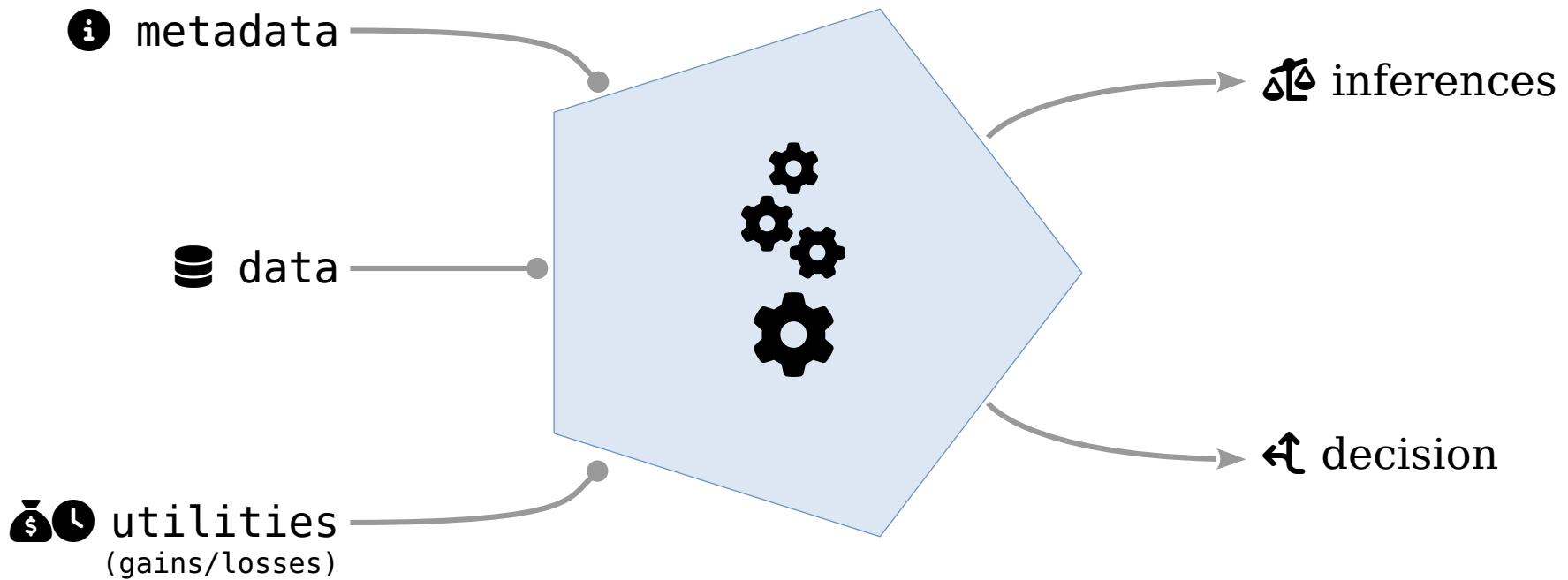
Optimal Predictor Machine



- ✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!
- ✓ mathematically guaranteed maximally-informative output
- ✓ learning is built-in



