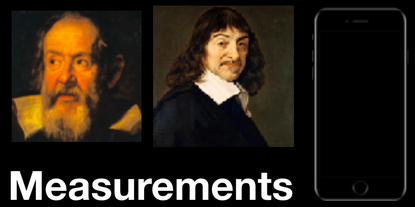
From Data to Learning

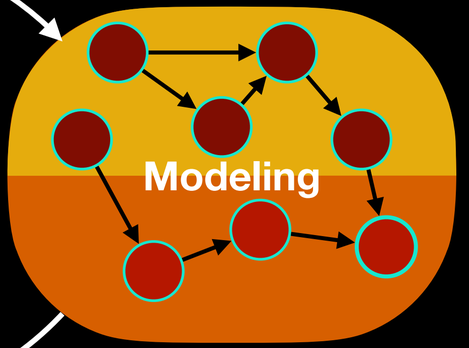
Compression



Measurements

Learning

Data



Modeling

Seen

Unseen

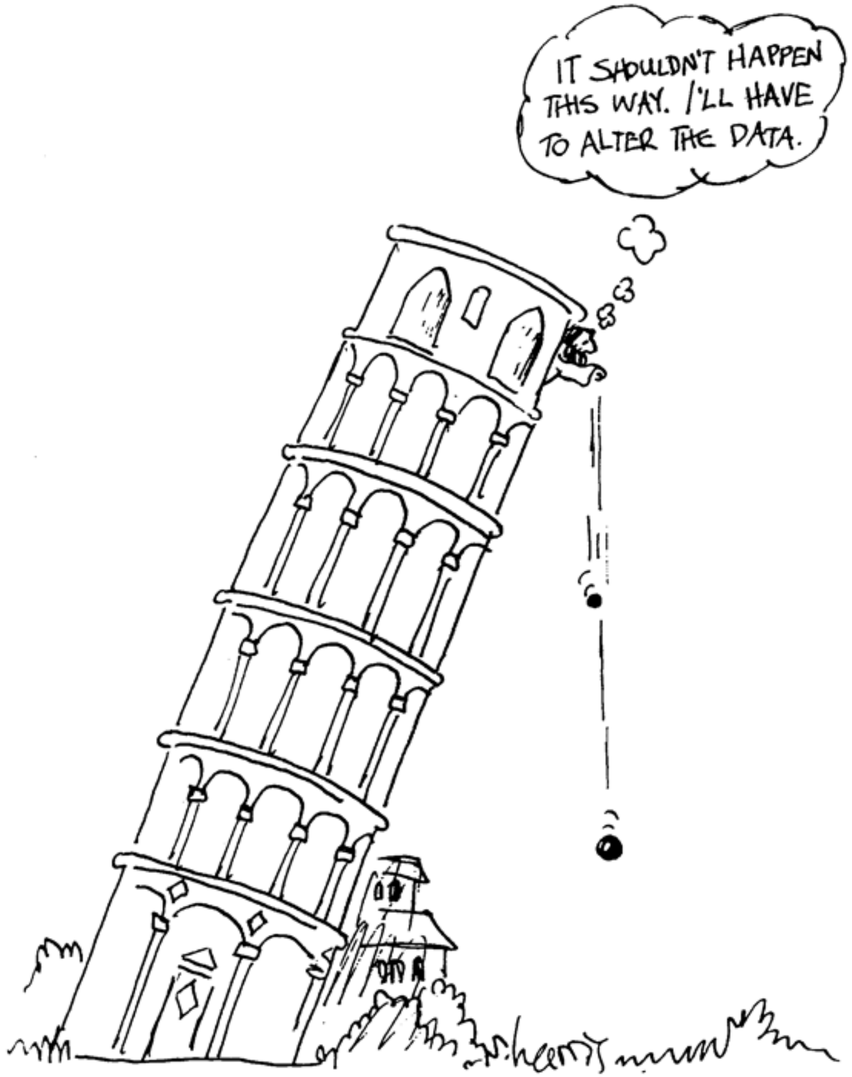
Inference

Projection

World

Reconstruction

Finding *constants* of nature that *generalize* across space and time



Galileo


676 EPITOMES ASTRONOMIÆ

LIBRI V.

PARS ALTERA.

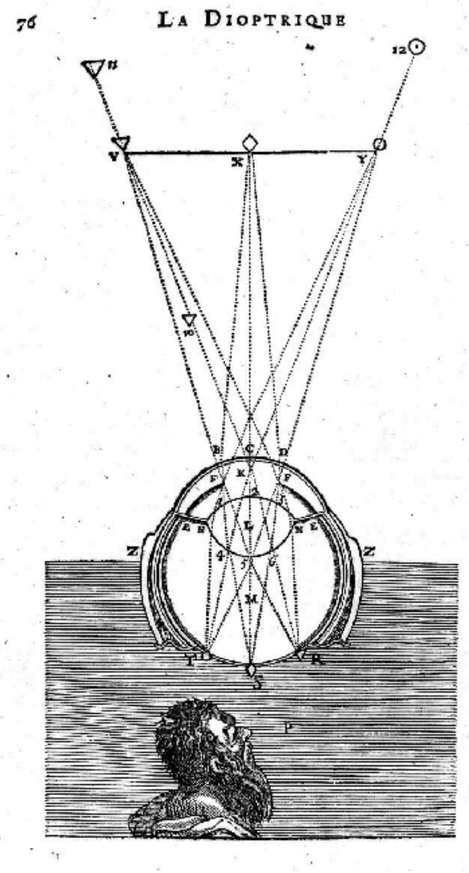
DE TERMINIS ASTRONOMICIS ET CALCVLO, EX ORBITA Eccentrica orientibus.

Quomodo appellatur Orbita cuiusque Planetæ?

APpellatur veteri voce Eccentricus subaudi, Circulus. Etsi enim orbitæ sunt Ellipticæ, vt hic PERI. quæ habent duo quasi centra A L. quæ physicè Foci os dicimus; & in eorum altero A. Sol ipse vt centrum mundi, inest: tamen etiam

punctum inter focos medium, vt B. à scriptoribus Conicis centrum figuræ, peculiari iure dicitur, & præter-

Kepler



Descartes

Empirical Laws are linear

Pascal's law (1653)

$$\Delta p = \rho g \Delta h$$

Hooke's law (1678)

$$F = -kx$$

Newton's law of viscosity (1701)

$$\tau = \mu \frac{du}{dy}$$

Ohm's law (1781)

$$I = V/R$$

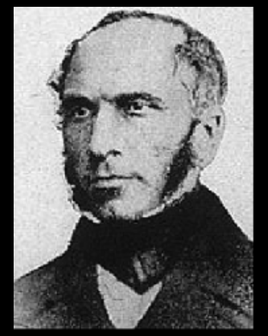
Fourier's law (1822)

$$q = -k \frac{dT}{dx}$$

Fick's law (1855)

$$J = -D \frac{dC}{dx}$$

Darcy's law (1856)



$$Q = \frac{kA}{\mu L} \Delta p$$

Ideal gas law (1834)

Amonton's law (1808)

Boyle's law (1662)

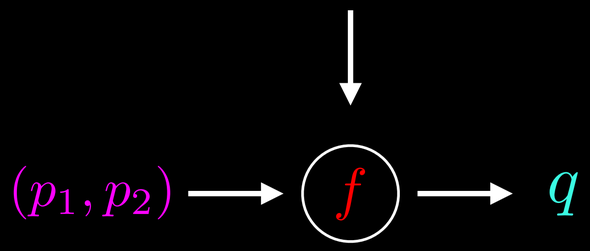
$$\frac{PV}{TN} = k_B$$

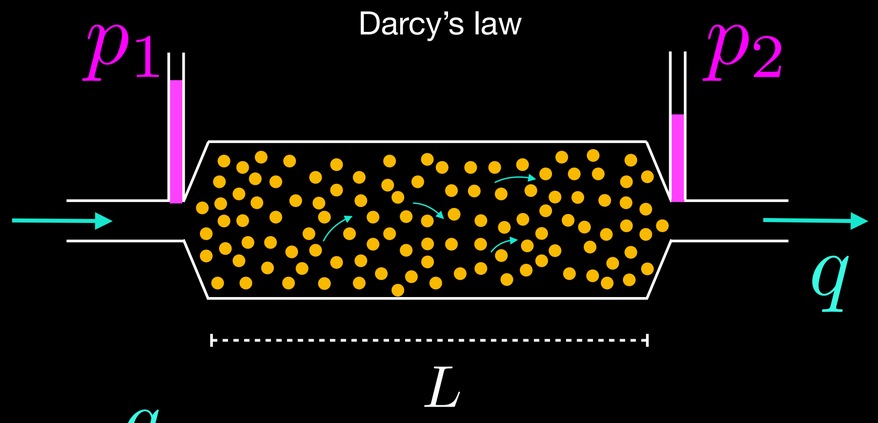
Charles's law (1787)

