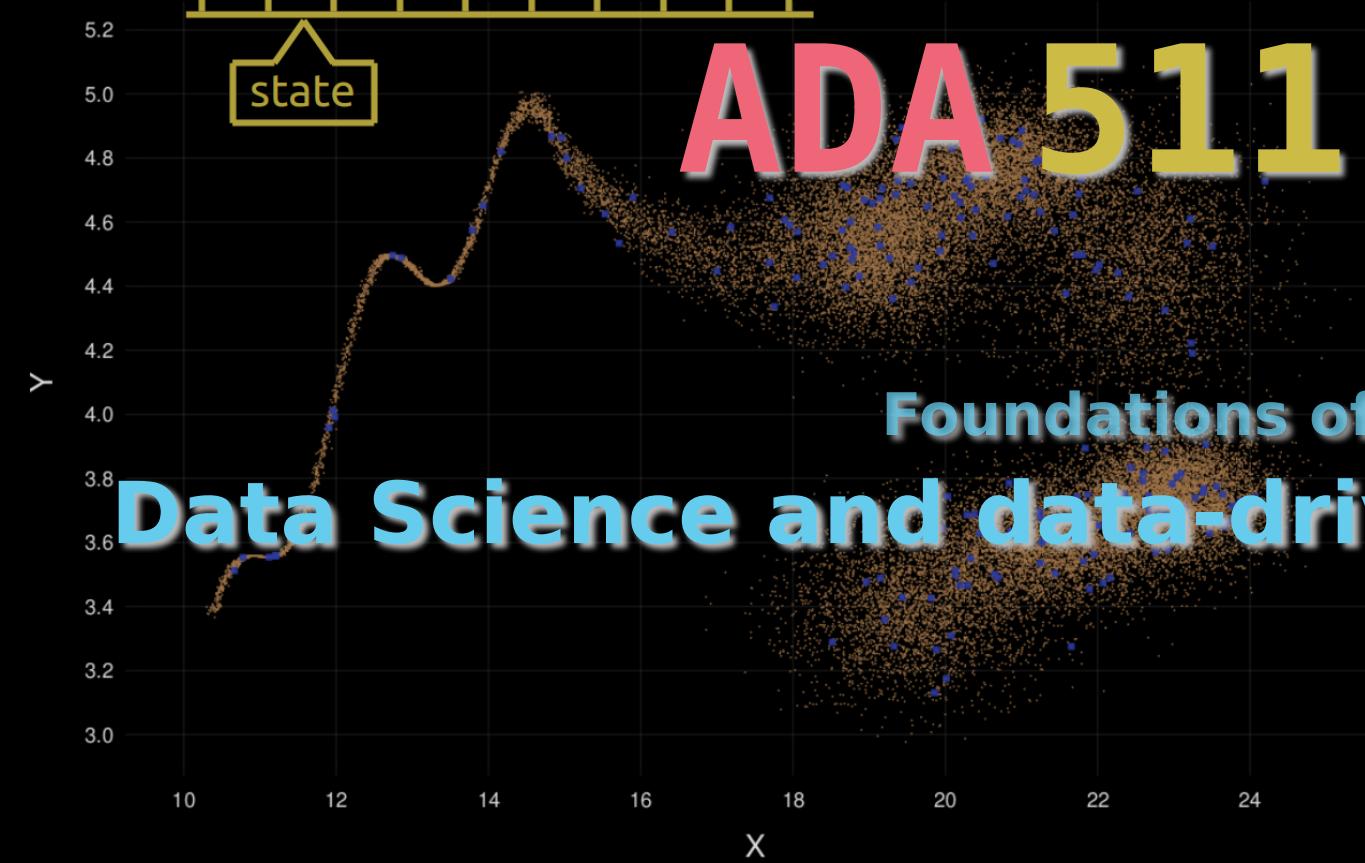


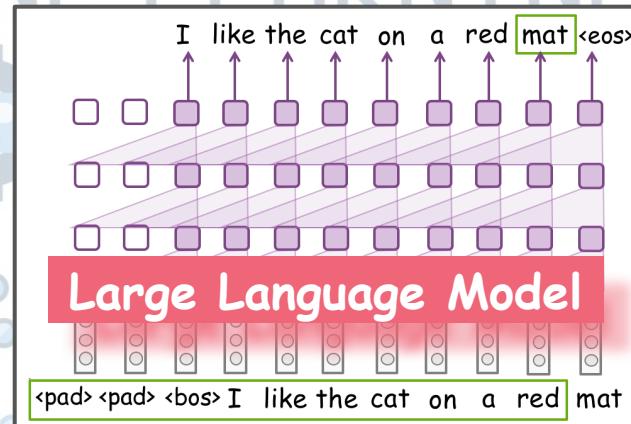
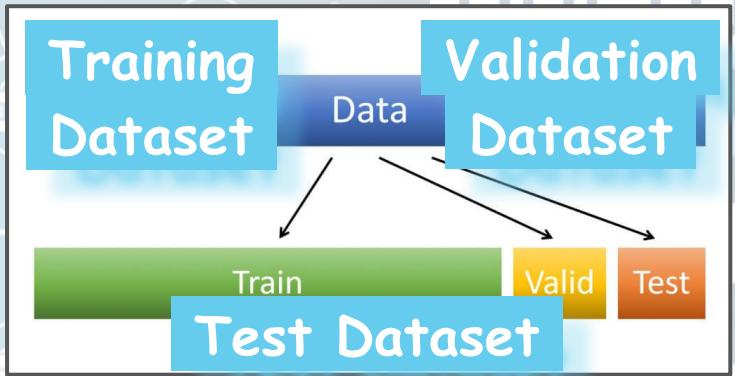
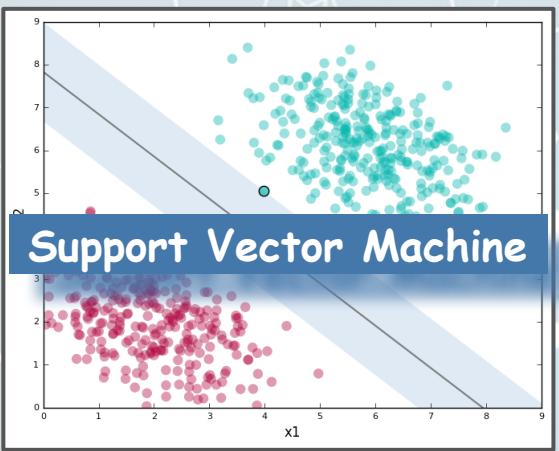
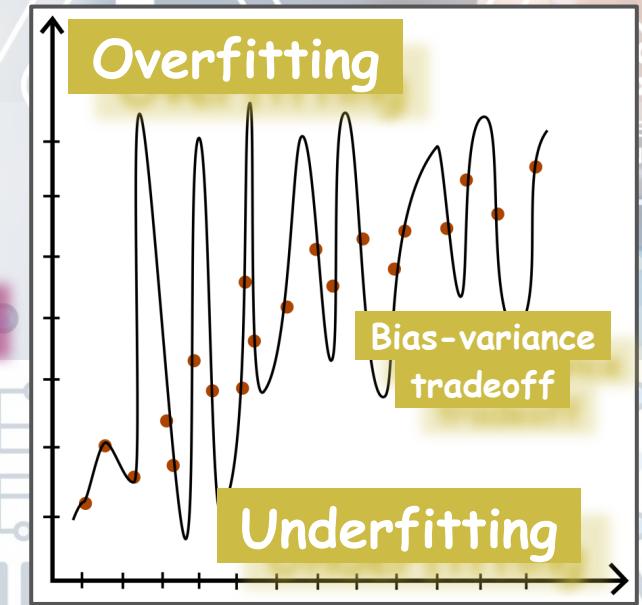