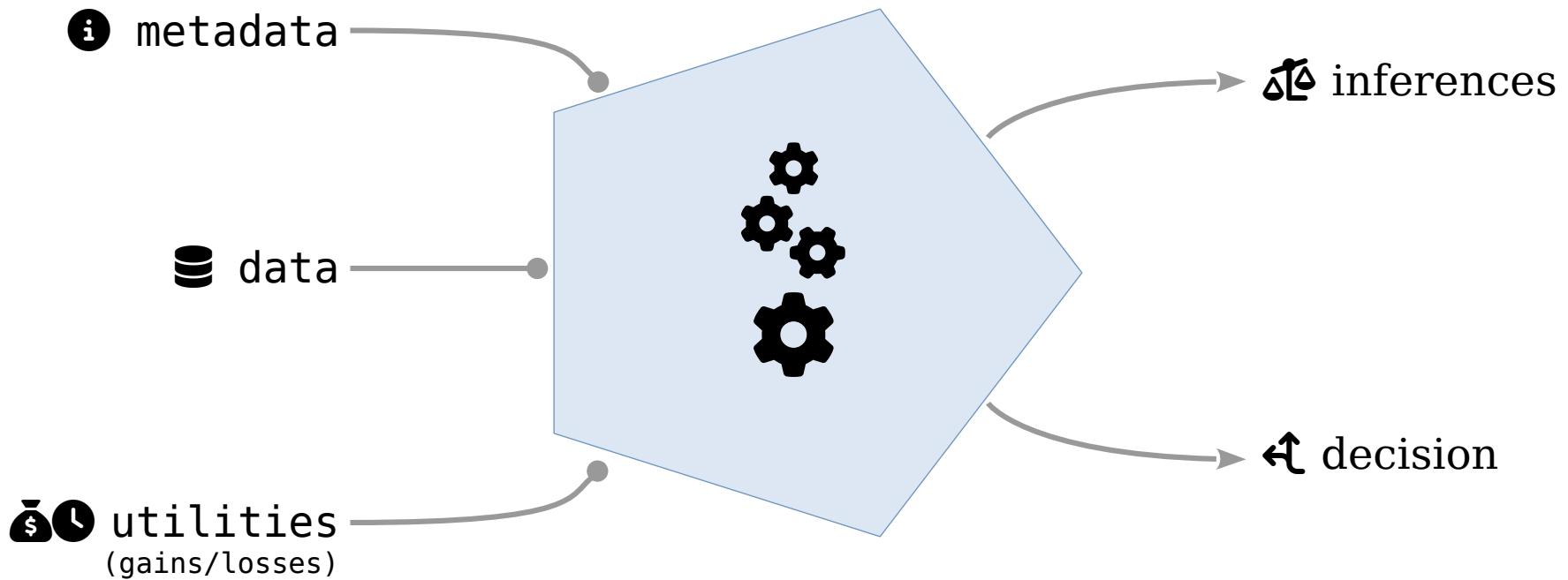
Optimal Predictor Machine



- ✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!
- ✓ mathematically guaranteed maximally-informative output
- ✓ learning is built-in
- ✓ no need to handle “un/supervised” or “discriminative/generative” differently



Optimal Predictor Machine



- ✓ mathematically guaranteed **optimal** outcome, by design → unbeatable!
- ✓ mathematically guaranteed maximally-informative output
- ✓ learning is built-in
- ✓ no need to handle “un/supervised” or “discriminative/generative” differently
- ✓ no need for data splits, cross-validation, under/overfit... (all implicit in how it works)





