

1	1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---	---

state

ADA 511

0.1

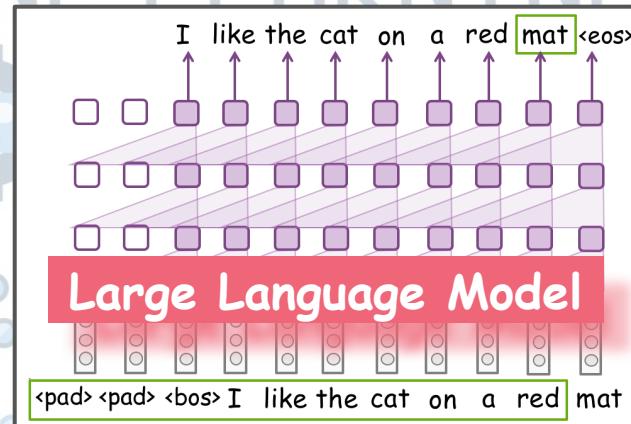
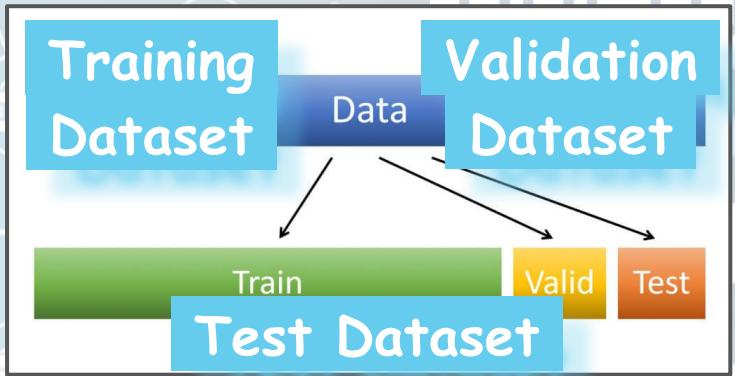
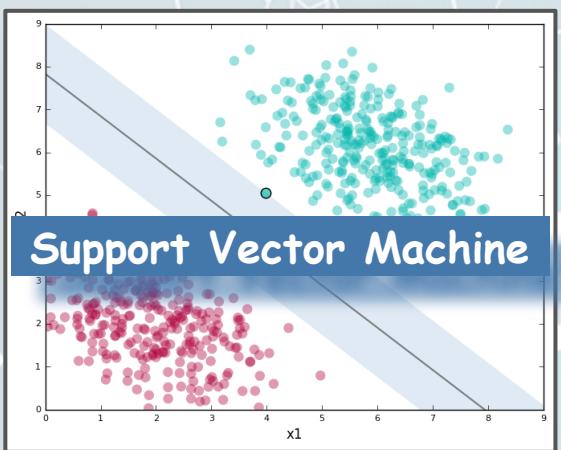
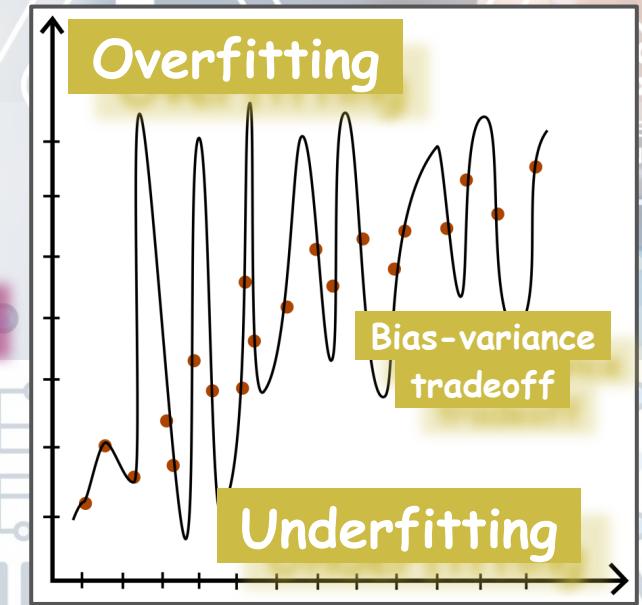
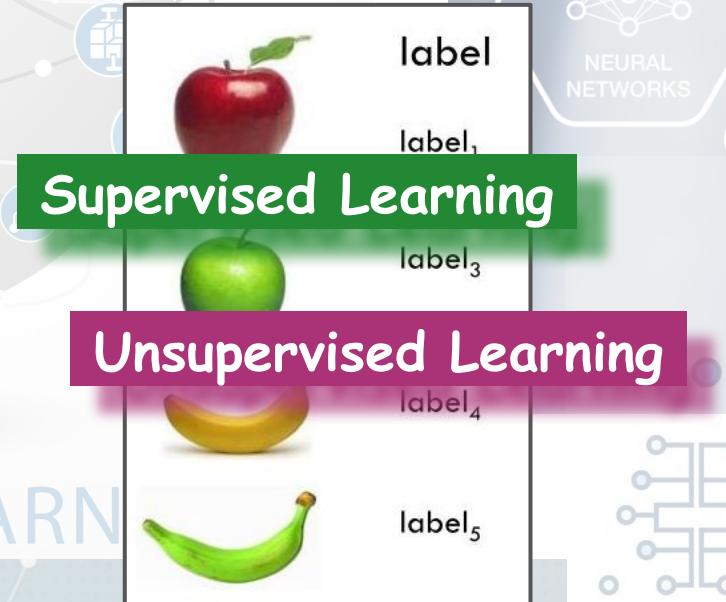
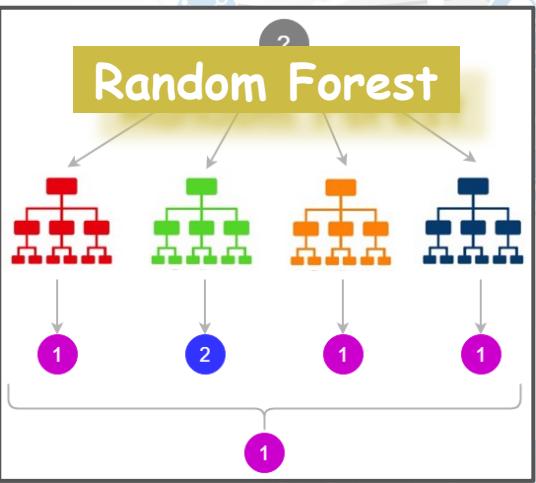
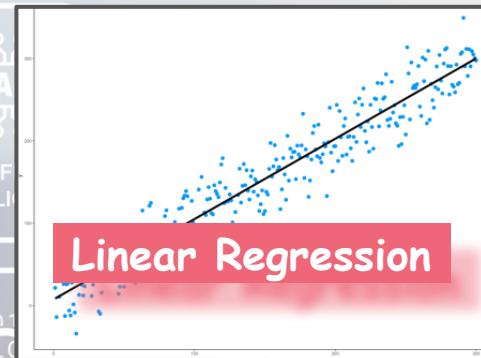
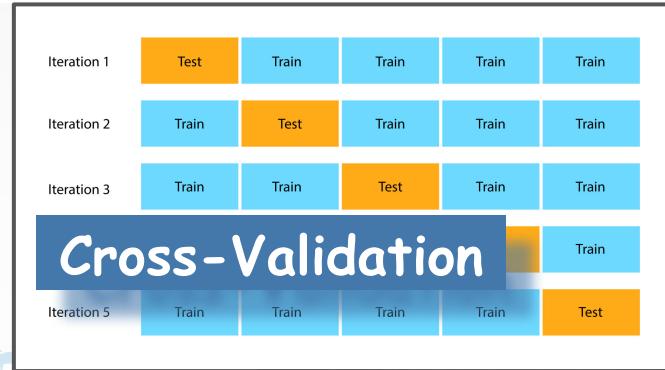
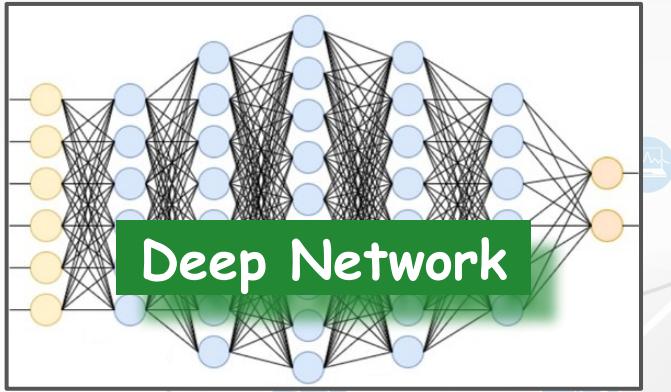
Y