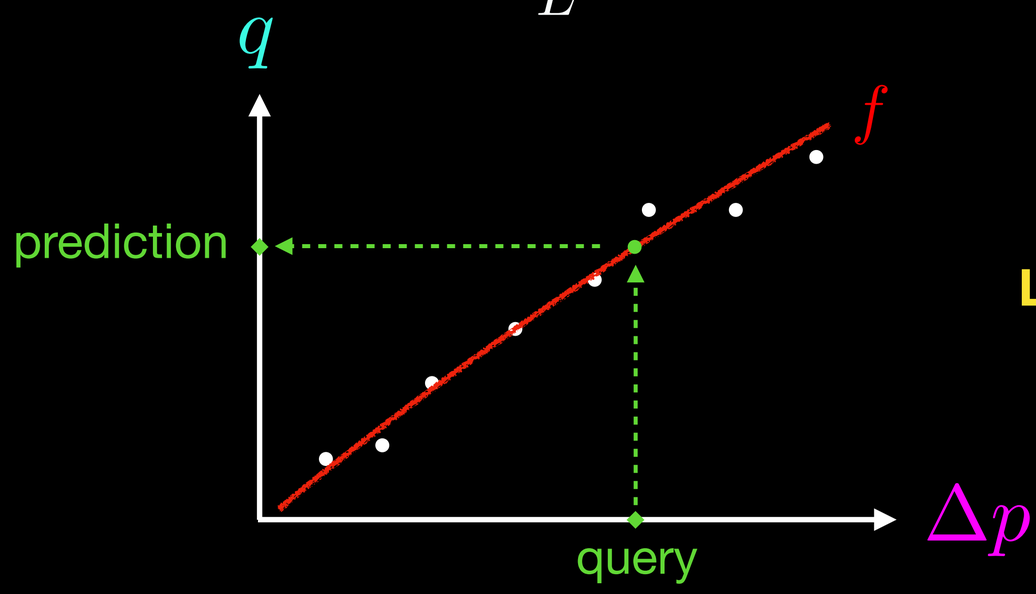
Avogadro's law (1811)

From Experiment to Law

p_1	p_2	q
1.3	1.0	22
1.6	1.5	23
3.4	2.4	46
4.8	3.5	67
6.7	4.5	83
…		
2.3	**1.4**	**?**







Linear regression

Machine or human learning

Training data: $\mathcal{D}_{\text{train}}$

p_1	p_2	q
1.3	1.0	22
1.6	1.5	23
3.4	2.4	46
4.8	3.5	67
6.7	4.5	83
...		

Example →

Training

Feature extractor

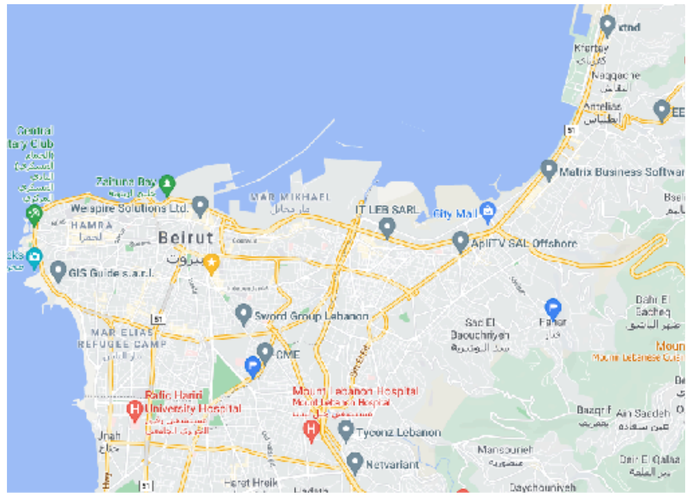
$(p_1, p_2) \rightarrow \phi \rightarrow \Delta p \rightarrow f \rightarrow q$

Predictor *Labels*

- Which predictors are possible?
- How good is the predictor?
- How can we find the best predictor?

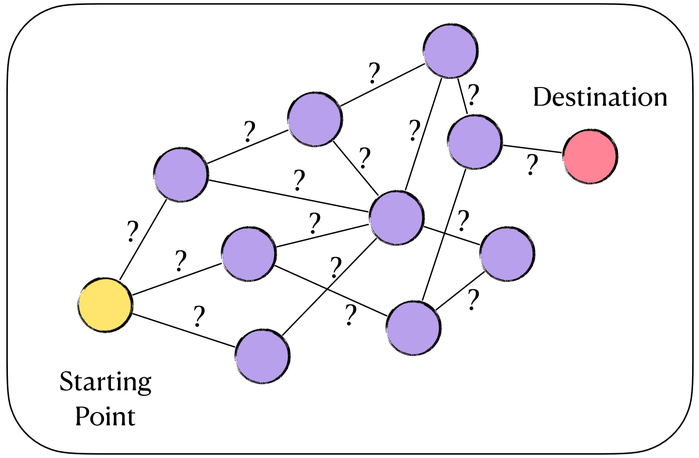
The 3 pillars of Artificial Intelligence

Real world

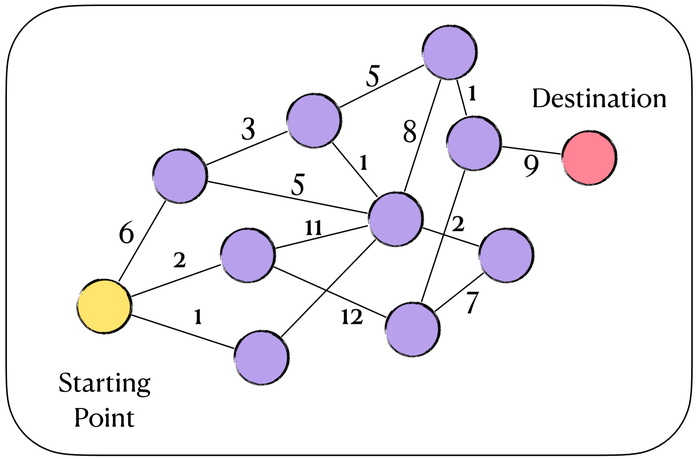


What's the shortest path from AUB to my home?

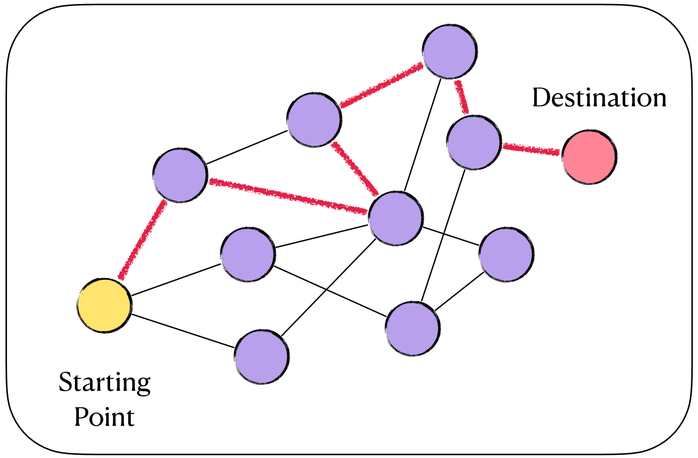
Learning



Modeling



Inference



The 3 pillars of Artificial Intelligence

Real world

Learning

$$\hat{p} = \arg\min_{p} \left\| \frac{\partial u}{\partial t} - \mathscr{L}(u(x,t);p) \right\|_2^2$$

Inference

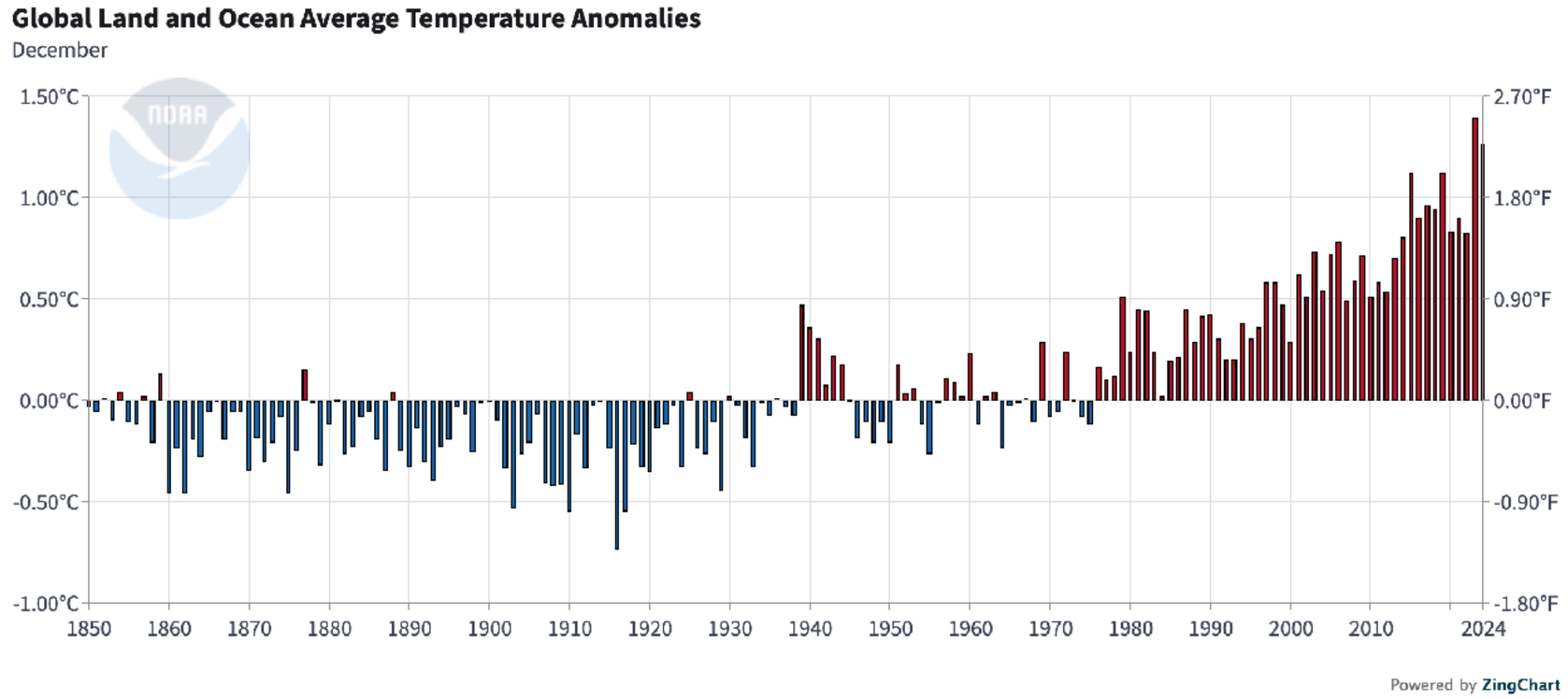
$$u(x_i, t_i) = \text{Integrator}\left(u(x_i, t_{i-1}), u(x_{i+1}, t_{i-1}), u(x_i, t_{i-1})\right)$$