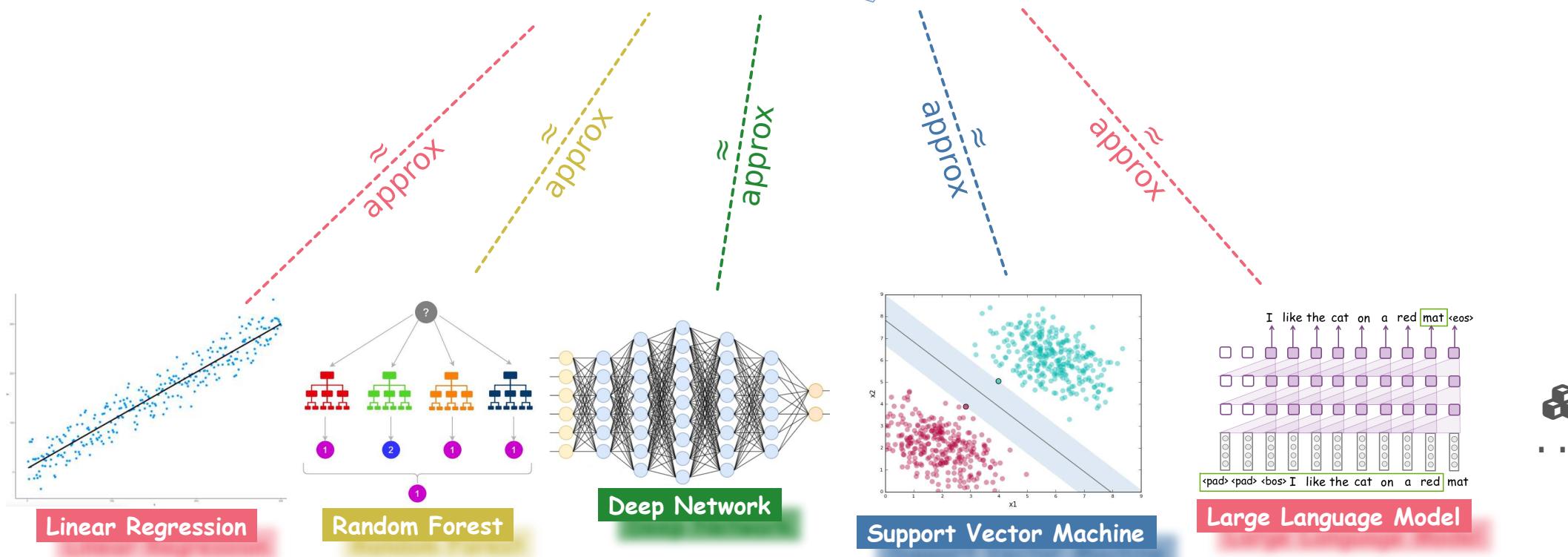
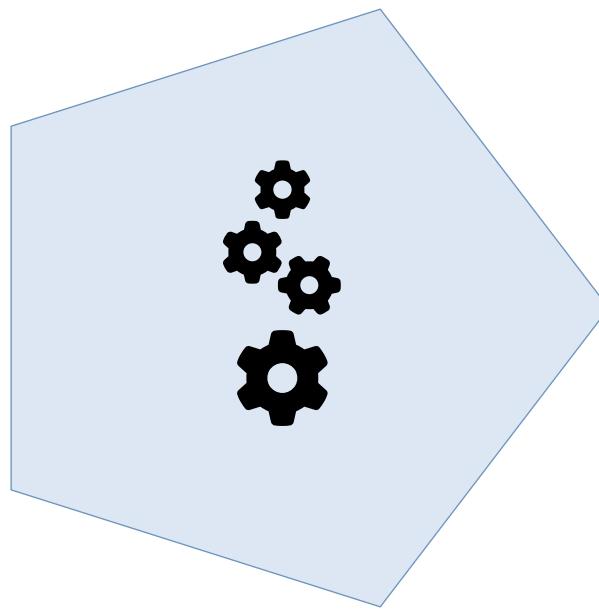
$P(\neg Y | X) = 1 - P(Y | X)$
 $P(Z \wedge Y | X) = P(Z | Y \wedge X) \cdot P(Y | X)$
 $P(Z \vee Y | X) = P(Z | X) + P(Y | X) - P(Z \wedge Y | X)$
choose $\arg \max_D U(D | Y \wedge X) \cdot P(Y | D \wedge X)$

the machine works by **4 fundamental mechanisms**

“laws of robotics”
for **AI** dealing with *uncertainty & decision-making*

Breaking the laws → logical inconsistency & suboptimality

Optimal Predictor Machine



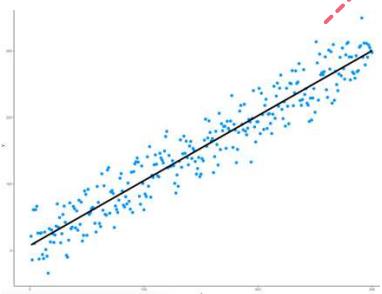
...

Optimal Predictor Machine

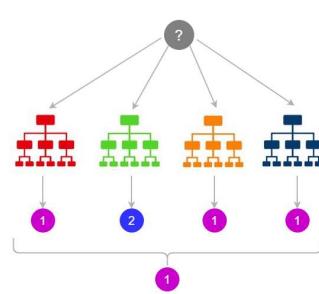
SLOW...

(with present technologies)

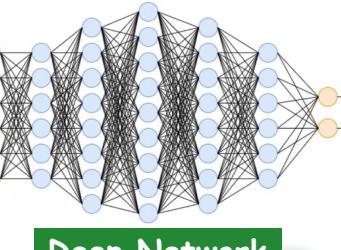
fast!