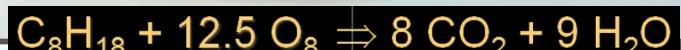
Foundations of

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



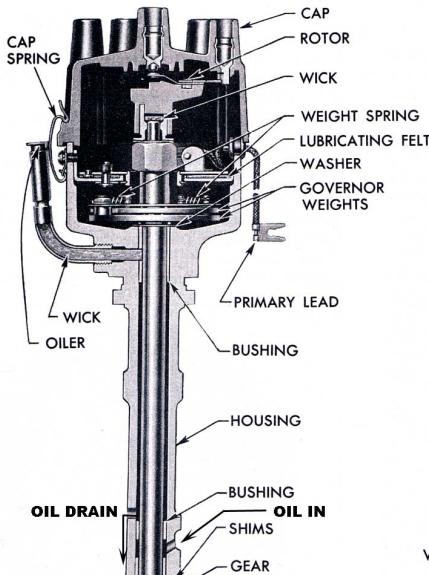
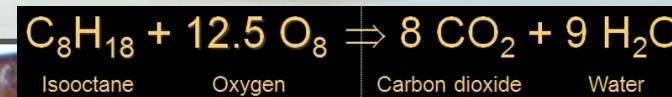
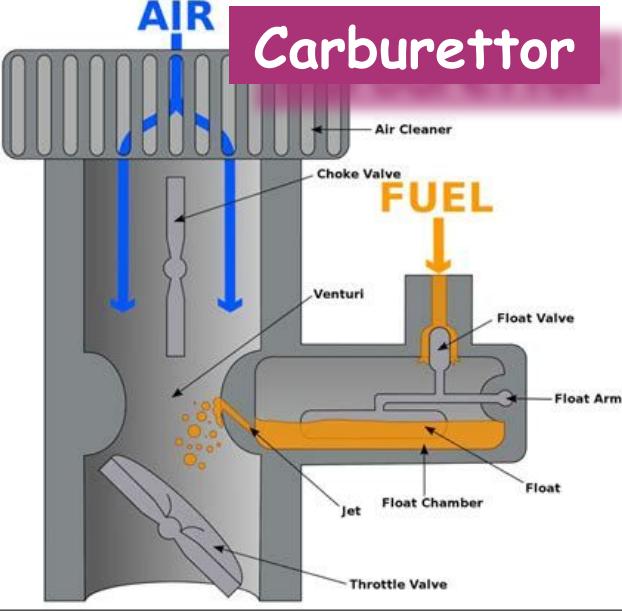
TRUCK



BIGTRUCK

AIR

Carburettor



Distributor



Spark plug

Fuel



Octanes



Valve



Silencer (muffler)



Engine oil

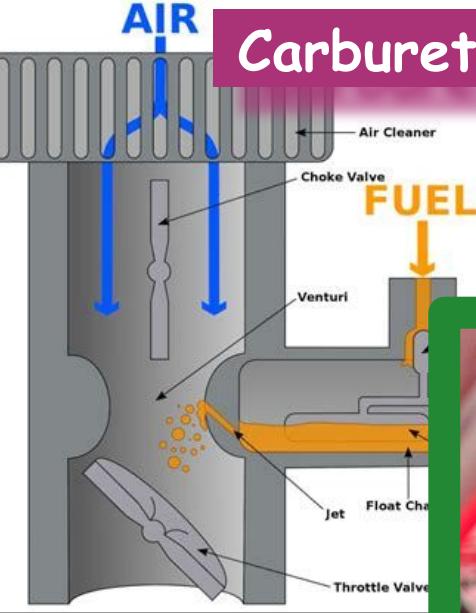
CAMPERVAN

TRUCK

Green emissions

AIR

Carburettor

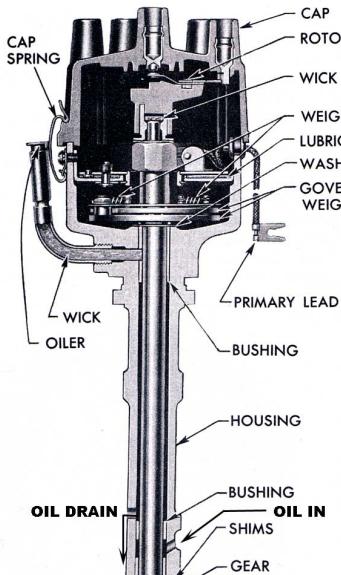


Fuel



Octanes

car mechanic



Distributor

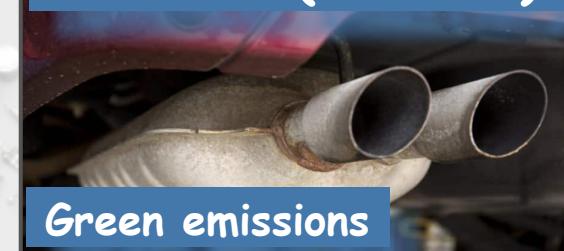


Spark plug

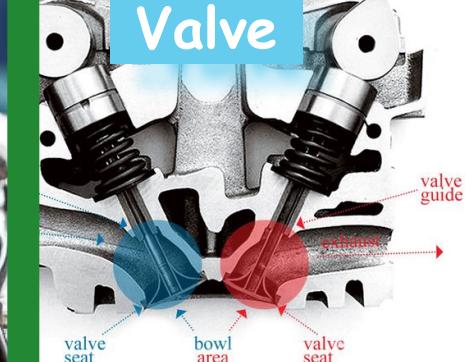
Engine oil



Silencer (muffler)



Green emissions



 cartwright



 car mechanic



 EV mechanic

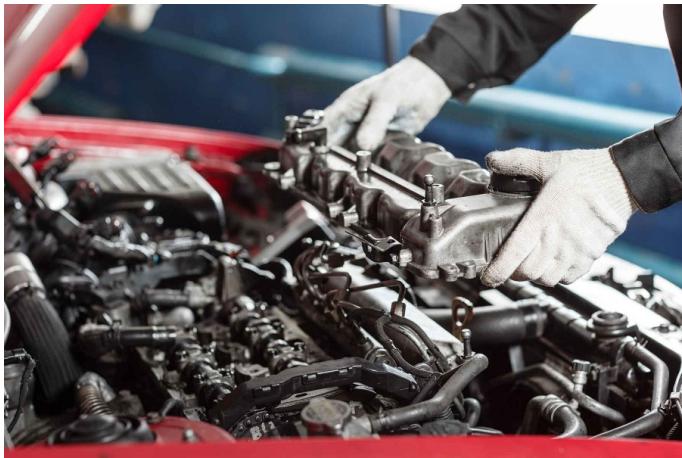


automotive engineer



?

car mechanic



🔍 *specific*

🔨 *practical*

☰ *procedural*

⚓ *technology-dependent*

...

automotive engineer



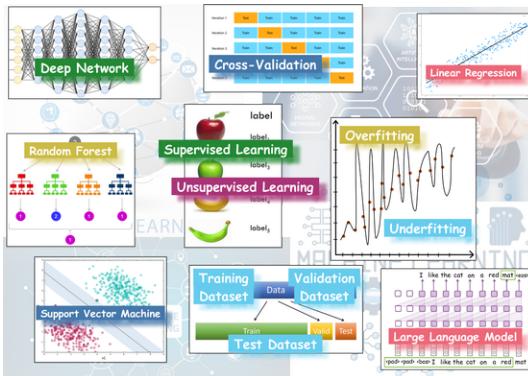
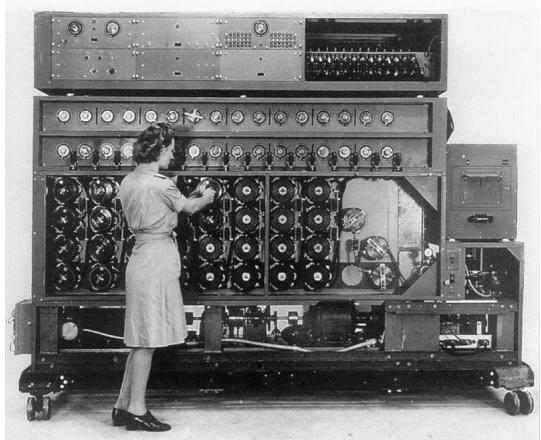
💡 *general*

Ἁ *principled*

🧭 *exploratory*

🐦 *technology-independent*

...



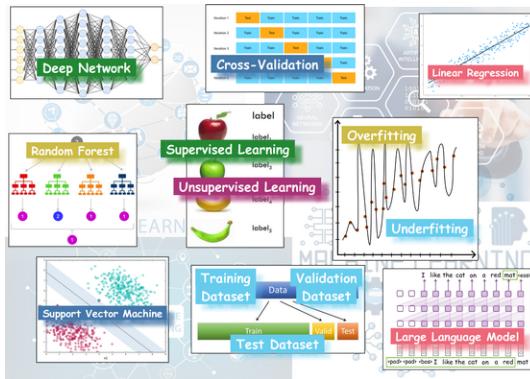
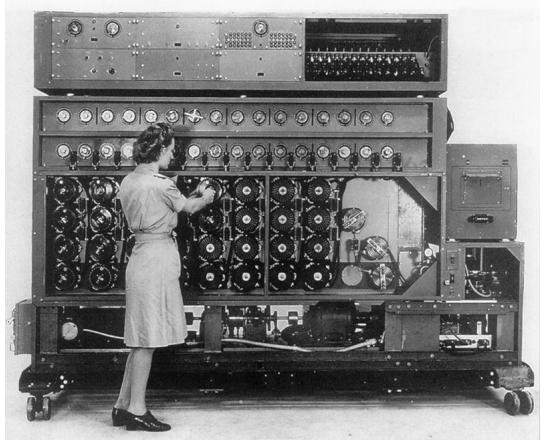
🔧 “data mechanic”

cartwright



🔌 EV mechanic





“data mechanic”

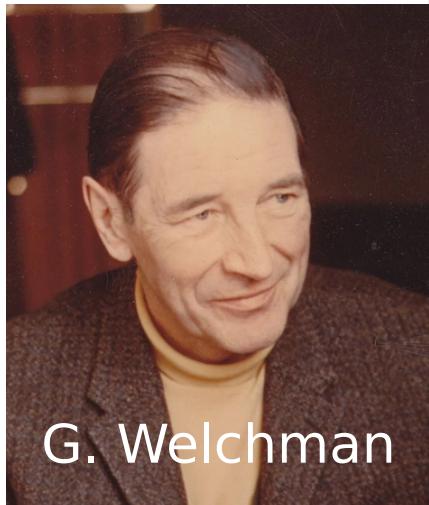


“data engineer”

[insert picture of you all here?]