Modeling

$$\frac{\partial u}{\partial t} = \mathscr{L}(u(x,t);p)$$

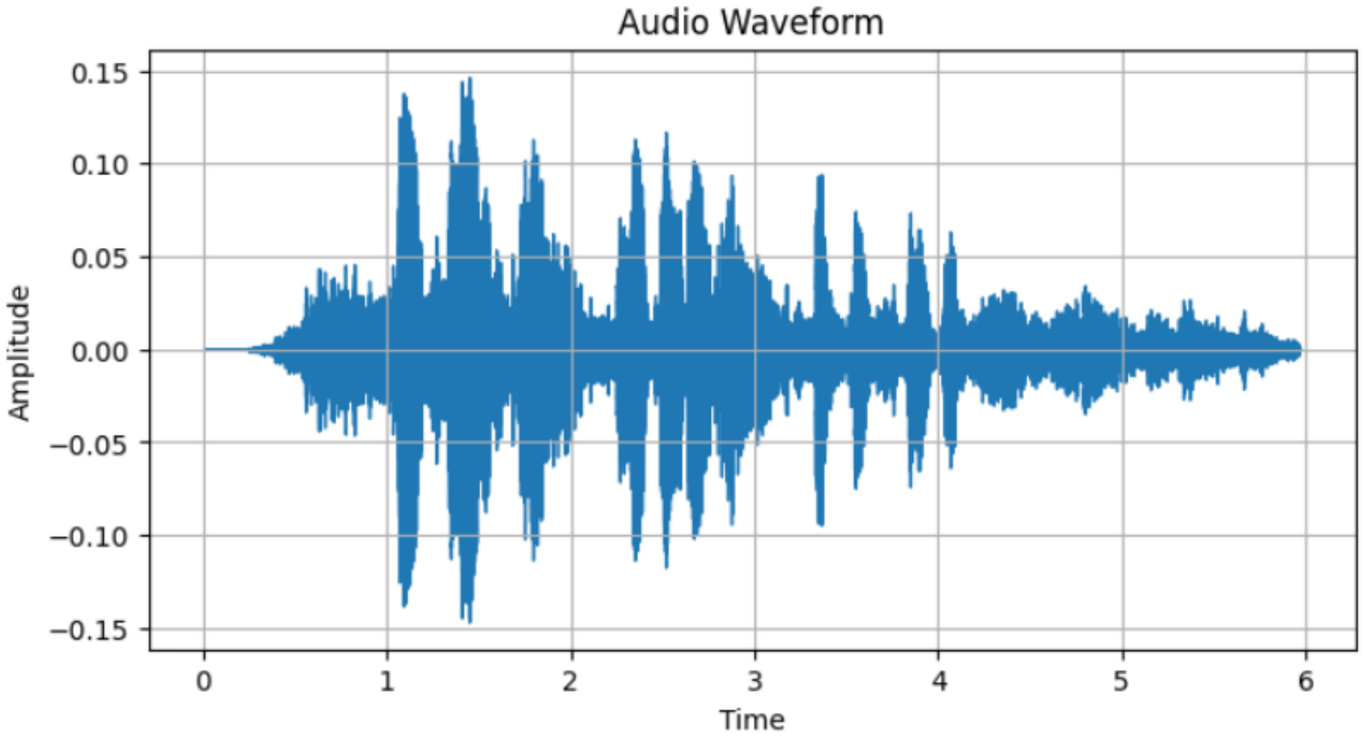
What **are*** Data?

Stocks



Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Temperature	0	1	-1	-2	-3	0	1	2	2	3	5	6	8	5	4	6

Audio



Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Amplitude	35	37	35	34	33	30	49	48	46	44	46	49	48	66	50	55

Images

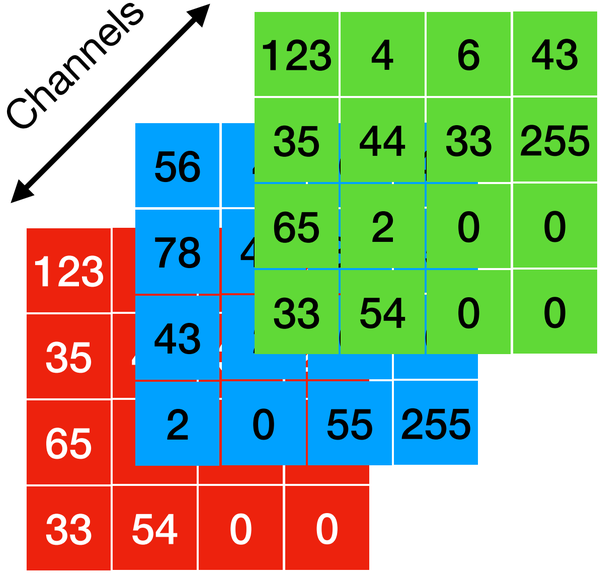


Image dimension: $h \times w \times 3$

Video

dimension: $T \times h \times w \times 3$

t_1

123	4	6	43
35	44	33	255
65	2	0	0
33	54	0	0

t_2

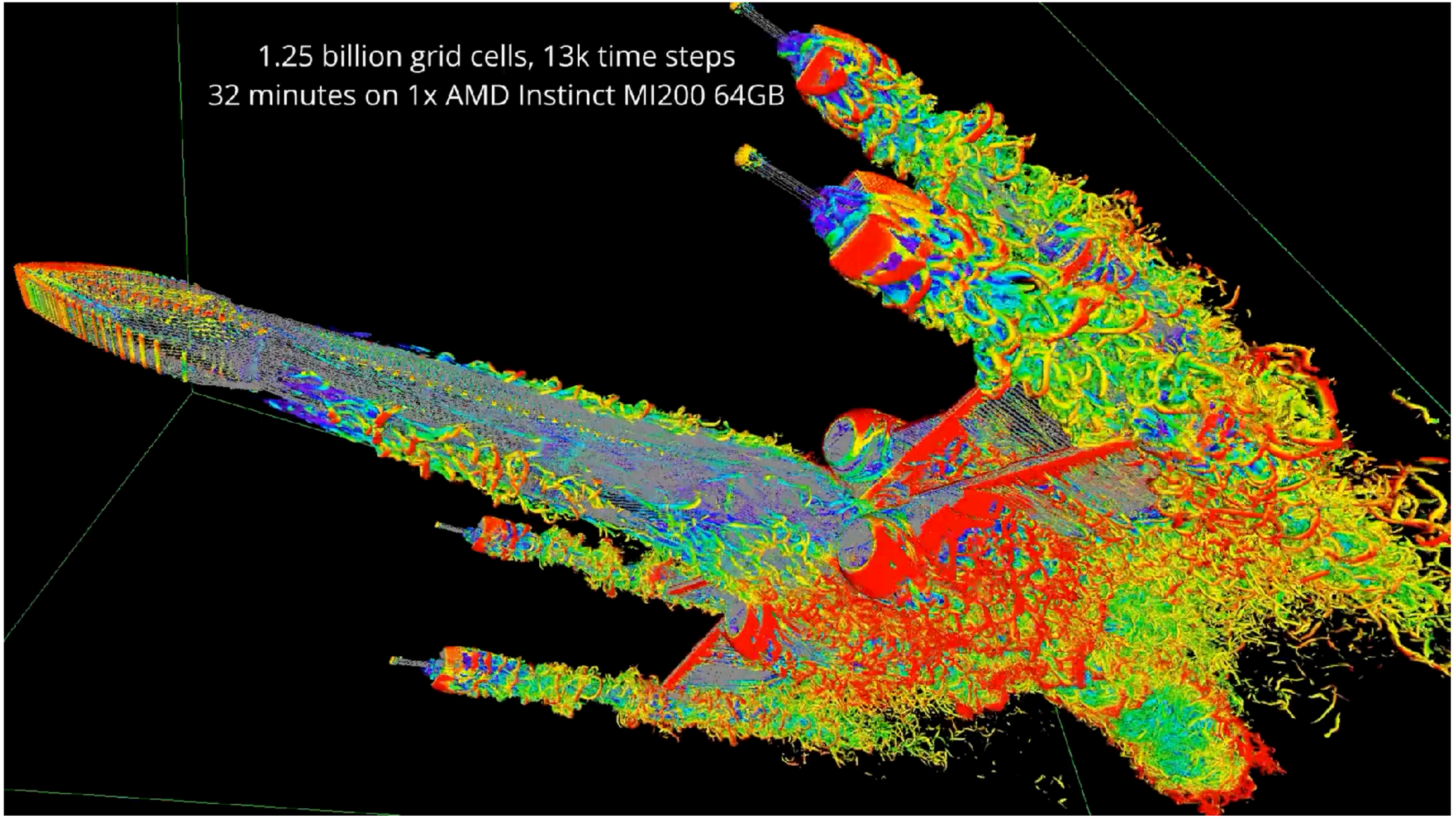
123	4	6	43
35	44	33	255
65	2	0	0
33	54	0	0