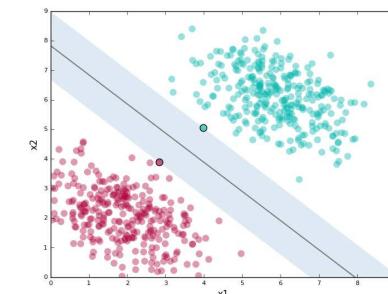
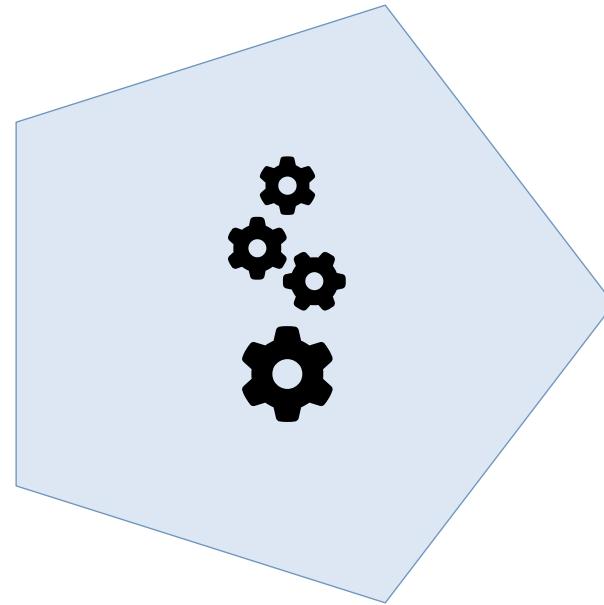
Linear Regression



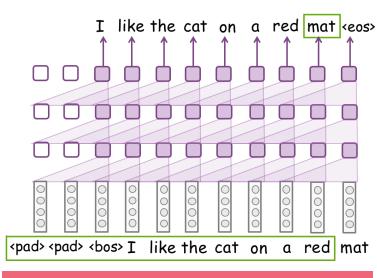
Random Forest



Deep Network



Support Vector Machine



Large Language Model

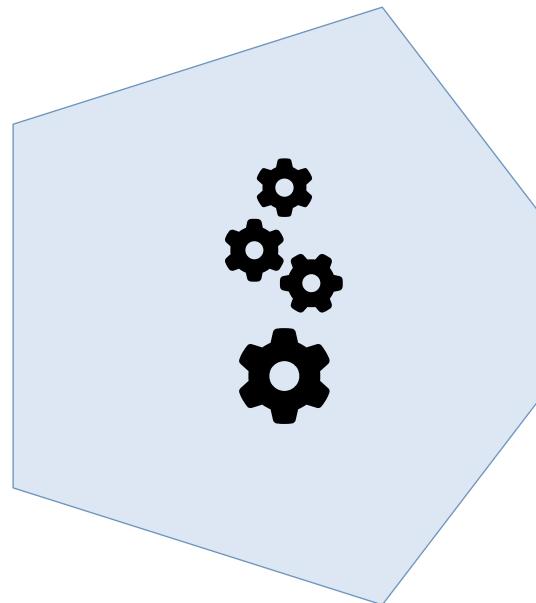


...