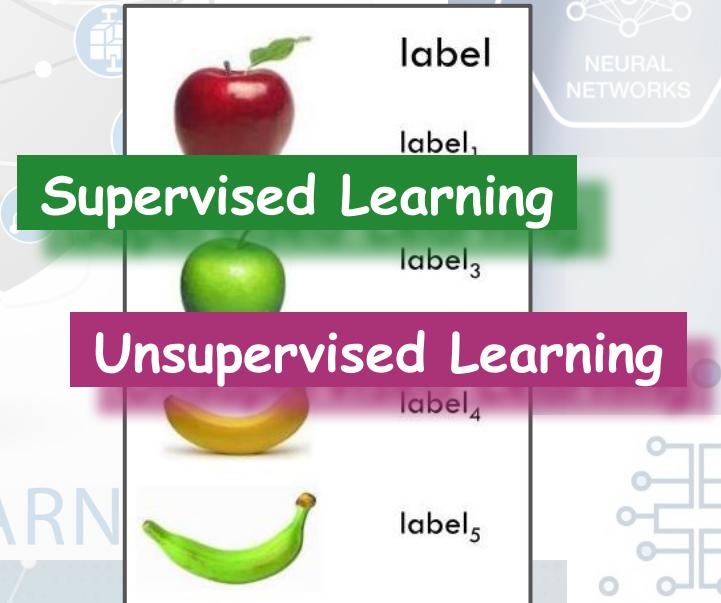
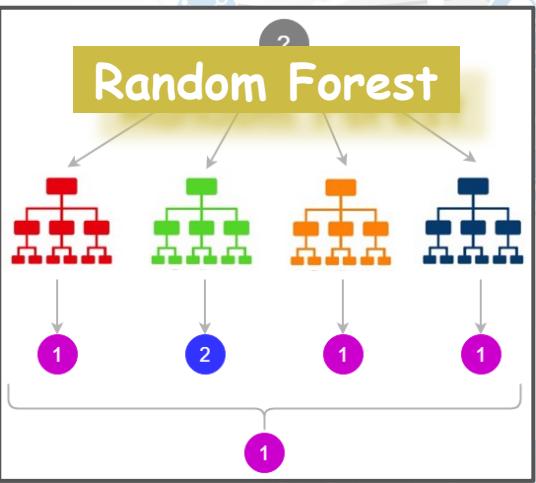
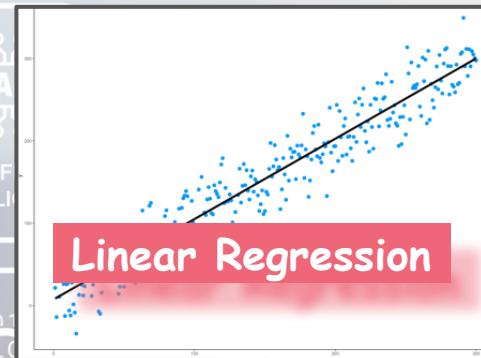
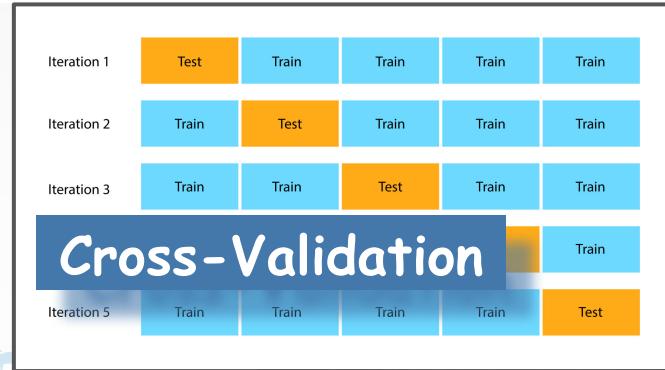
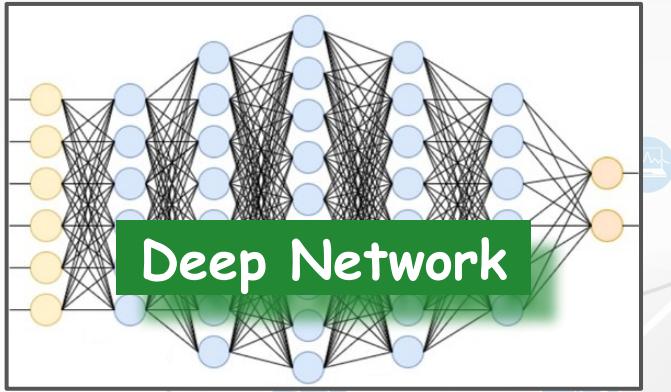
ADA 511