. . . . . .

T

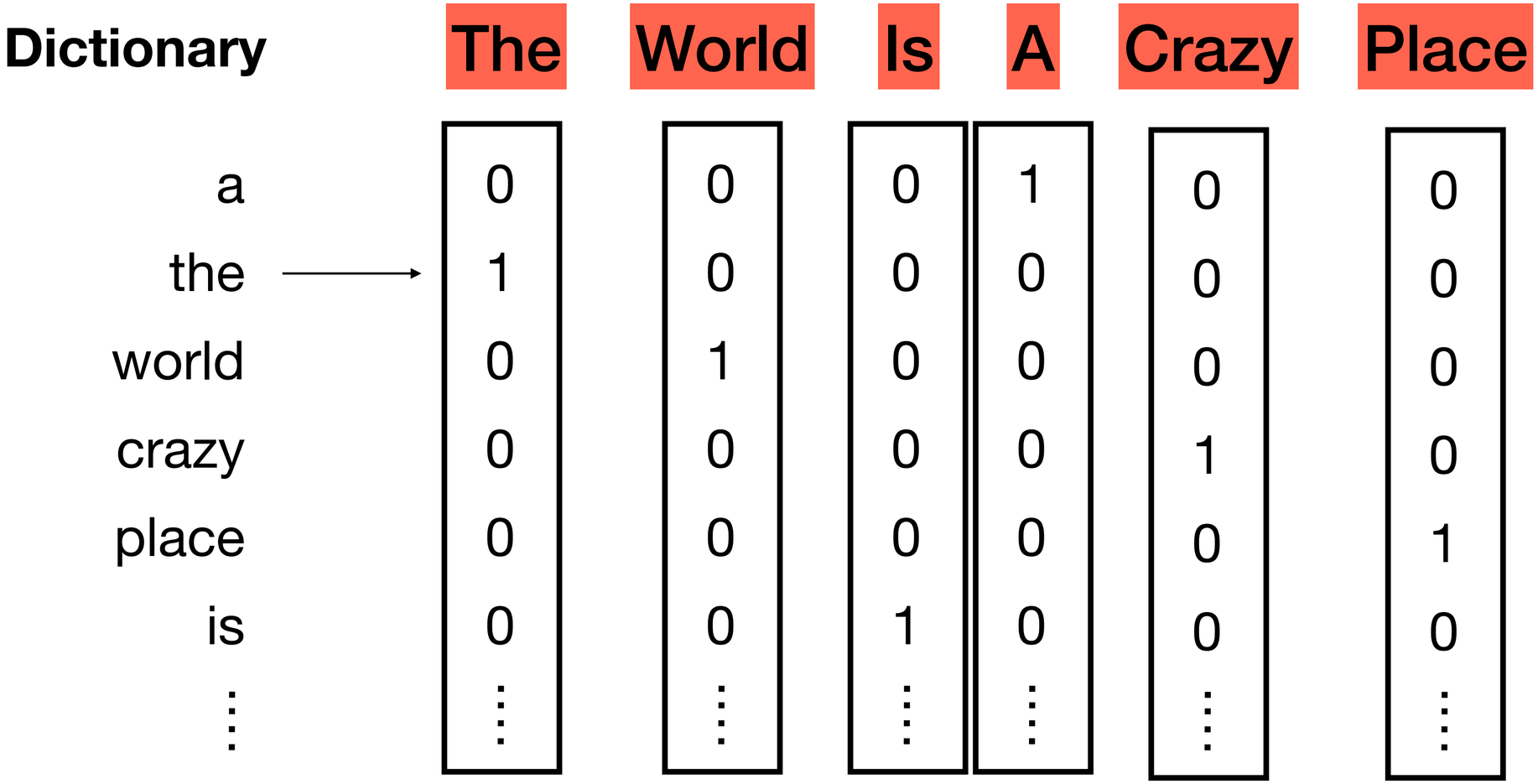
123	4	6	43
35	44	33	255
65	2	0	0
33	54	0	0

Fluid dynamics



Text

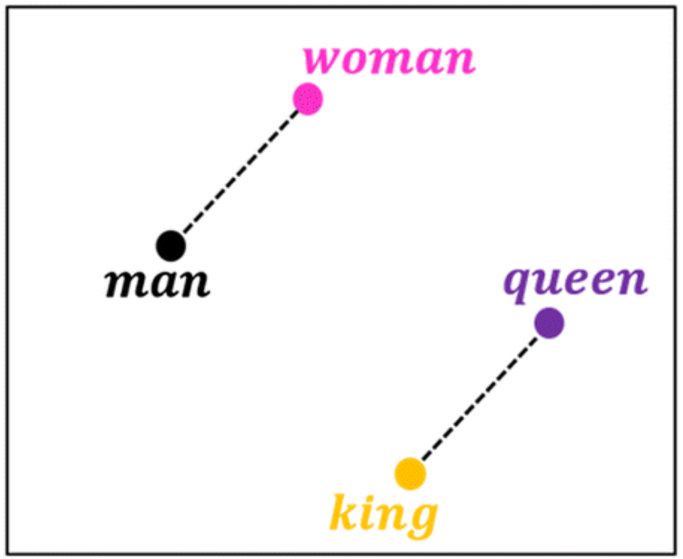
One-hot word representation



Text

Word Embedding Representation

man	→	0.6	−0.2	0.8	0.9	−0.1	−0.9	−0.7
woman	→	0.7	0.3	0.9	−0.7	0.1	−0.5	−0.4
king	→	0.5	−0.4	0.7	0.8	0.9	−0.7	−0.6
queen	→	0.8	−0.1	0.8	−0.9	0.8	−0.5	−0.9