Alan Turing



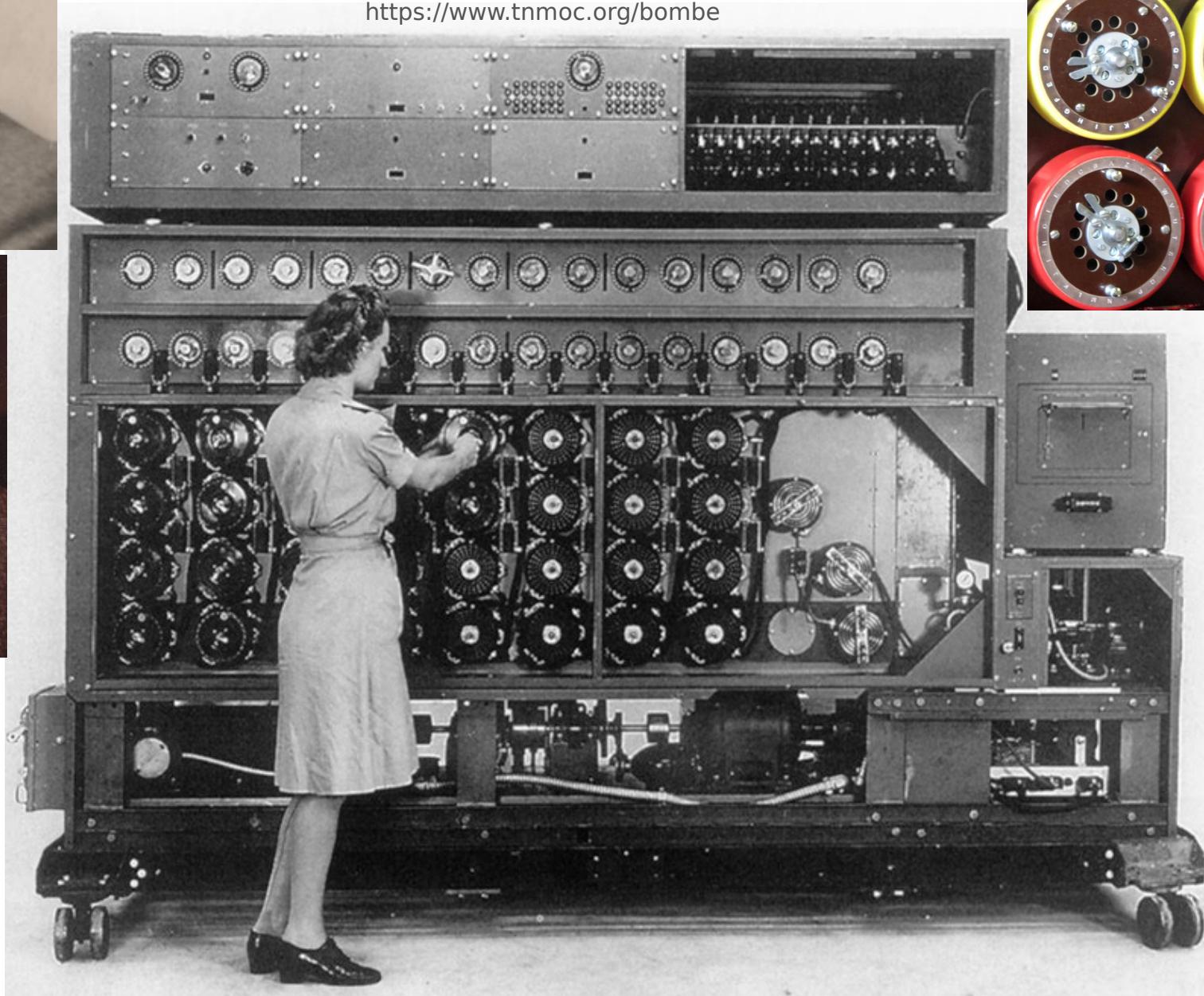
G. Welchman

The Bombe a.k.a. *Agnes* electromechanical computer

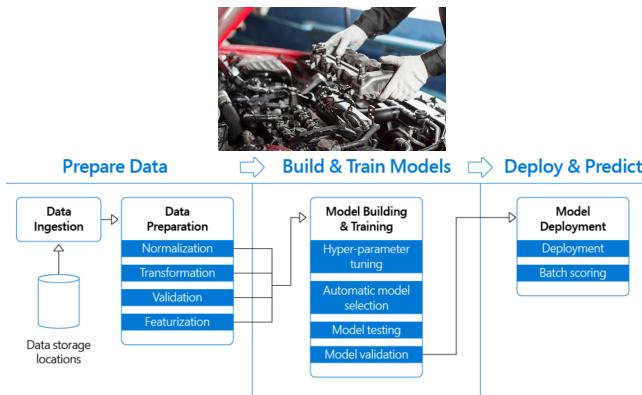
Decoded encrypted Nazi communications during World War II
(efficiency rate: 98.8%)

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

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



“data mechanic”



🔍 *specific*

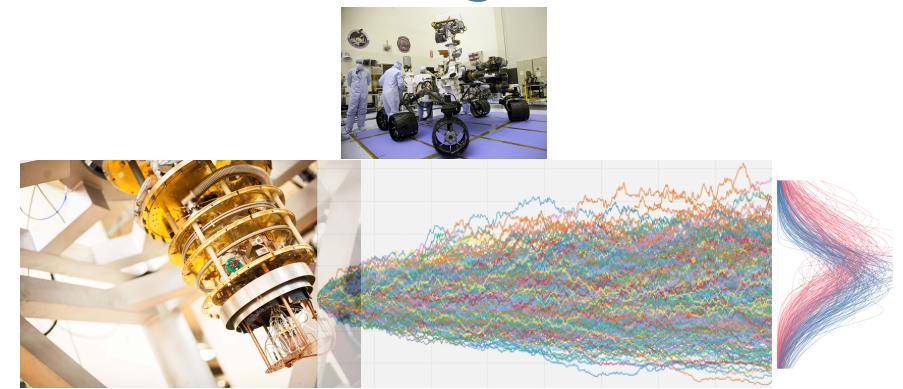
🔨 *practical*

Ξ *procedural*

⚓ *technology-dependent*

...

“data engineer”



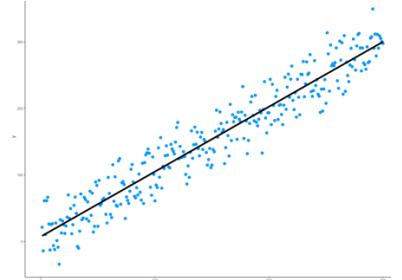
👀 *general*

📐 *principled*

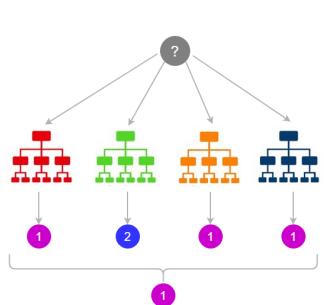
🧭 *exploratory*

🐦 *technology-independent*

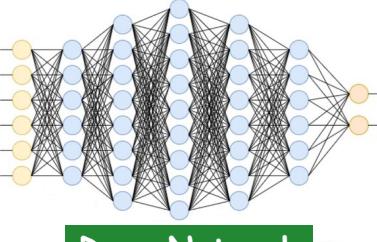
...