Foundations of

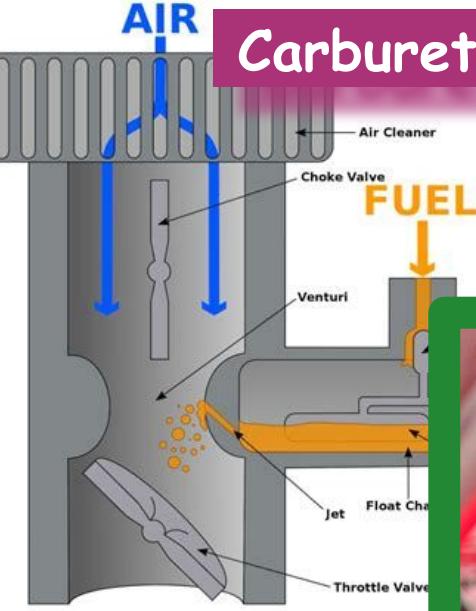
Data Science and data-driven engineering

$$P(Y|X)$$



AIR

Carburettor

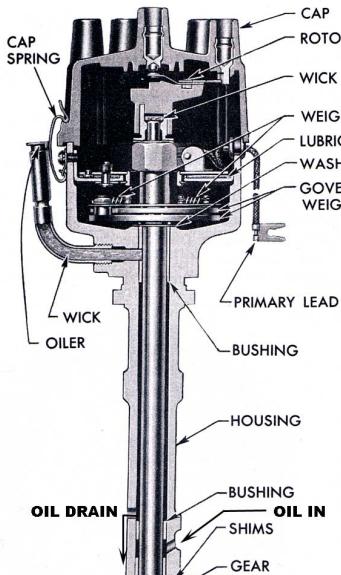


Fuel



Octanes

car mechanic



Distributor



Spark plug

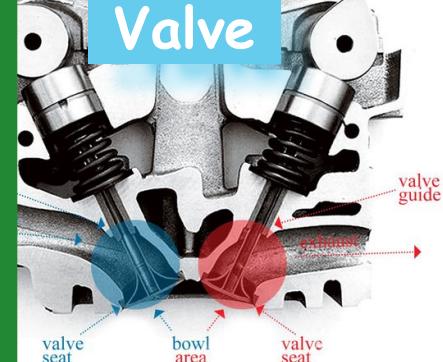
Engine oil



Silencer (muffler)



Green emissions



Valve

 cartwright



 car mechanic



 EV mechanic

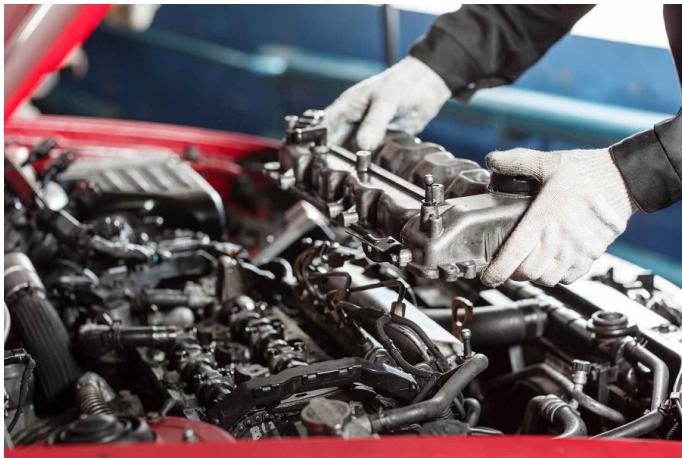


automotive engineer



?