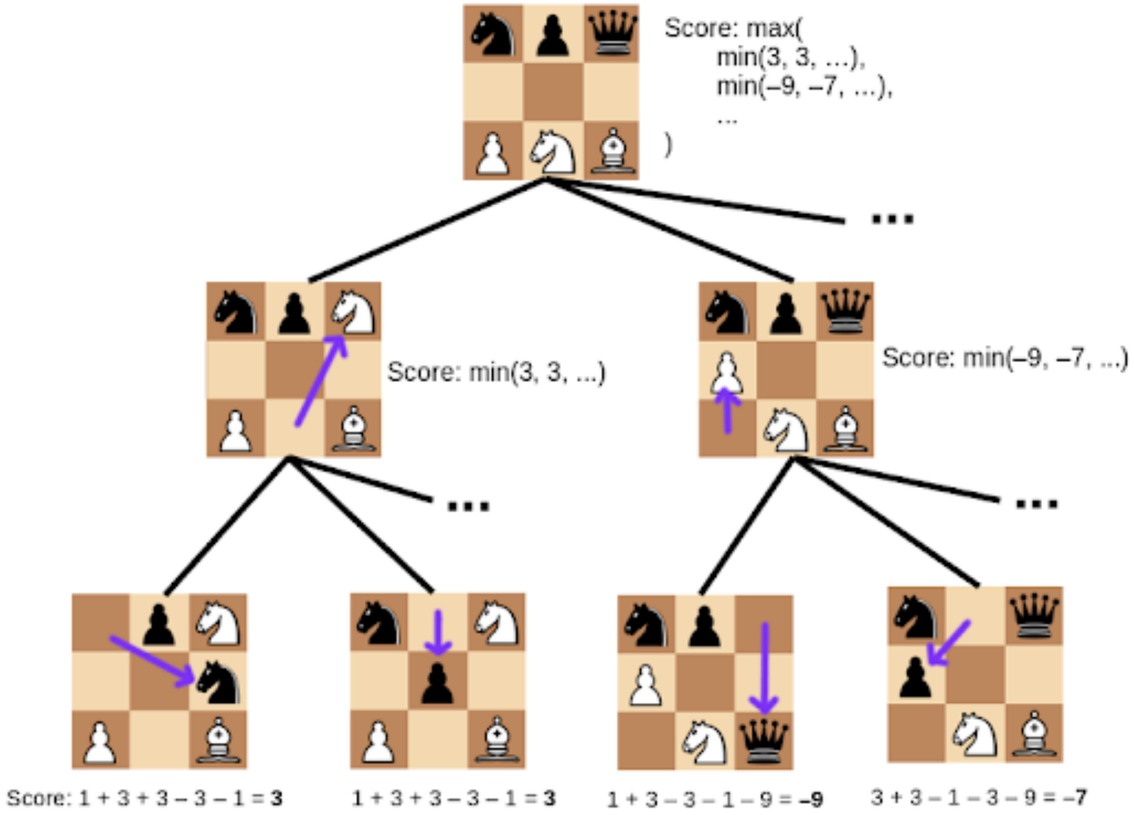
Words

Embedding

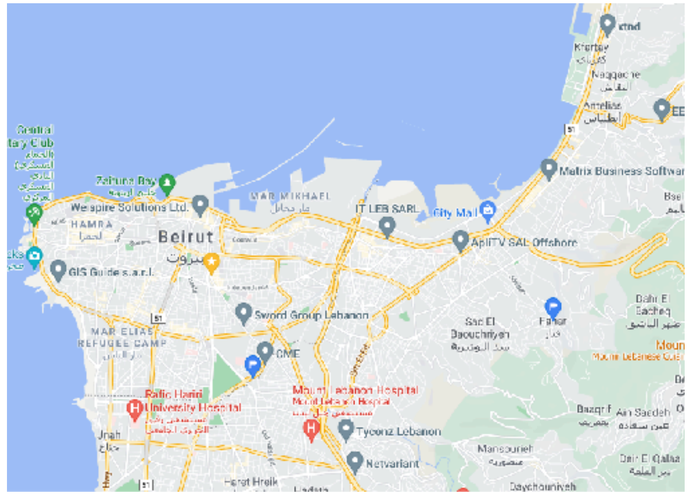
Semantic Space

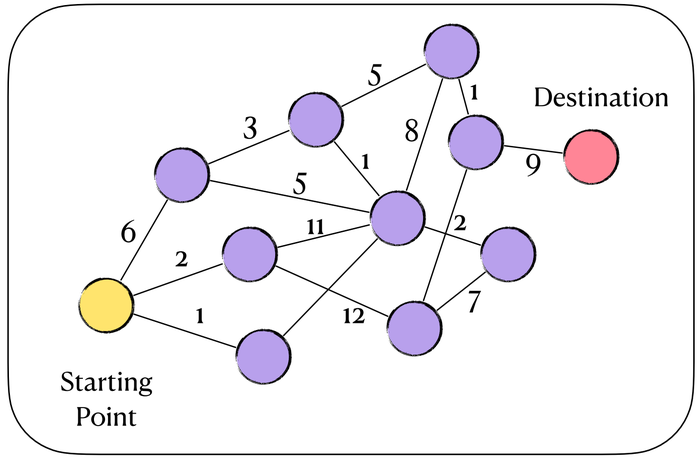
https://www.analyticsvidhya.com/blog/2021/07/feature-extraction-and-embeddings-in-nlp-a-beginners-guide-to-understand-natural-language-processing/