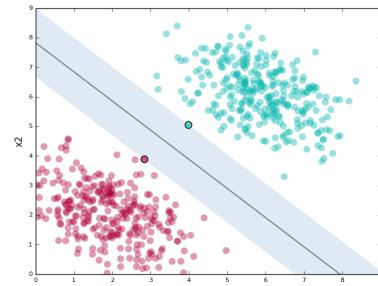
Linear Regression



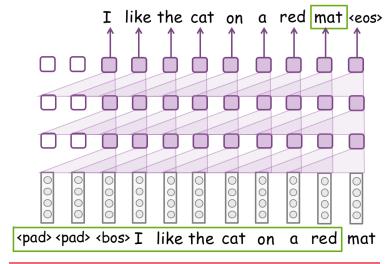
Random Forest



Deep Network



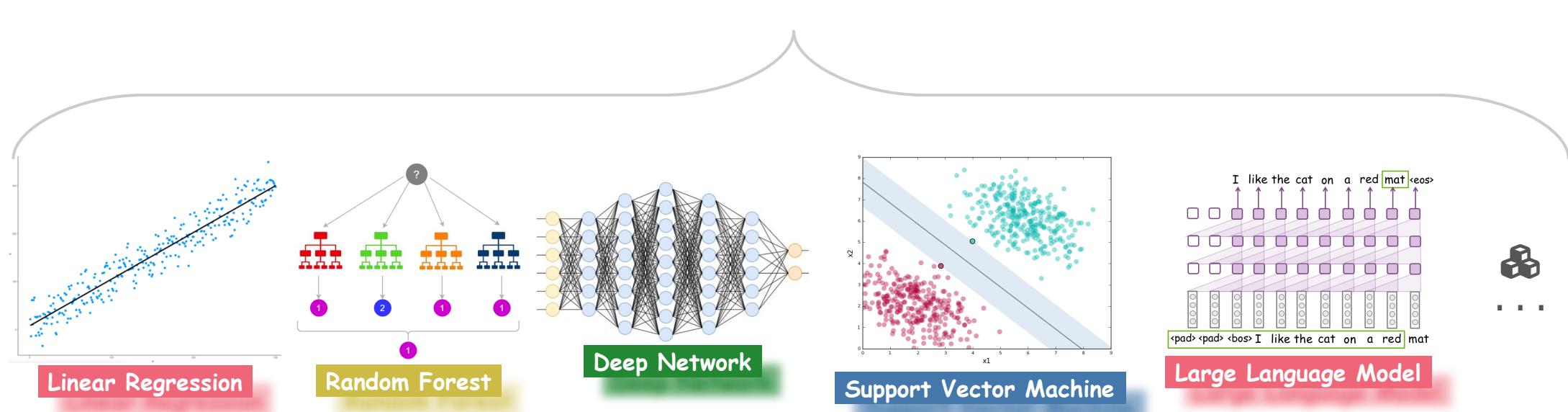
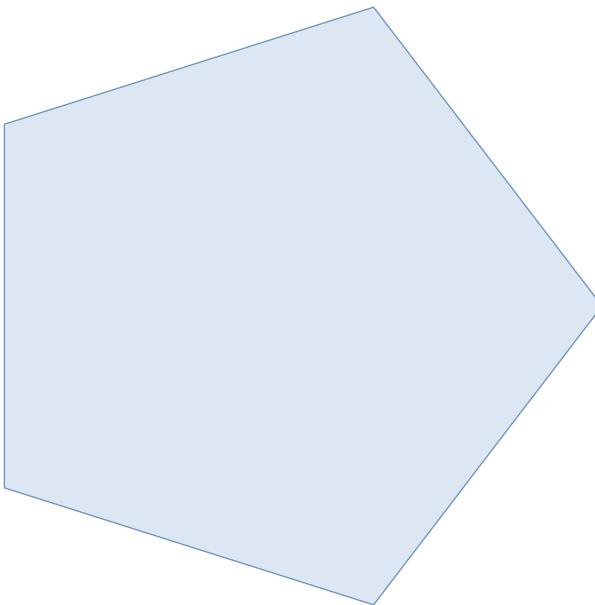
Support Vector Machine



Large Language Model



Optimal Predictor Machine



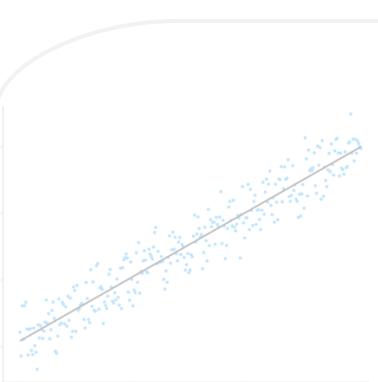
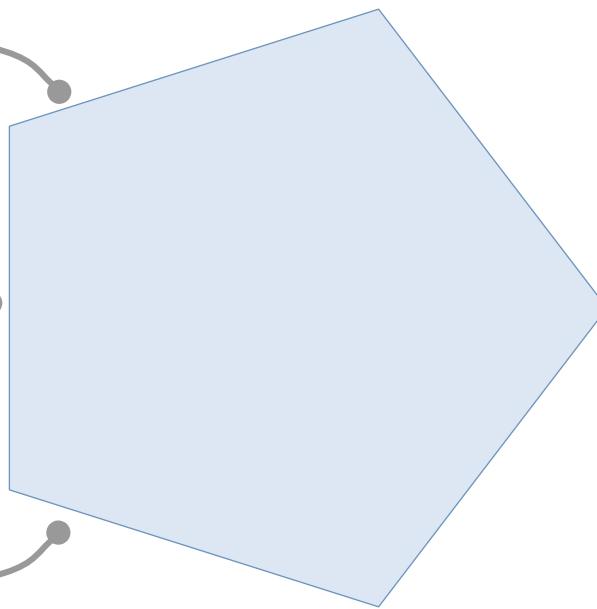
Optimal Predictor Machine

problem-dependent

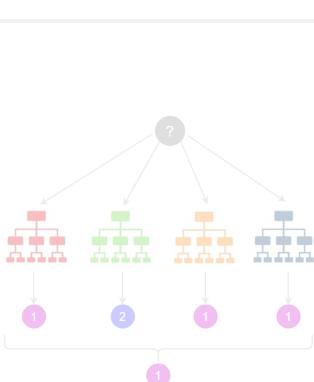
i metadata

data

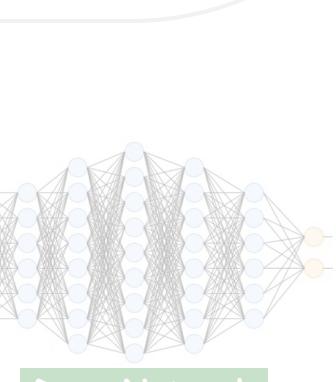
\$⌚ utilities
(gains/losses)



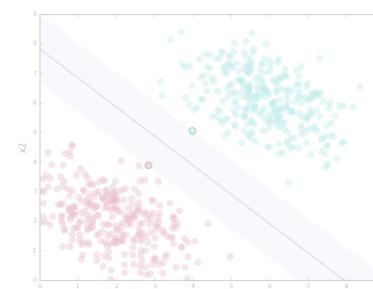
Linear Regression



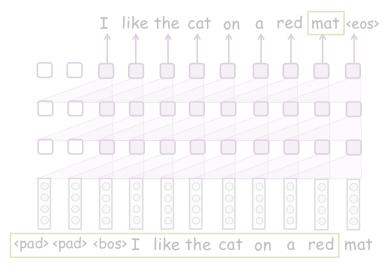
Random Forest



Deep Network



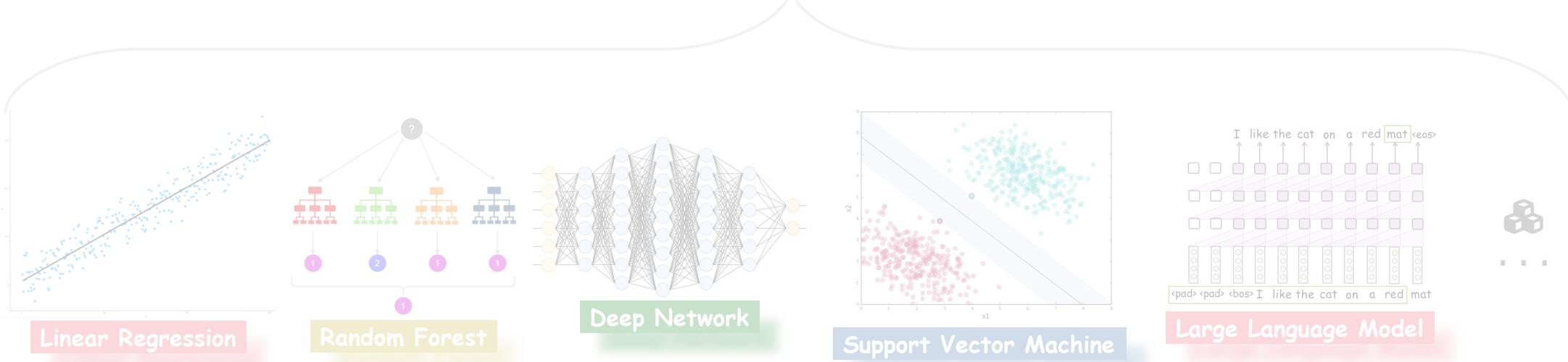
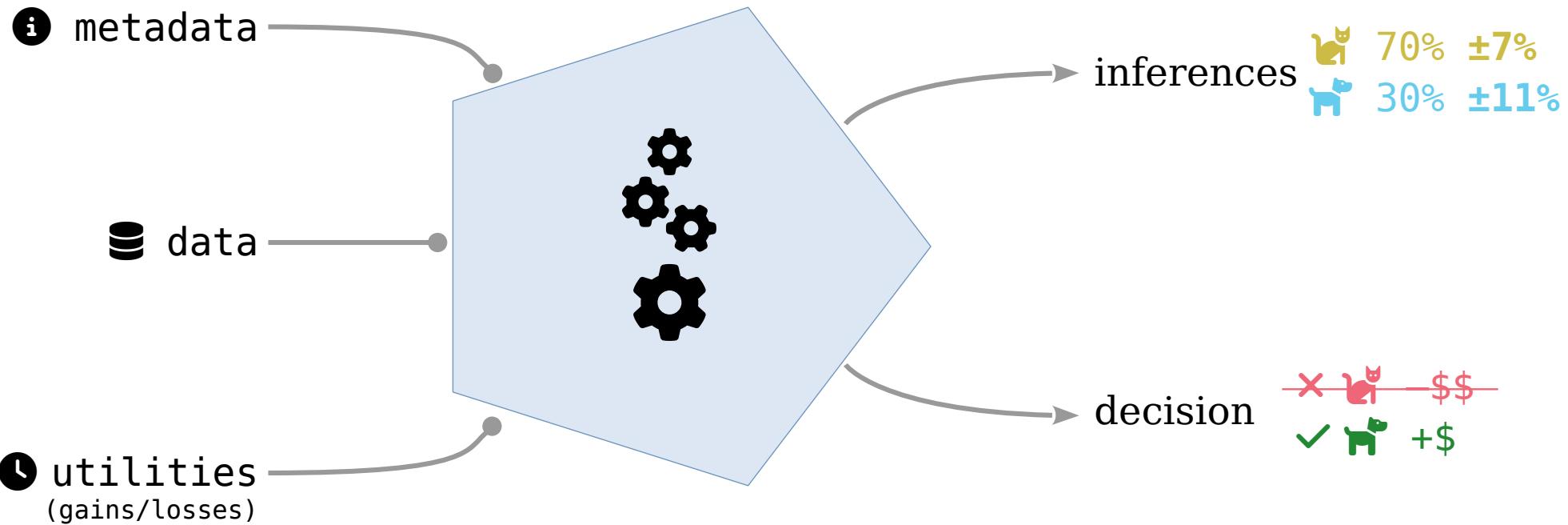
Support Vector Machine