car mechanic



🔍 *specific*

🔨 *practical*

☰ *procedural*

⚓ *technology-dependent*

...

automotive engineer



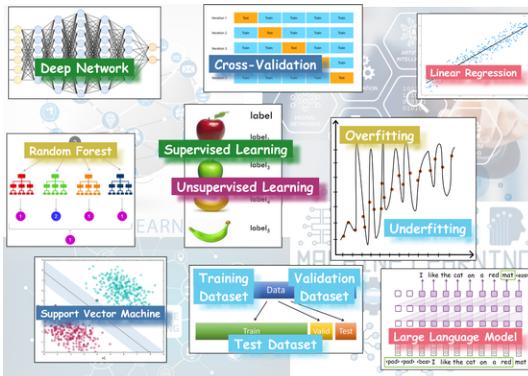
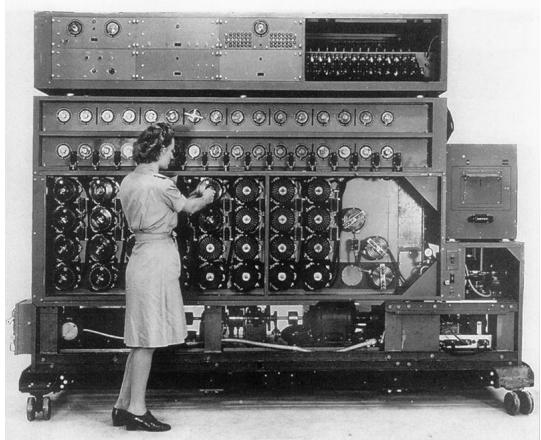
💡 *general*

Ἁ *principled*

🧭 *exploratory*

🐦 *technology-independent*

...



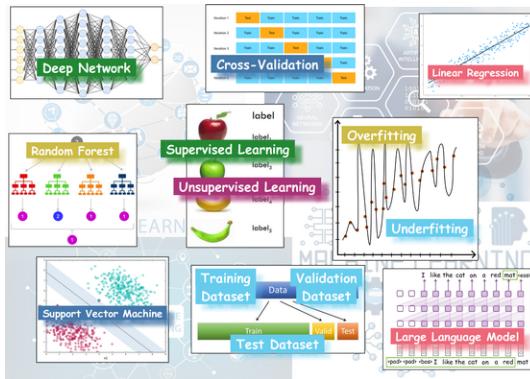
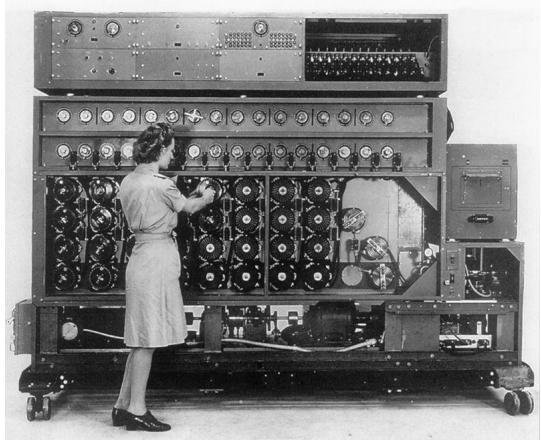
🔧 “data mechanic”

cartwright



🔌 EV mechanic





“data mechanic”

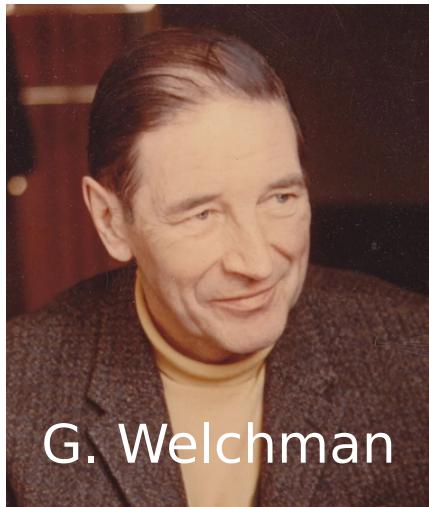


“data engineer”

[insert picture of you all here?]