Graph Representation



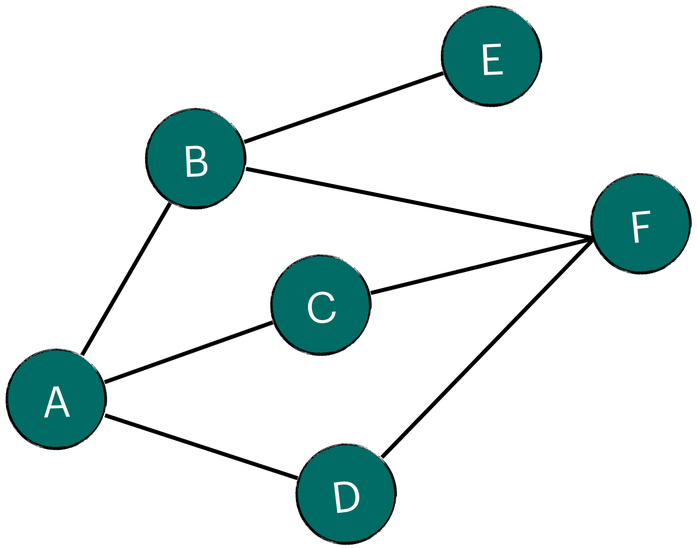
Real world





Graph Representation

Graph with Vertices and Edges



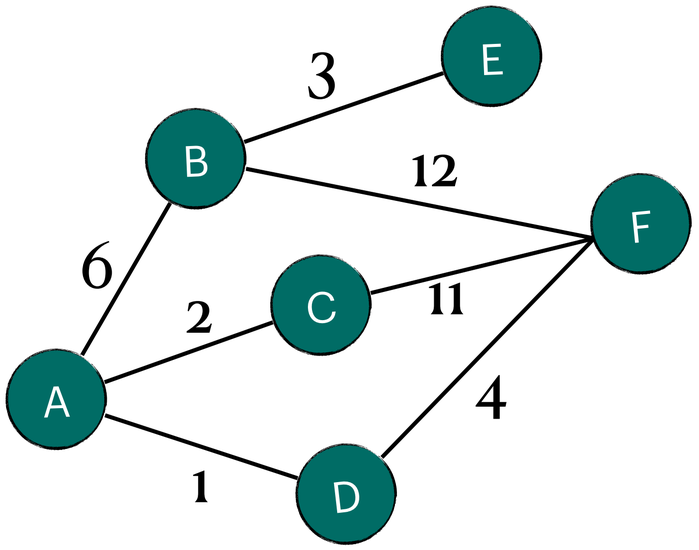
Adjacency Matrix

	A	B	C	D	E	F
A	0	1	1	1	0	0
B	1	0	0	0	1	1
C	1	0	0	0	0	1
D	1	0	0	0	0	1
E	0	1	0	0	0	0
F	0	1	1	1	0	0

https://www.cs.mtsu.edu/~xyang/3080/adjacencyMatrix.html

Graph Representation

Weighted Graph



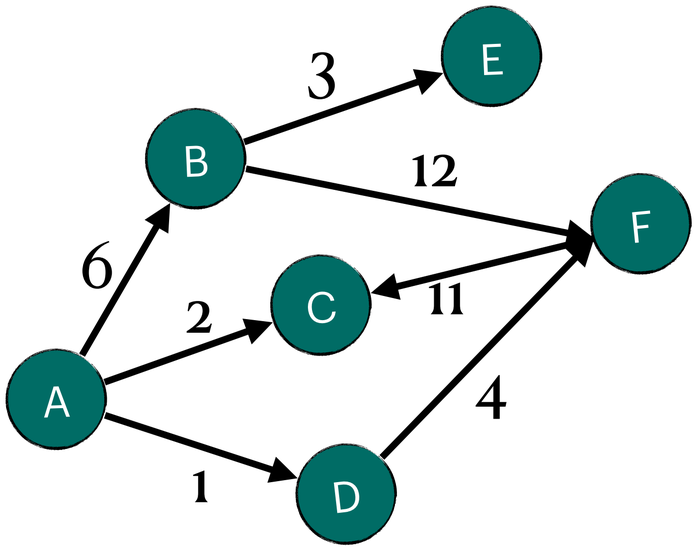
Adjacency Matrix

	A	B	C	D	E	F
A	0	6	2	1	0	0
B	6	0	0	0	3	12
C	2	0	0	0	0	11
D	1	0	0	0	0	4
E	0	3	0	0	0	0
F	0	12	11	4	0	0

https://www.cs.mtsu.edu/~xyang/3080/adjacencyMatrix.html

Graph Representation

Weighted Directed Graph



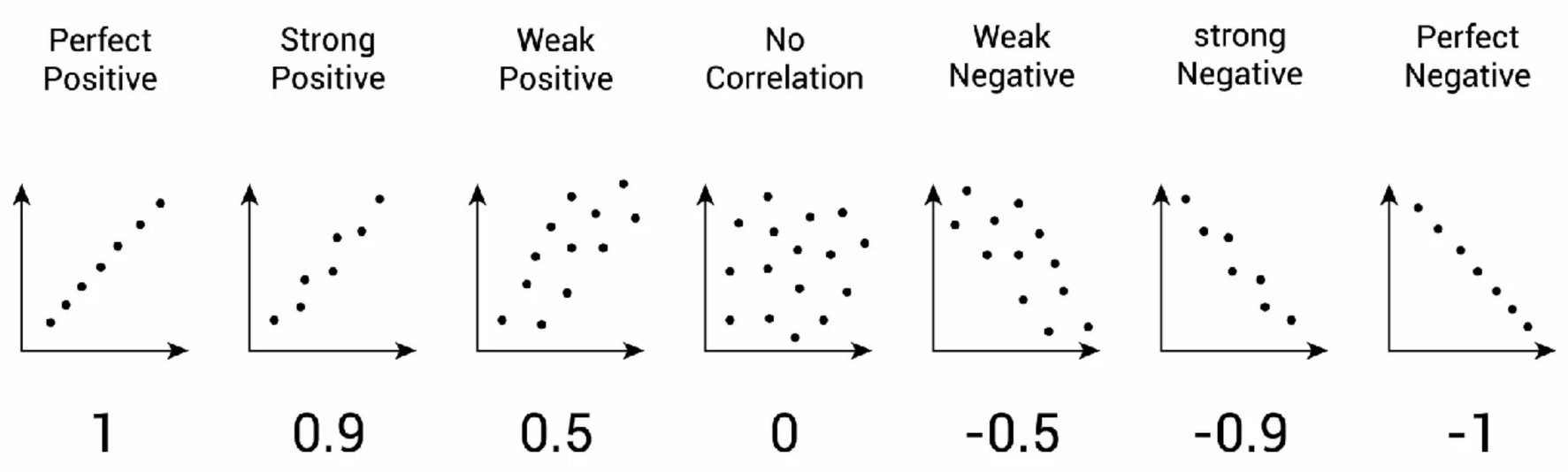
Adjacency Matrix

	A	B	C	D	E	F
A	0	6	2	1	0	0
B	0	0	0	0	3	12
C	0	0	0	0	0	0
D	0	0	0	0	0	4
E	0	0	0	0	0	0
F	0	0	11	0	0	0

https://www.cs.mtsu.edu/~xyang/3080/adjacencyMatrix.html

Data Analysis

Correlation coefficient (2 variables)



$$r_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_summarizingdata/bs704_summarizingdata8.html

Correlation Matrix (multiple variables)

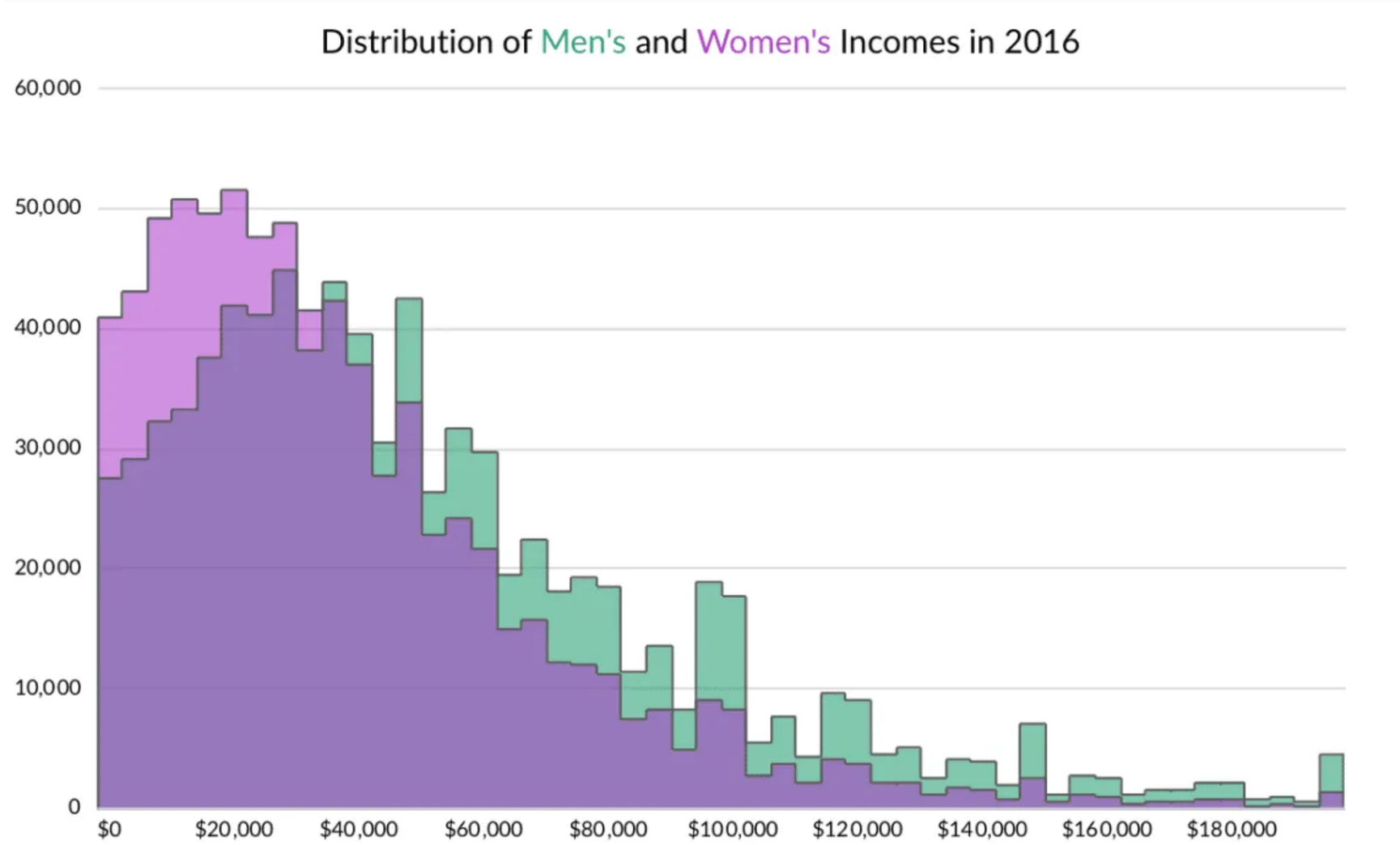
	Hours spent studying	Exam score	IQ score	Hours spent sleeping	School rating
Hours spent studying	1.00	0.82	0.48	-0.22	0.36
Exam score	0.82	1.00	0.33	-0.04	0.23
IQ score	0.08	0.33	1.00	0.06	0.02
Hours spent sleeping	-0.22	-0.04	0.06	1.00	0.12
School rating	0.36	0.23	0.02	0.12	1.00

$$r_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}}$$

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}$$

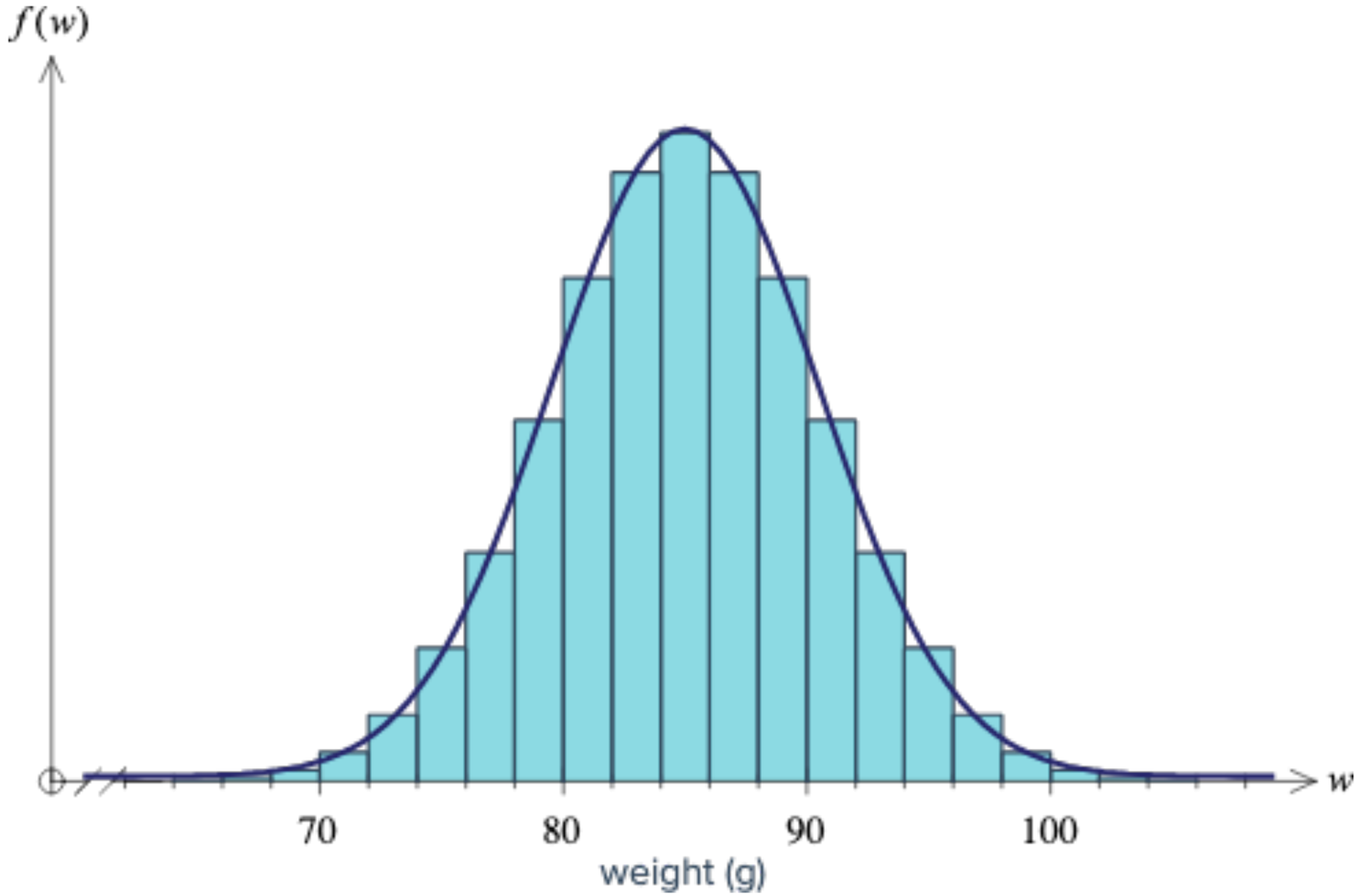
https://www.statology.org/how-to-read-a-correlation-matrix/