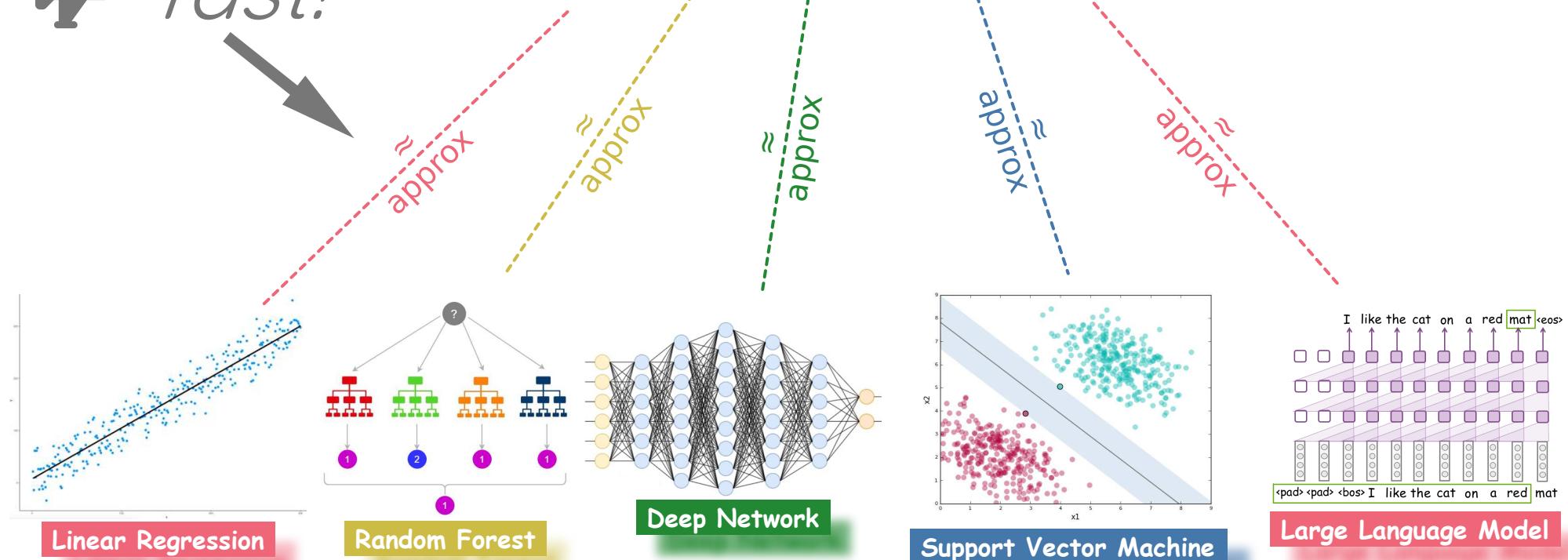
Optimal Predictor Machine

SLOW...

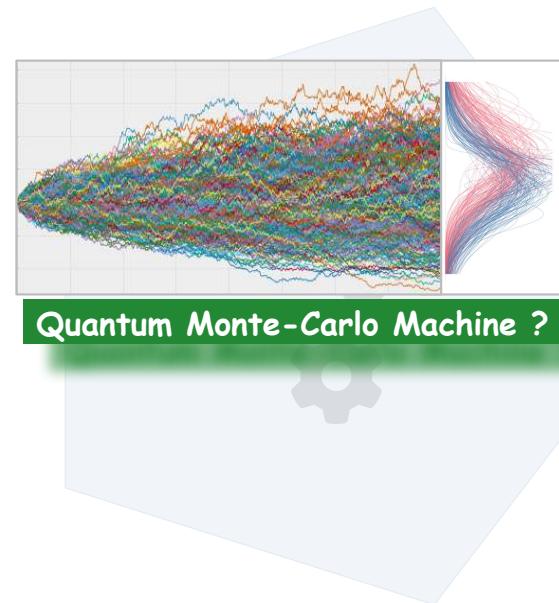
(with present technologies)



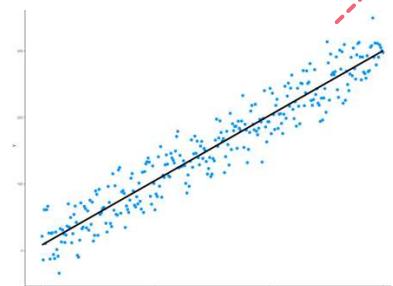
fast!



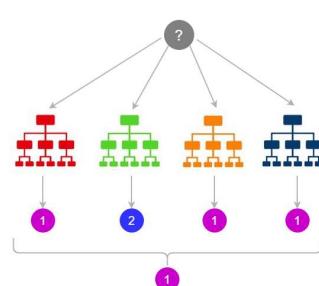
 *fast!* →
(with future technologies?)