Alan Turing



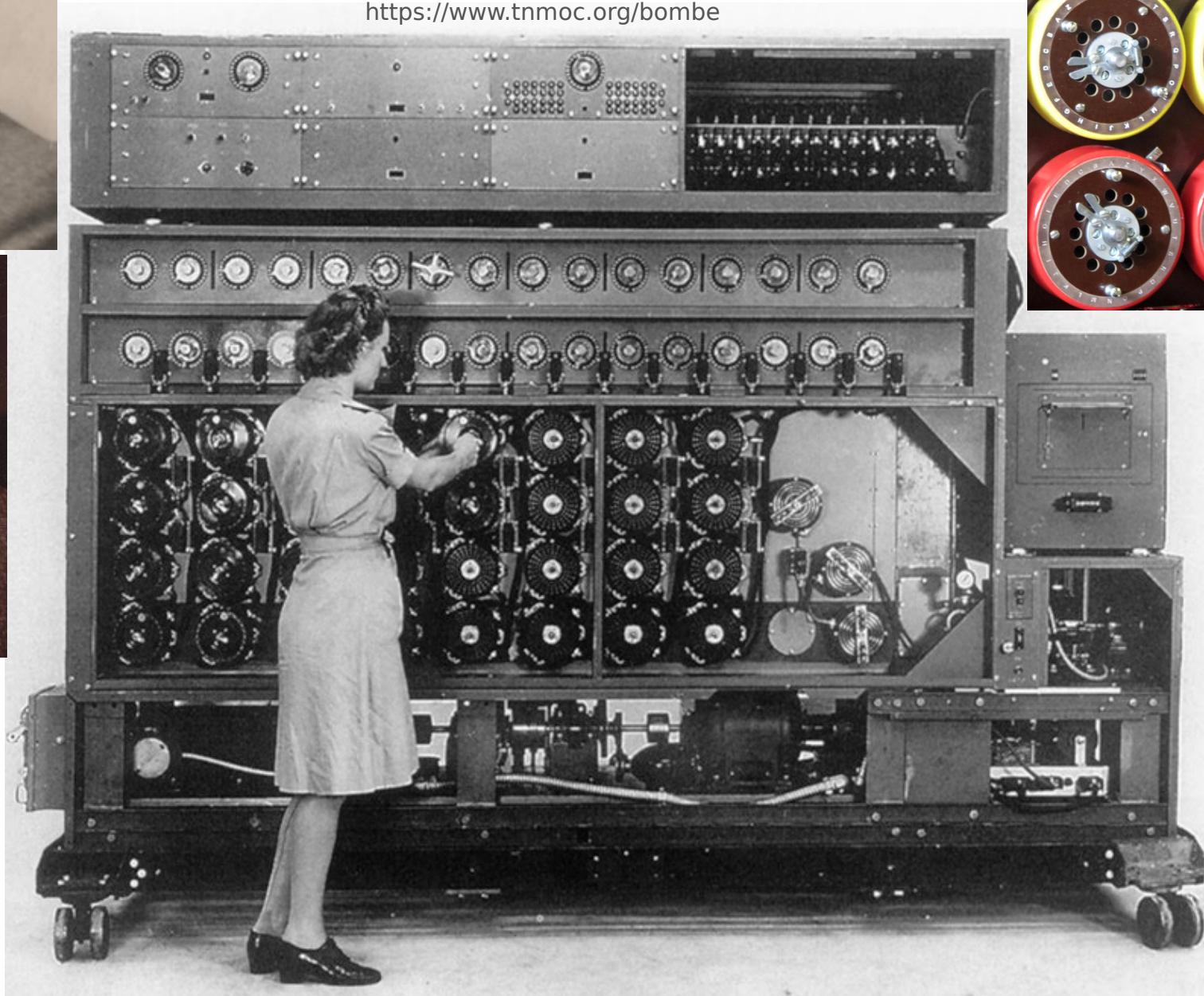
G. Welchman

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

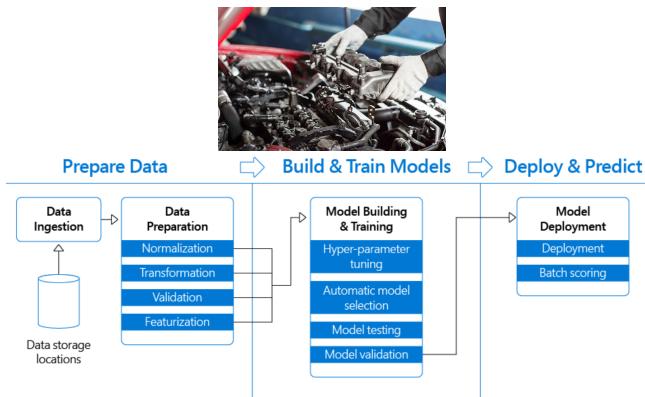
Decoded encrypted Nazi communications during World War II
(efficiency rate: 98.8%)

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

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



“data mechanic”



🔍 *specific*

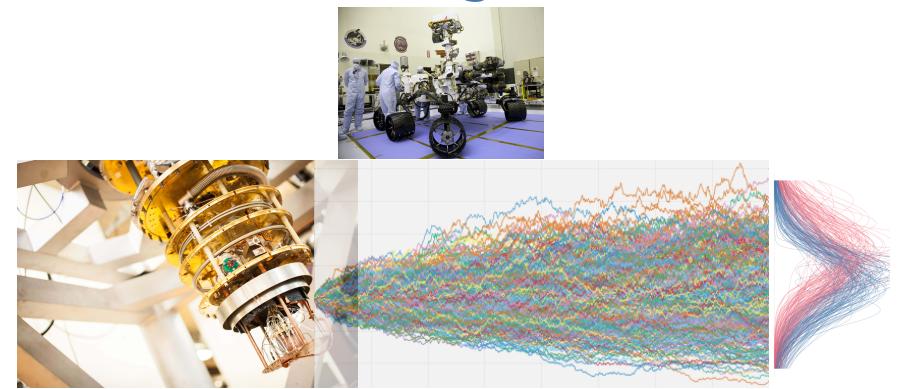
🔨 *practical*

Ξ *procedural*

⚓ *technology-dependent*

...