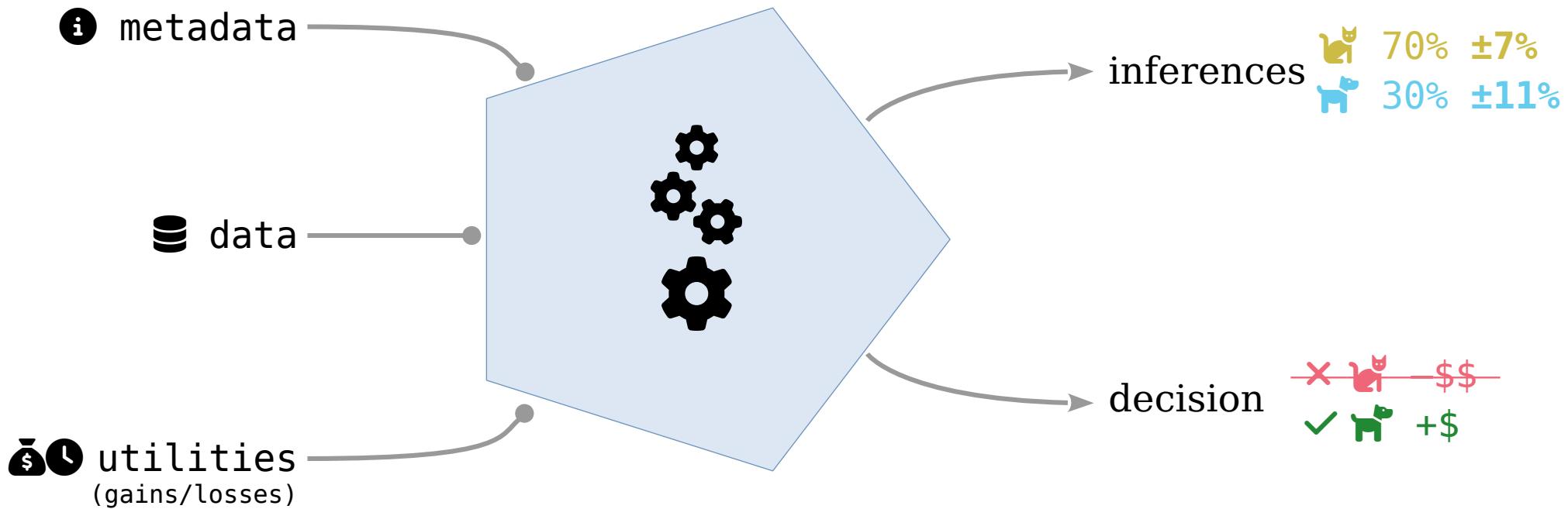
Large Language Model

Optimal Predictor Machine

problem-dependent

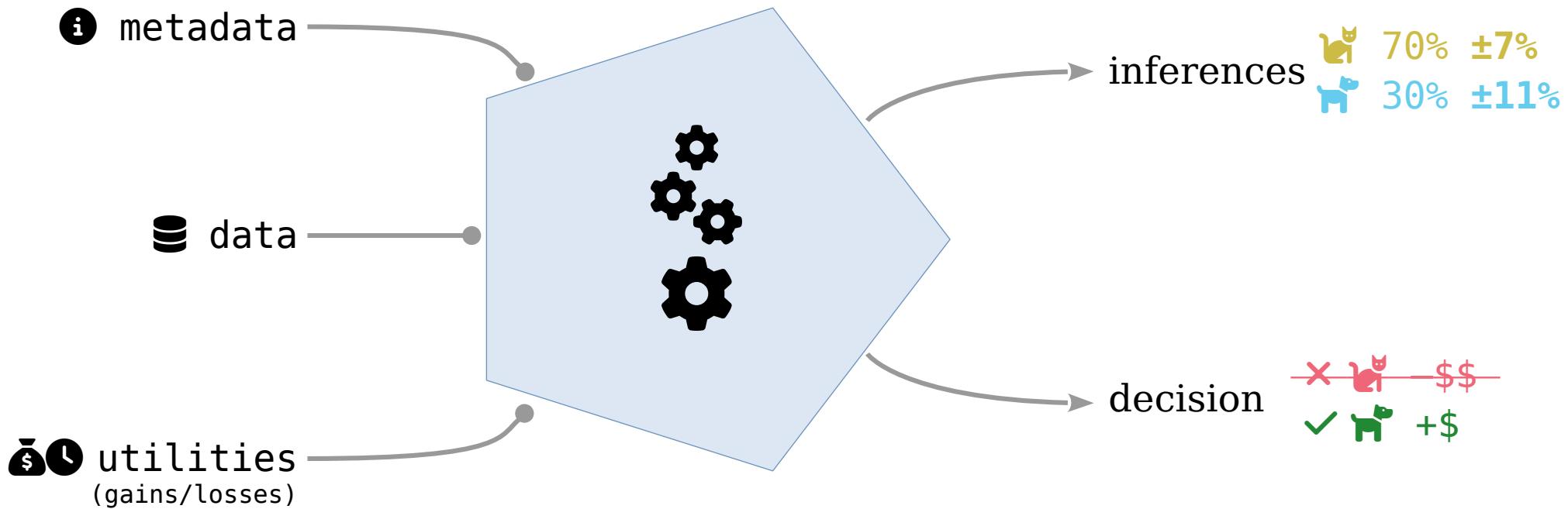


Optimal Predictor Machine



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

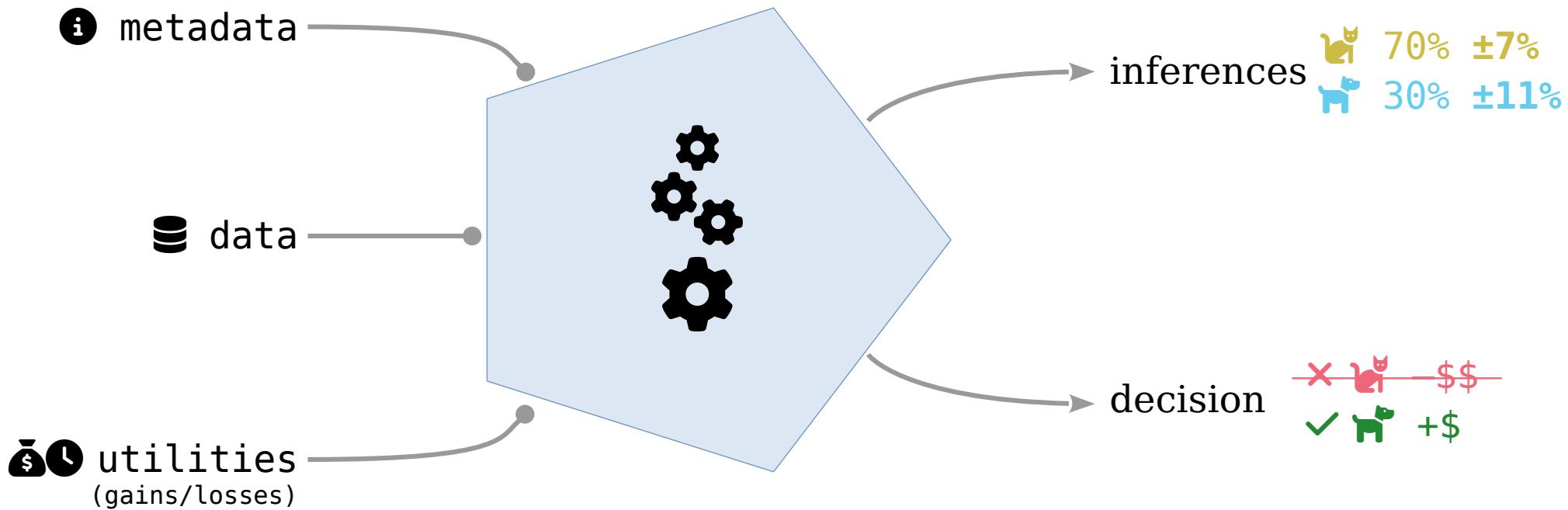
Optimal Predictor Machine



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