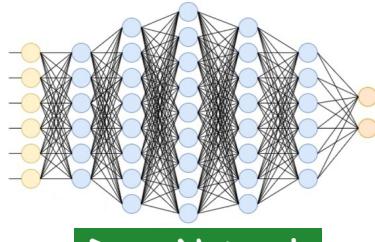
✗ **SLOW...**



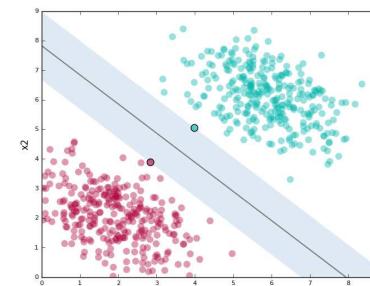
Linear Regression



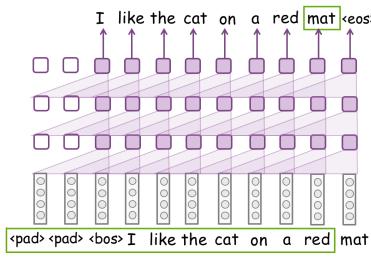
Random Forest



Deep Network



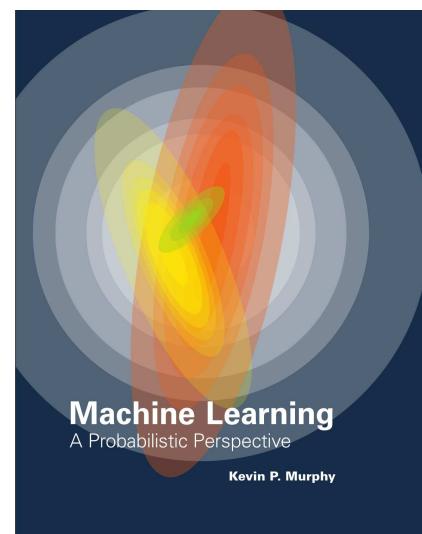
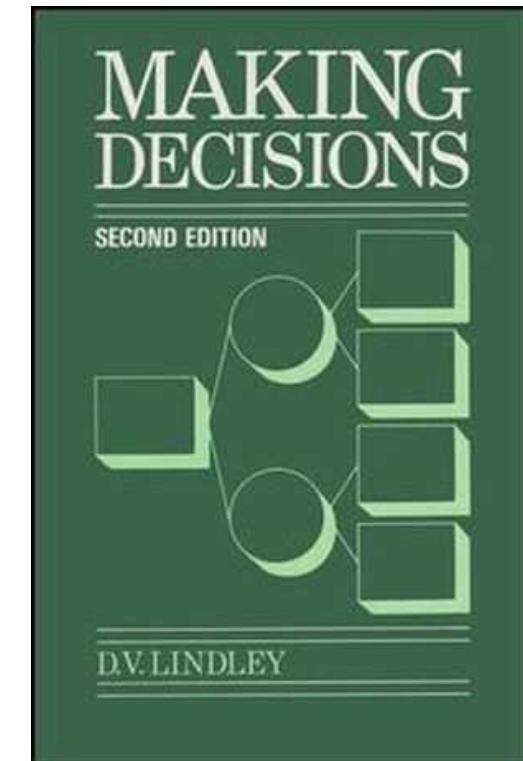
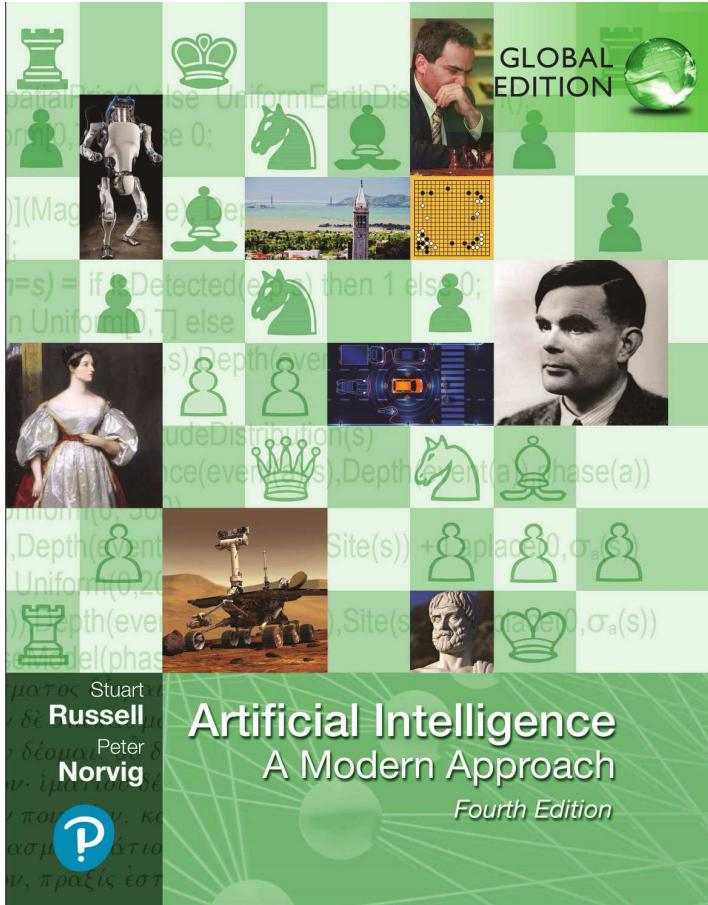
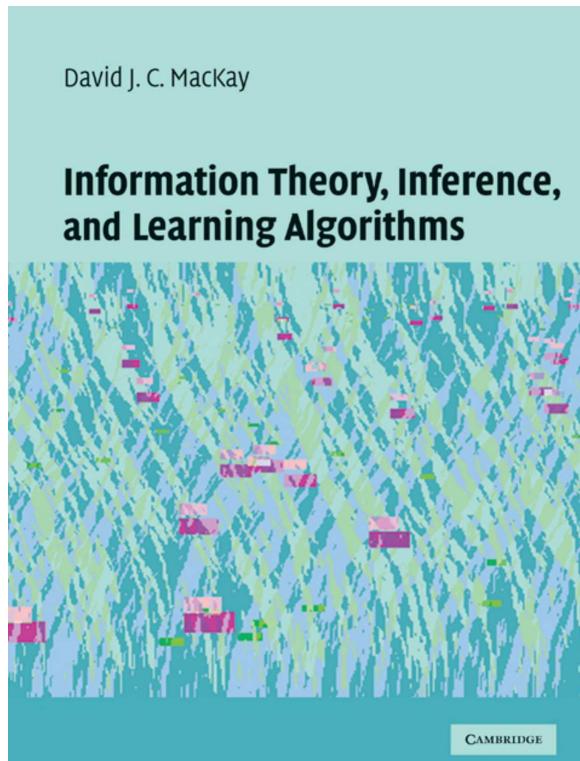
Support Vector Machine