“data engineer”



👀 *general*

📘 *principled*

🧭 *exploratory*

🐦 *technology-independent*

...



problem-dependent

i metadata

$$P(\neg Y | X) = 1 - P(Y | X)$$

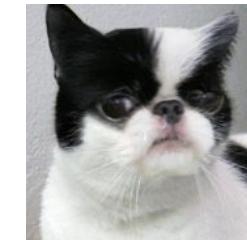
$$P(Z \wedge Y | X) = P(Z | Y \wedge X) \cdot P(Y | X)$$

$$P(Z \vee Y | X) = P(Z | X) + P(Y | X) - P(Z \wedge Y | X)$$

data



$$\text{choose } \arg \max_D U(D | Y \wedge X) \cdot P(Y | D \wedge X)$$



inferences



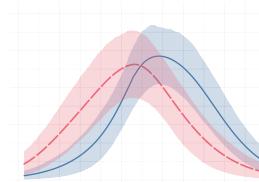
70% $\pm 7\%$

30% $\pm 11\%$

- most likely cat
- however, treat as dog

\$⌚ utilities
(gains/losses)

~~x~~ cat $\$ \$$
✓ dog $+ \$$



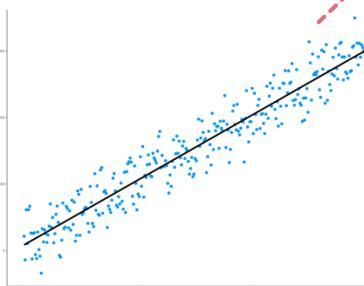
\approx approx

\approx approx

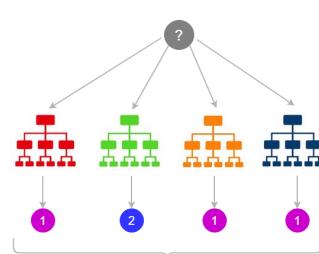
\approx approx

\approx approx

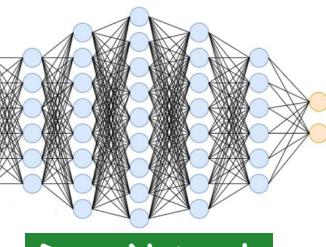
\approx approx



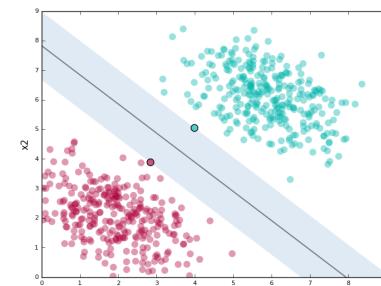
Linear Regression



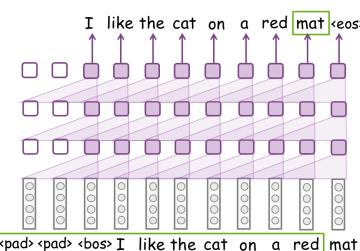
Random Forest



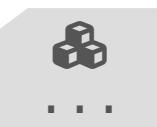
Deep Network



Support Vector Machine



Large Language Model



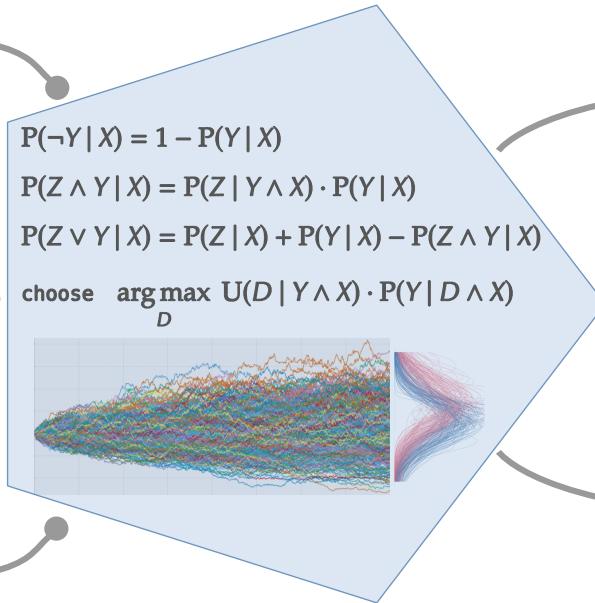
Optimal Predictor Machine

problem-dependent

i metadata

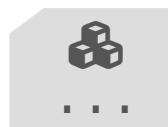
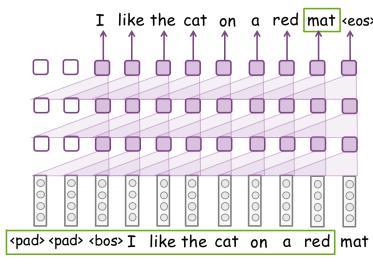
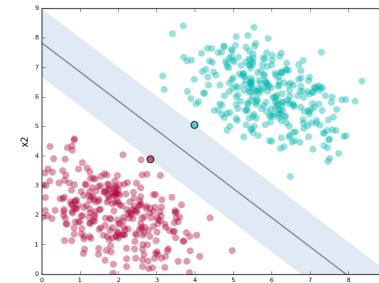
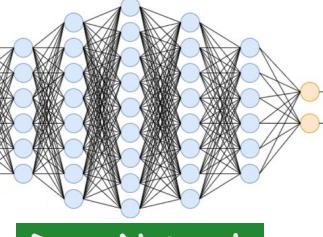
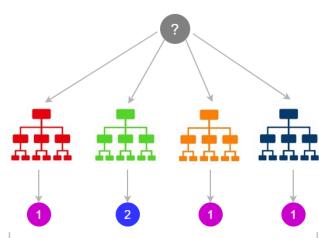
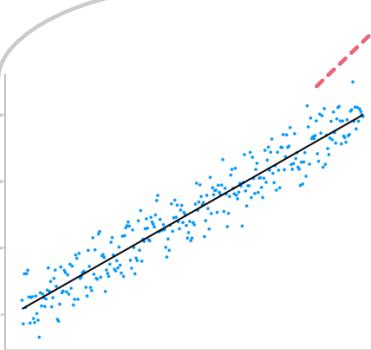
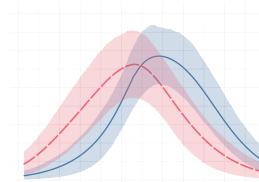
data