Histogram



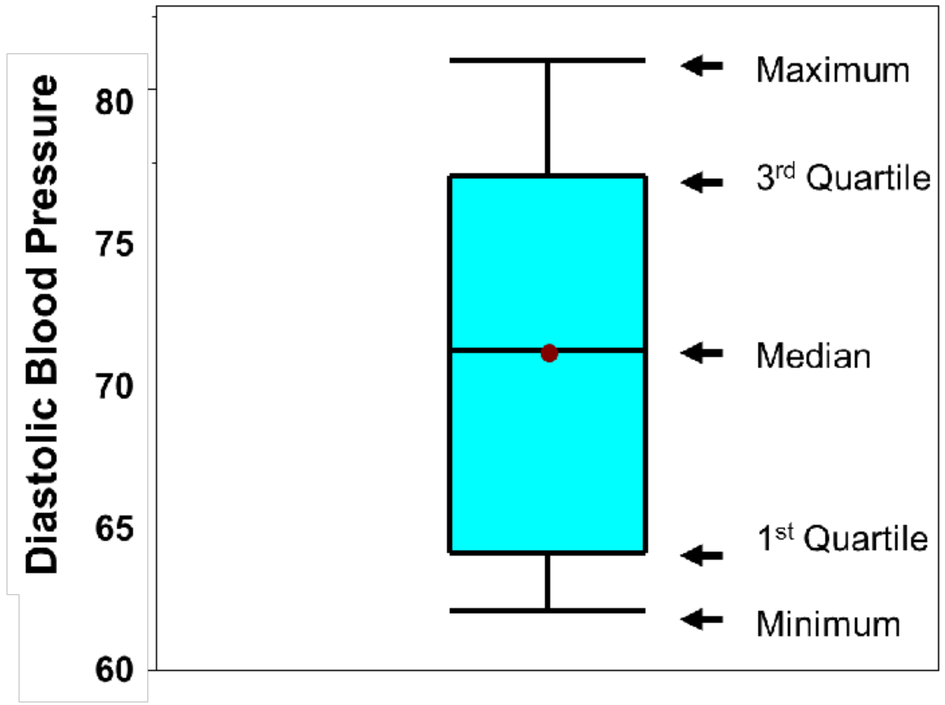
https://policyviz.com/2018/11/27/histogram-design-decisions/

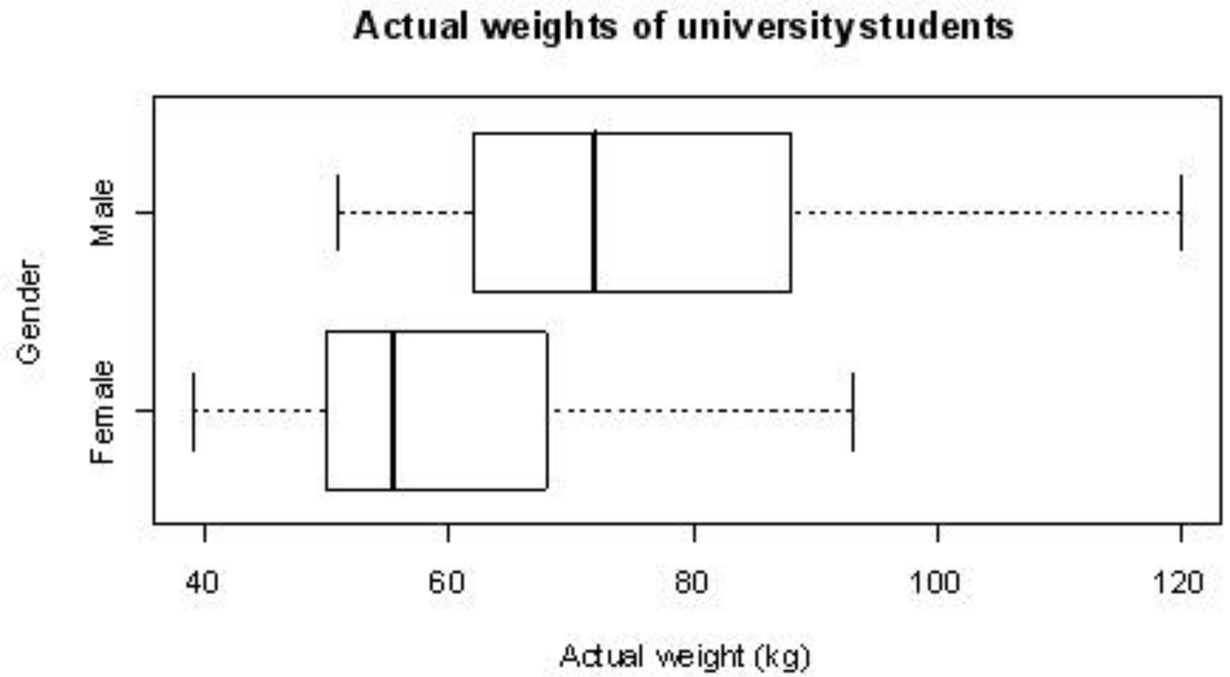
Probability Density Function



https://mathspace.co/textbooks/syllabuses/Syllabus-826/topics/Topic-18358/subtopics/Subtopic-279214/?activeTab=theory

Box and Whisker Plot





https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_summarizingdata/bs704_summarizingdata8.html