Large Language Model



⋮



- Topic plan (subject to changes) -

💡 Goals & strategy

🤖 Inference I (and some AI)

- *1st connection with machine learning*

>Data I

⚙️ Inference II

- *2nd connection with machine learning*

Data II

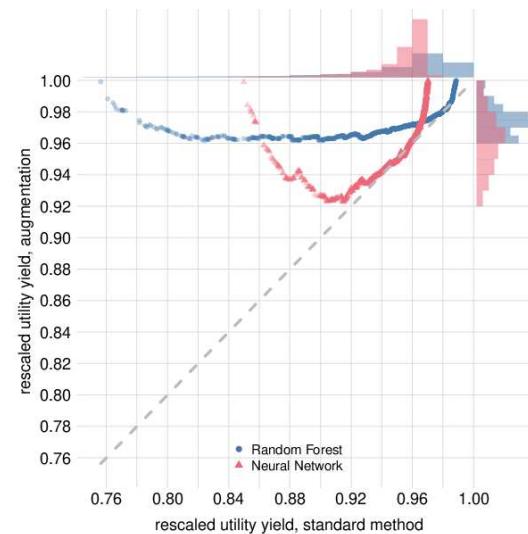
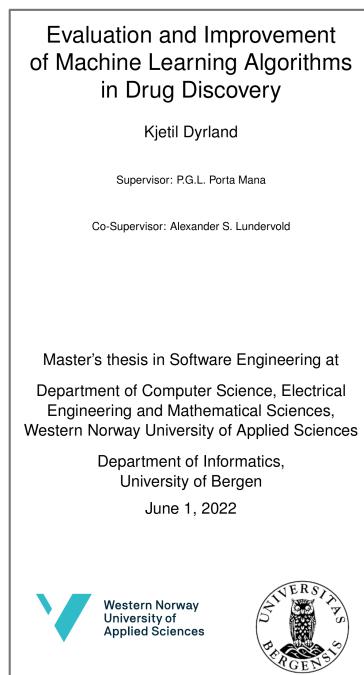
💰 Utility theory (gains & losses)

↔ Decision-making

👉 Machine learning
... and how to improve it!

- Projects & ideas -

- Implement utilities in present ML algorithms
- Predict max achievable accuracy of some ML algorithms
- Push the operative boundaries of the Optimal Machine
- ...



Canvas

<https://hvl.instructure.com/courses/25074>

Course notes

<https://pglpm.github.io/ADA511/>

Software:

R

<https://www.r-project.org/>

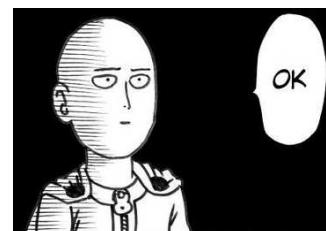
Python

<https://www.python.org/>

One Punch Man

<https://onepunchman.fandom.com/>

The screenshot shows the course page for 'ADA511 0.1 Data science and data-driven engineering'. The left sidebar contains a navigation menu with sections like 'Dear student and aspiring data engineer', 'Preface', 'An invitation', 'Inference I', and 'Data I'. The main content area features a title 'ADA511 0.1 Data science and data-driven engineering' with a sub-section 'If you can't join 'em, beat 'em.' by U.Schreiber. It also includes a 'Dear student and aspiring data engineer' section and a detailed description of the course goals.



Thank you!