utilities
(gains/losses)



- most likely cat
- however, treat as dog

✗ cat \$\$\$
✓ dog +\$



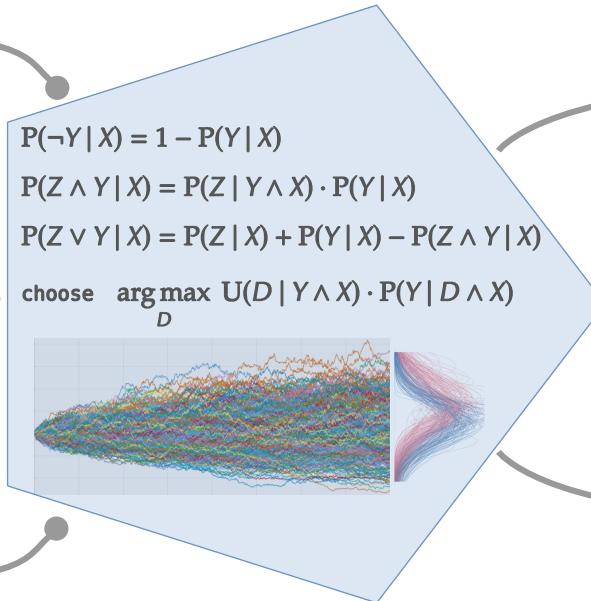
Optimal Predictor Machine

problem-dependent

i metadata

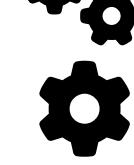
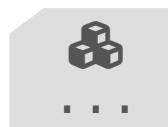
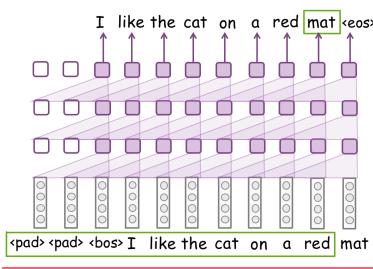
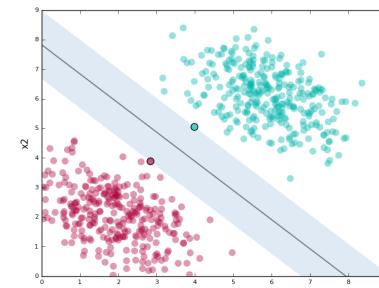
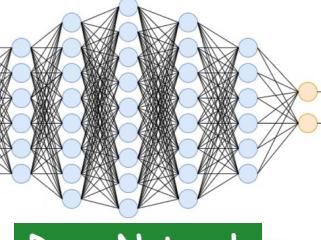
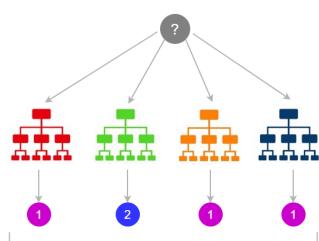
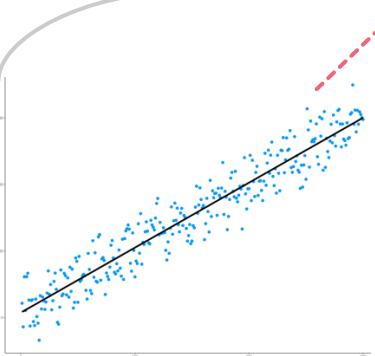
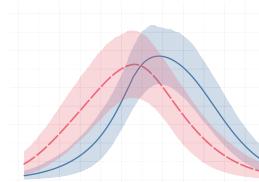
data

\$⌚ utilities
(gains/losses)