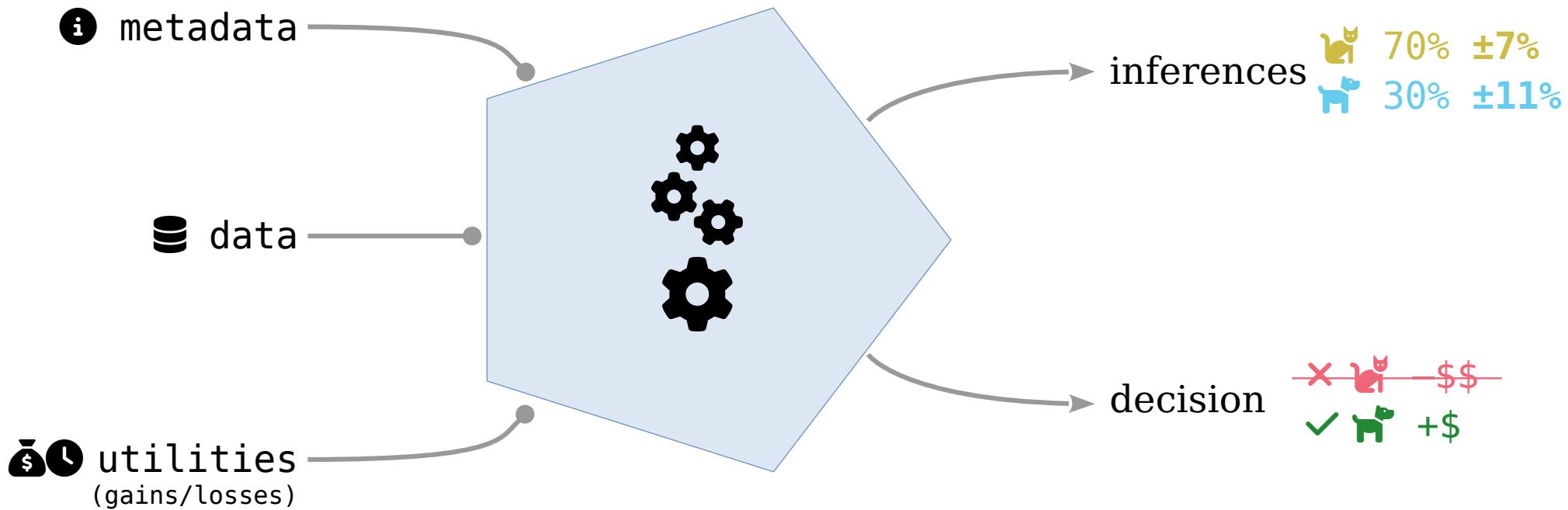
Optimal Predictor Machine



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



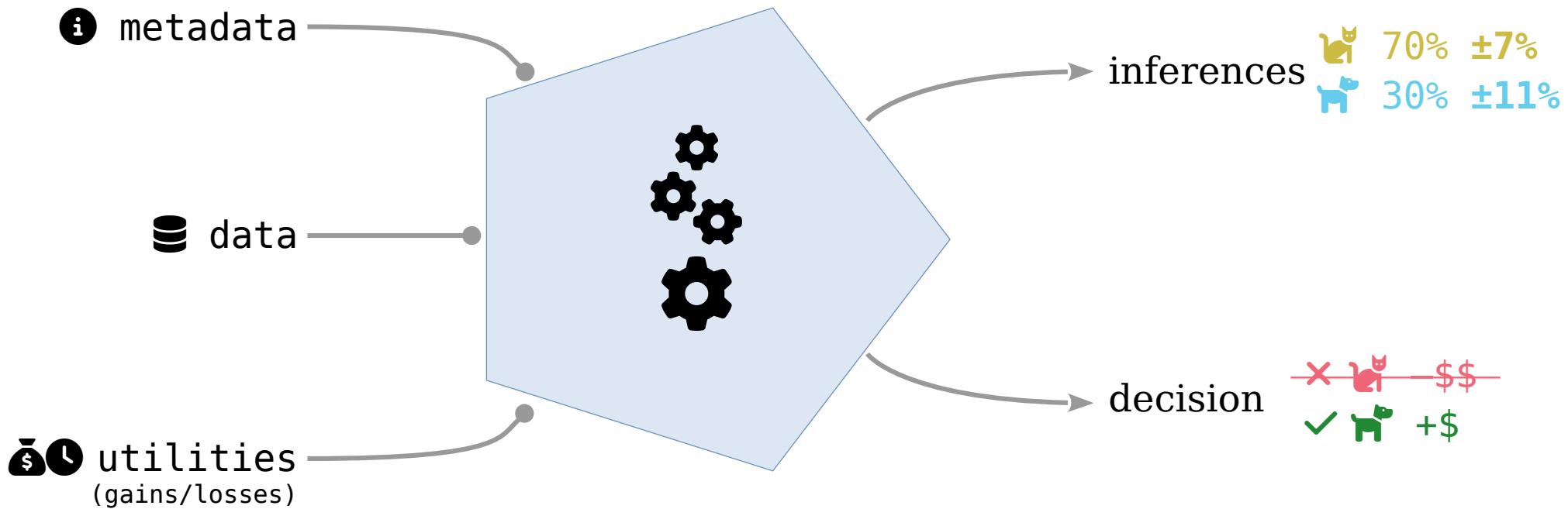
Optimal Predictor Machine



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



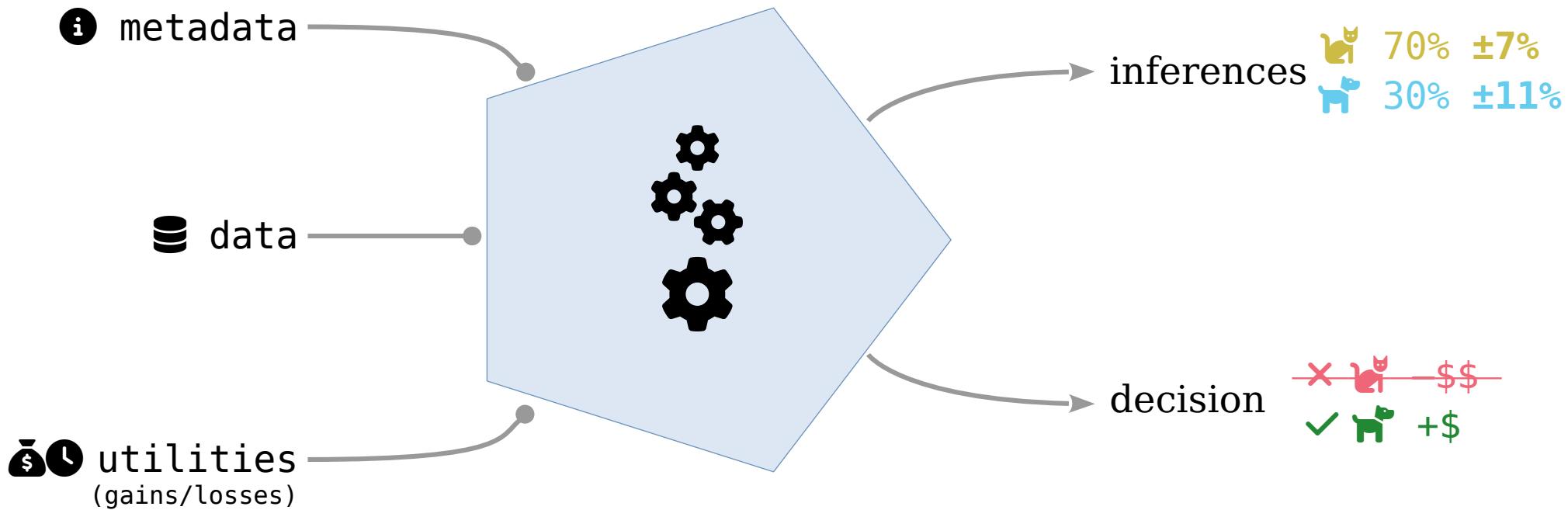
Optimal Predictor Machine



- ✓ 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)