- most likely cat
- however, treat as dog

✗ cat \$\$\$
✓ dog +\$



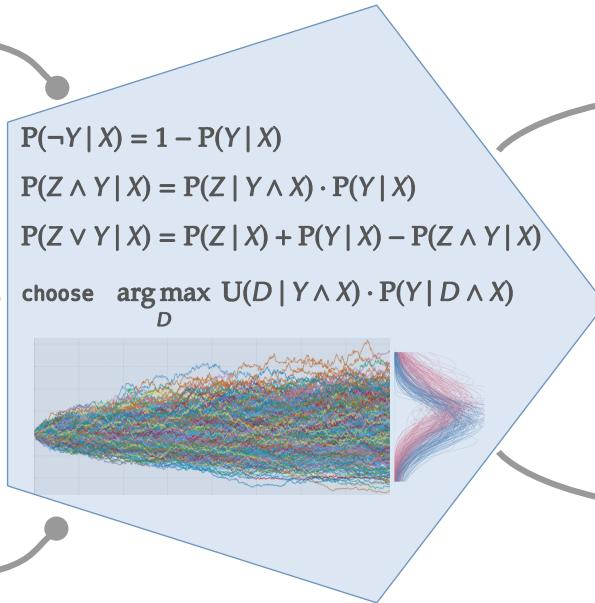
Optimal Predictor Machine

problem-dependent

i metadata

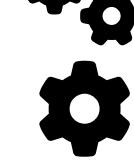
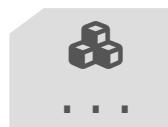
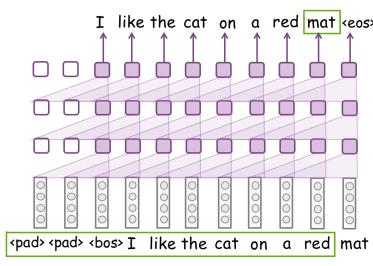
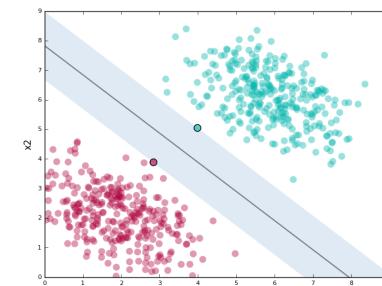
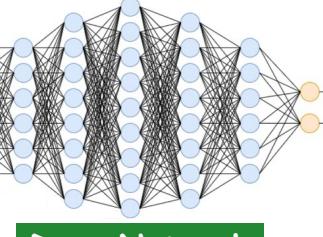
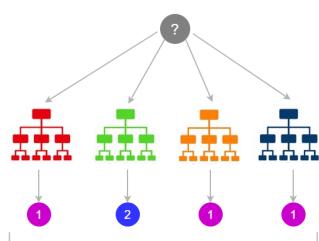
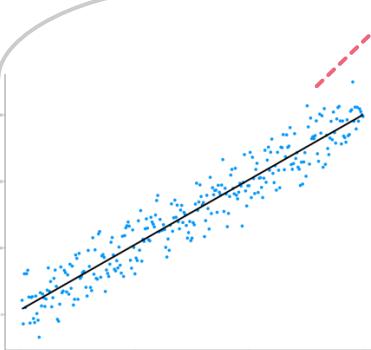
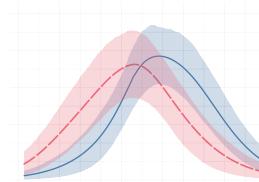
data

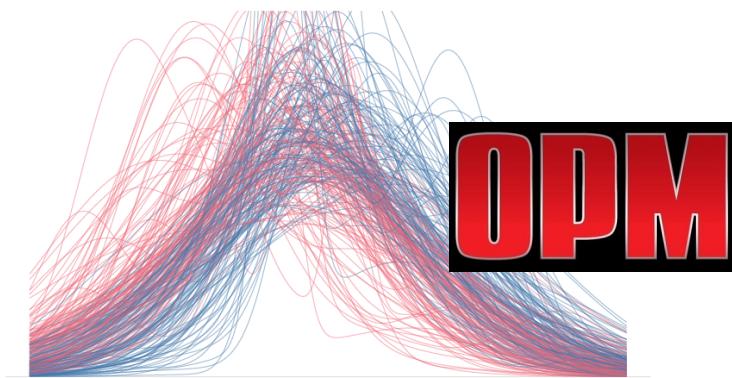
\$⌚ utilities
(gains/losses)

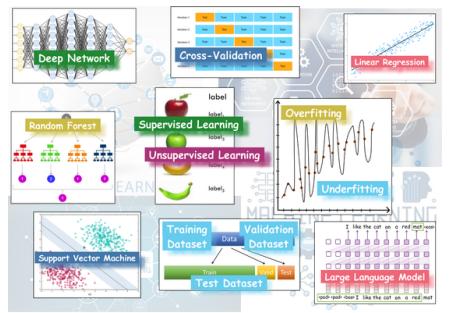


- most likely cat
- however, treat as dog

✗ cat \$\$\$
✓ dog +\$



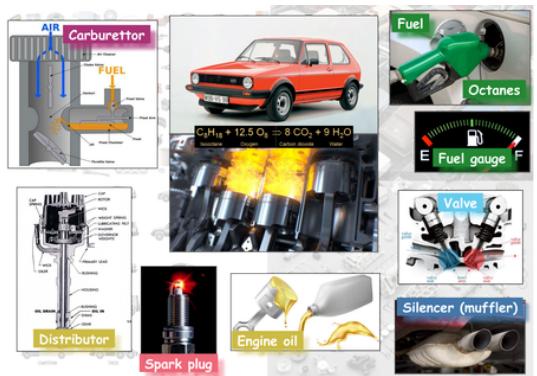




cartwright



car mechanic



EV mechanic

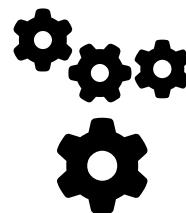




: -\$
 : +\$



2



Optimal Predictor Machine

Optimal Predictor Machine

Optimal Predictor Machine

Optimal Predictor Machine

Optimal Predictor Machine



Optimal Predictor Machine

Optimal Predictor Machine

Optimal Predictor Machine