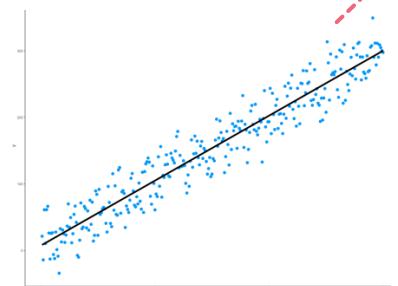
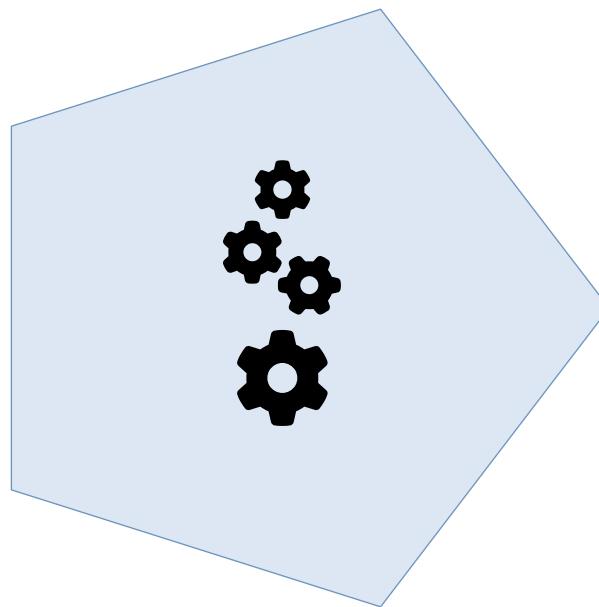




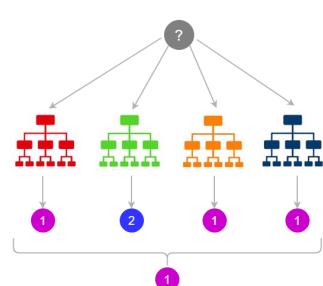
$P(\neg Y | X) = 1 - P(Y | X)$
 $P(Z \wedge Y | X) = P(Z | Y \wedge X) \cdot P(Y | X)$
 $P(Z \vee Y | X) = P(Z | X) + P(Y | X) - P(Z \wedge Y | X)$
choose $\arg \max_D U(D | Y \wedge X) \cdot P(Y | D \wedge X)$

the machine works by **4 fundamental mechanisms**
“laws of robotics”
for **AI** dealing with *uncertainty & decision-making*

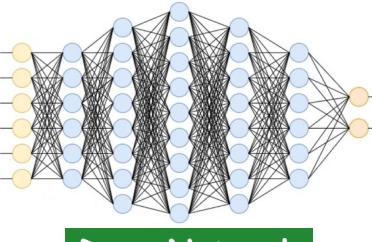
Optimal Predictor Machine



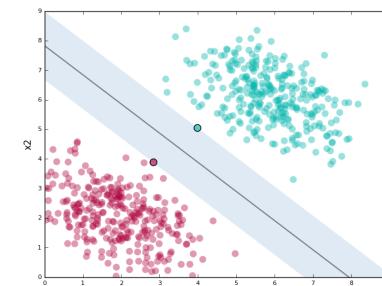
Linear Regression



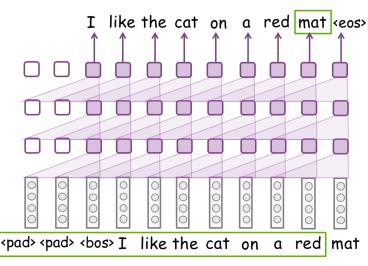
Random Forest



Deep Network



Support Vector Machine



Large Language Model



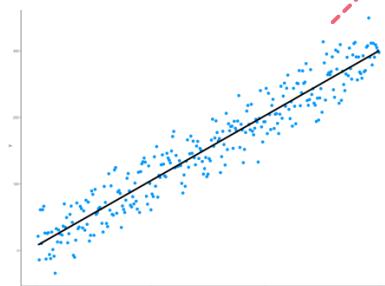
...

Optimal Predictor Machine

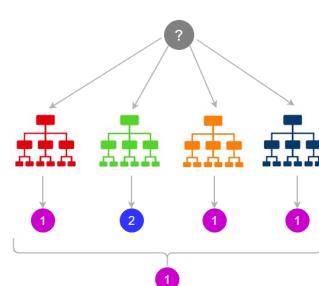
~~Σ SLOW...~~

(with present technologies)

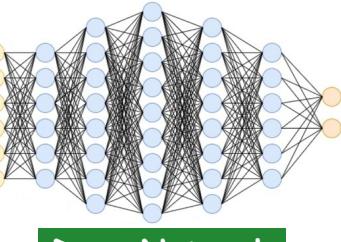
✈️ *fast!*



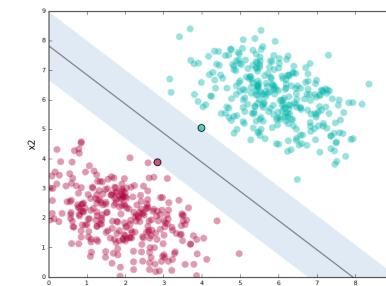
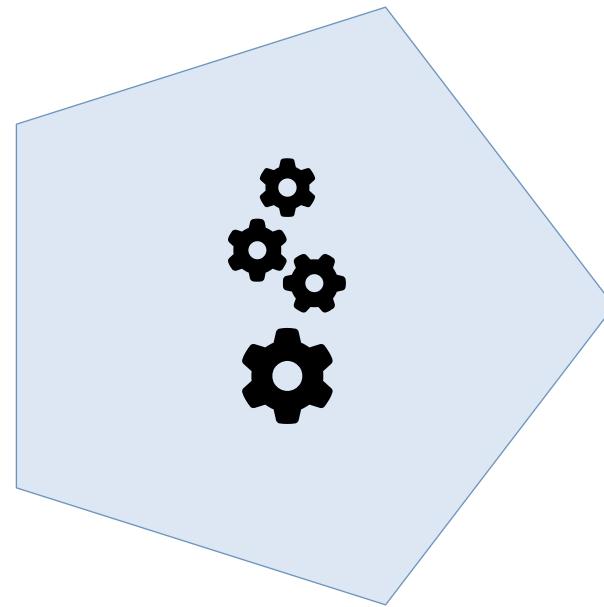
Linear Regression



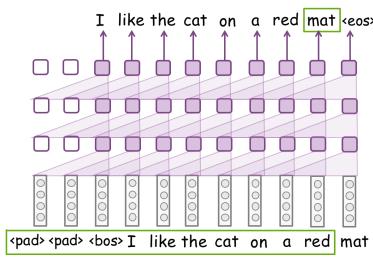
Random Forest



Deep Network



Support Vector Machine

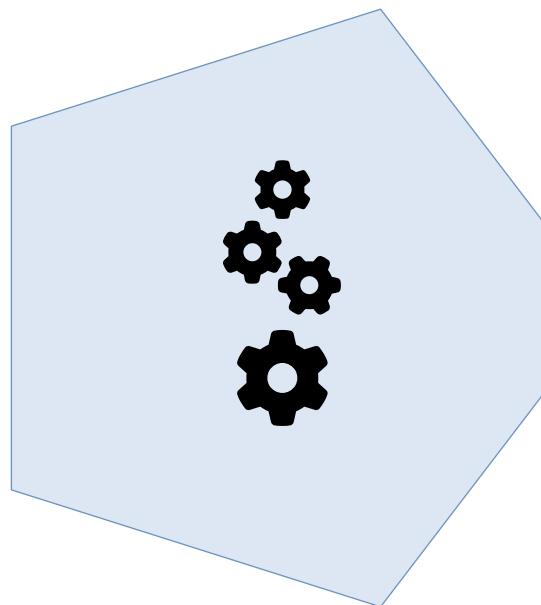


Large Language Model

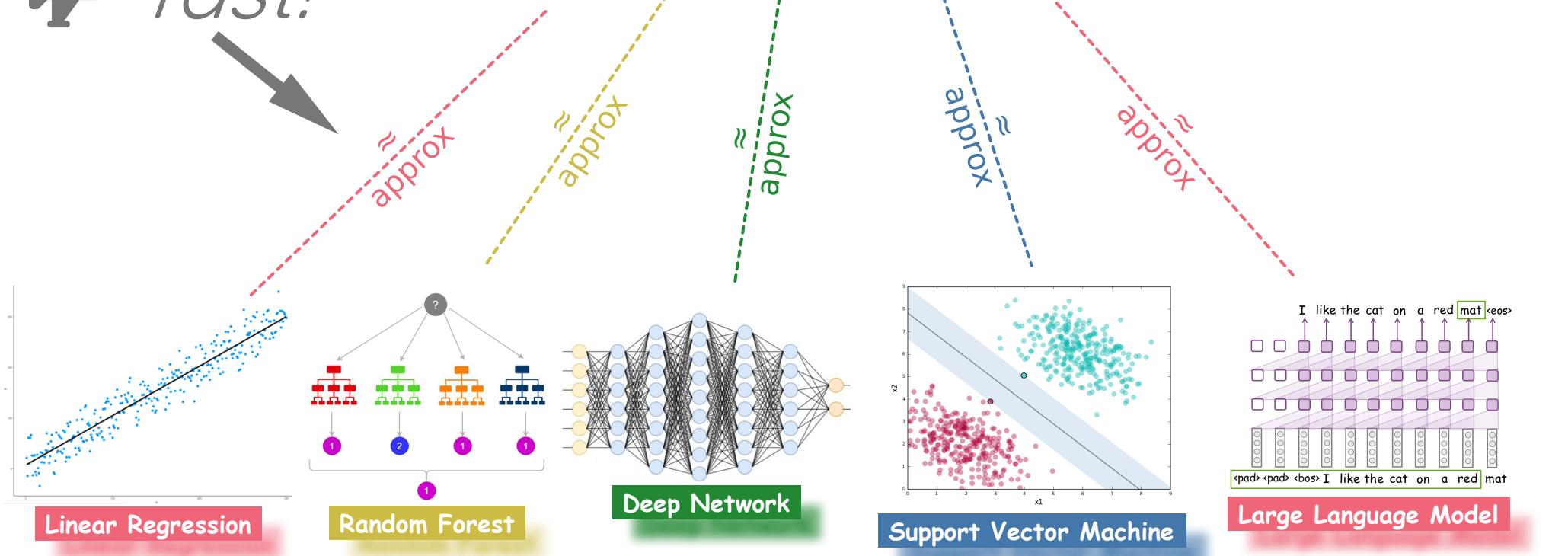


Optimal Predictor Machine

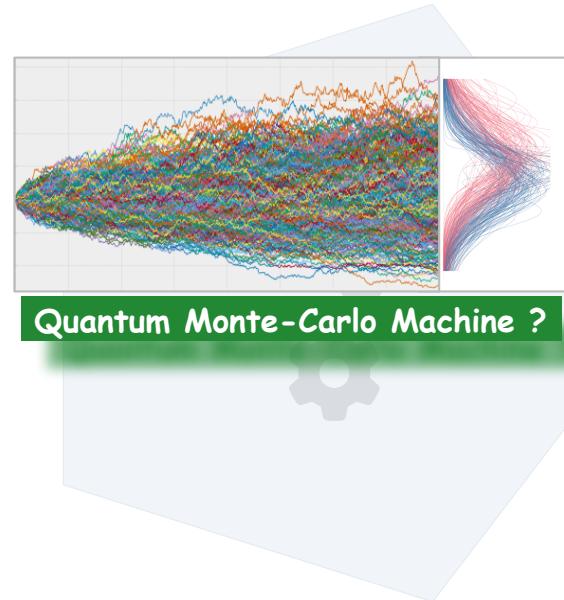
Σ SLOW...
(with present technologies)



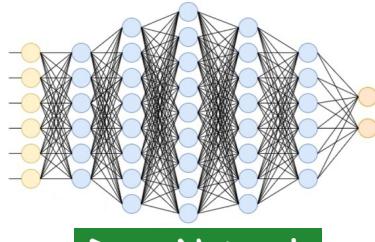
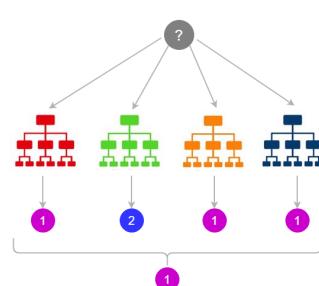
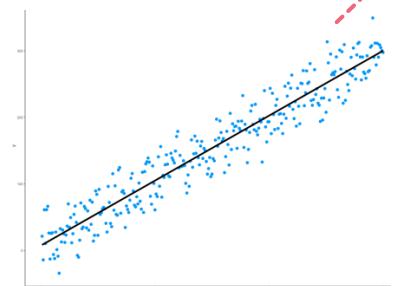
\oplus fast!



 *fast!* →
(with future technologies?)



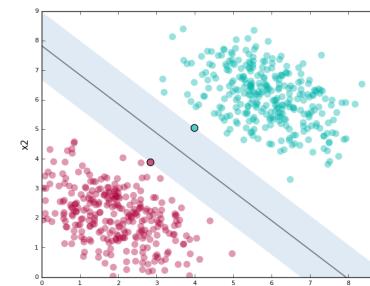
✗ SLOW...



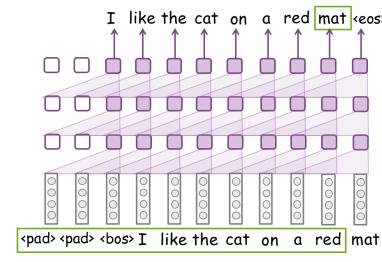
≈ approx

≈ approx

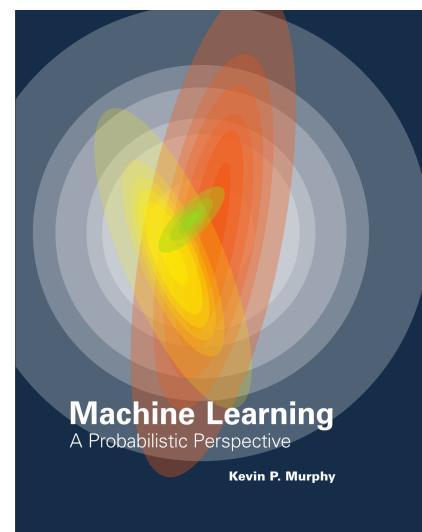
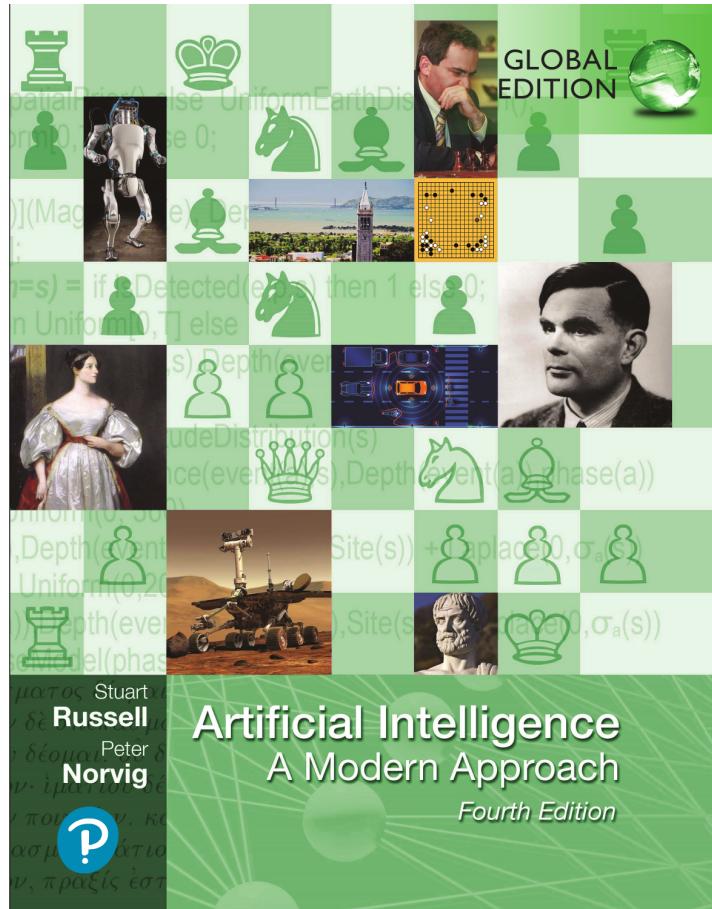
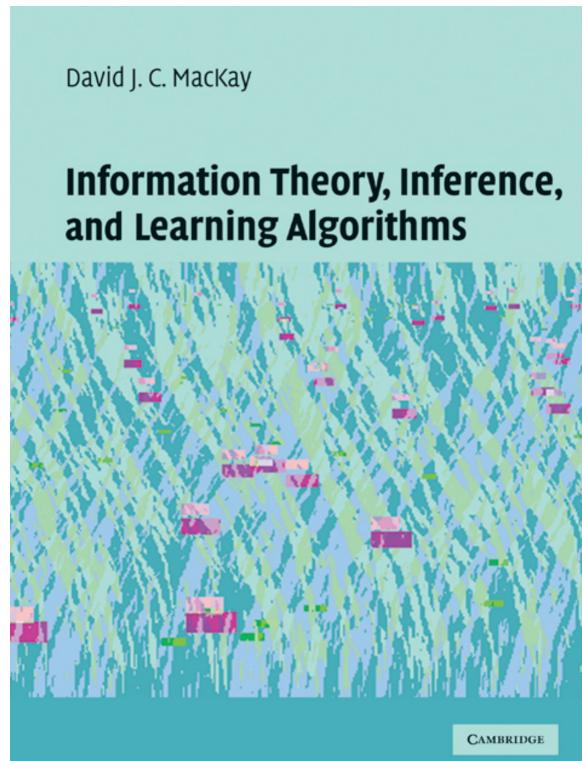
≈ approx



Support Vector Machine



⋮



- Time plan (subject to changes) -

💡 Goals & strategy

🤖 Inference I

- *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!