Discovering Functions From Data

Machine Learning

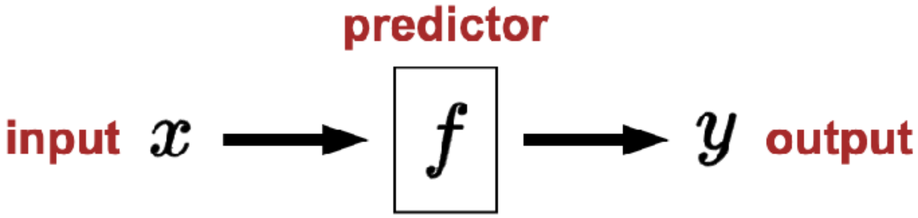
$$f(x) = 2^x$$

x `0, 1, 2, 3, 4, 5`

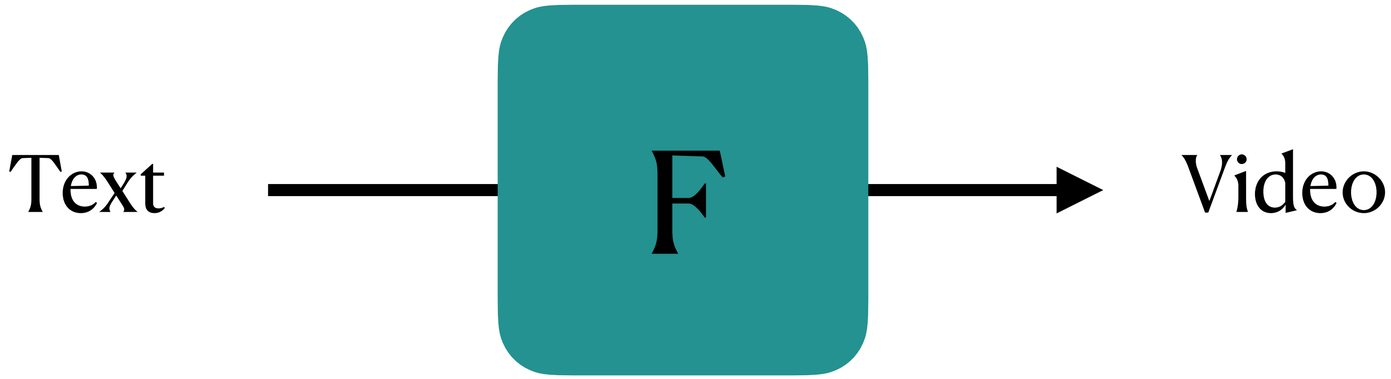
y `1, 2, 4, 8, 16, ?`

$x \quad y$

$$f(x) = 2x$$



Sora - Text-to-Video

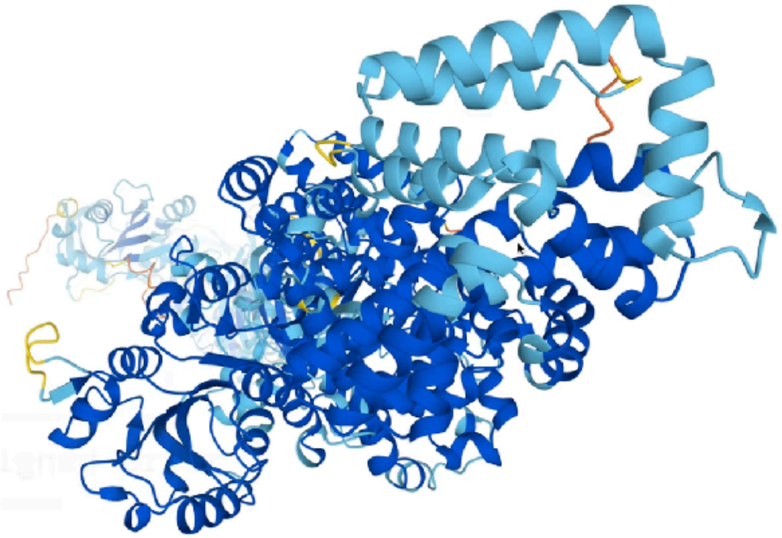


ChatGPT

AlphaFold

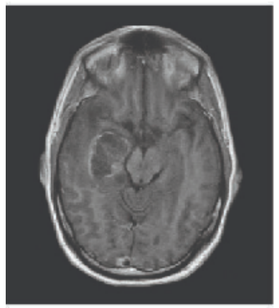
AlphaFold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.

Amino Acid Sequence → F → 3D protein structure



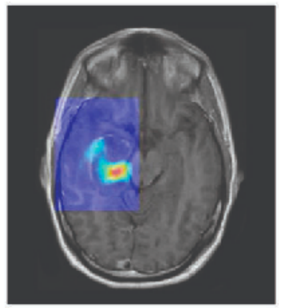
Healthcare

CT Scan



F

Cancer



Healthcare

CT S ancer

Mindaugas Galvosas, MD @MGalvosas · Jan 18

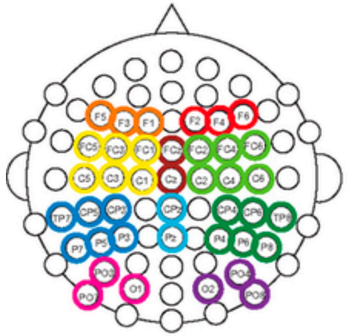
FDA just cleared the first AI device detecting all major skin cancer by DermoSensor.

Its pivotal trial showed sensitivity of 96% and a 97% chance of accurately identifying a skin lesion as benign.

Notably, the approval process began back in 2016.

42 248 1K 257K

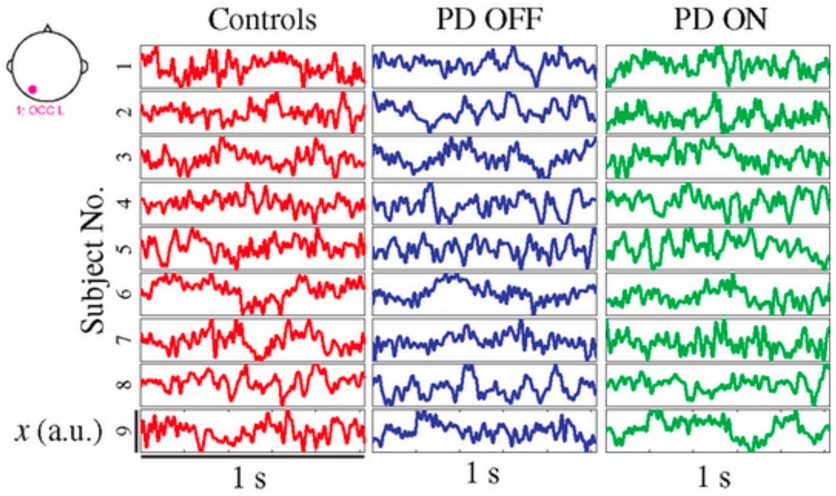
Neuroscience



EEG

F

Behavior
Mechanism
Disease

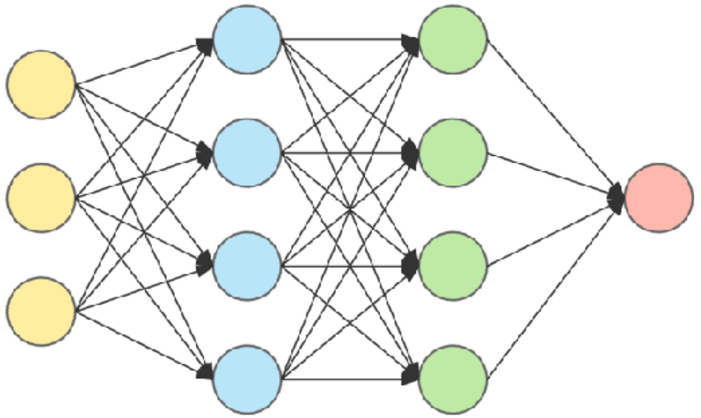


Lainscsek, Claudia, et al. "Non-linear dynamical analysis of EEG time series distinguishes patients with Parkinson's disease from healthy individuals." *Frontiers in neurology* 4 (2013): 200.

Neural Networks

Input

Output

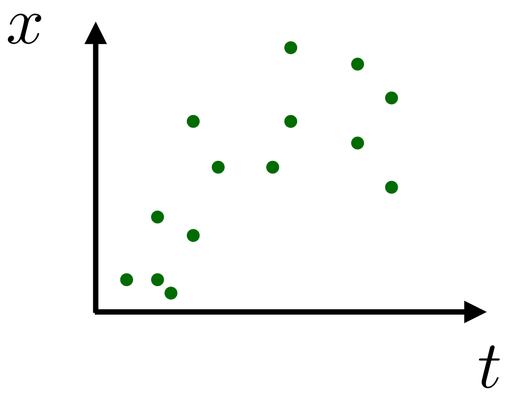


$$f \circ f$$

Applications in Scientific Computing

Symbolic Regression

Time series



F

Equation

$$\frac{dx}{dt} = f(x)$$

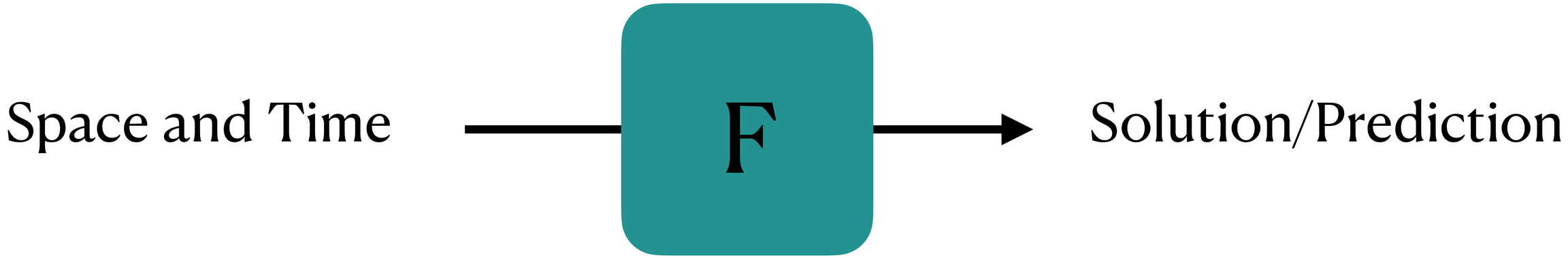
+ Physical Constraints

e.g. Conservation of Energy

???

Applications in Scientific Computing

Physics Informed Neural Networks



+ Physical Constraints

e.g. Conservation of Energy

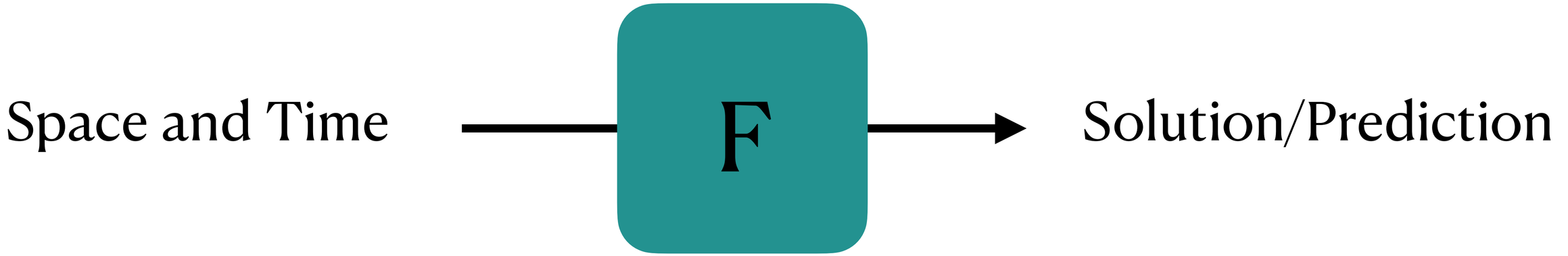
???

Google Research BLOG ›

A decoder-only foundation model for time-series forecasting

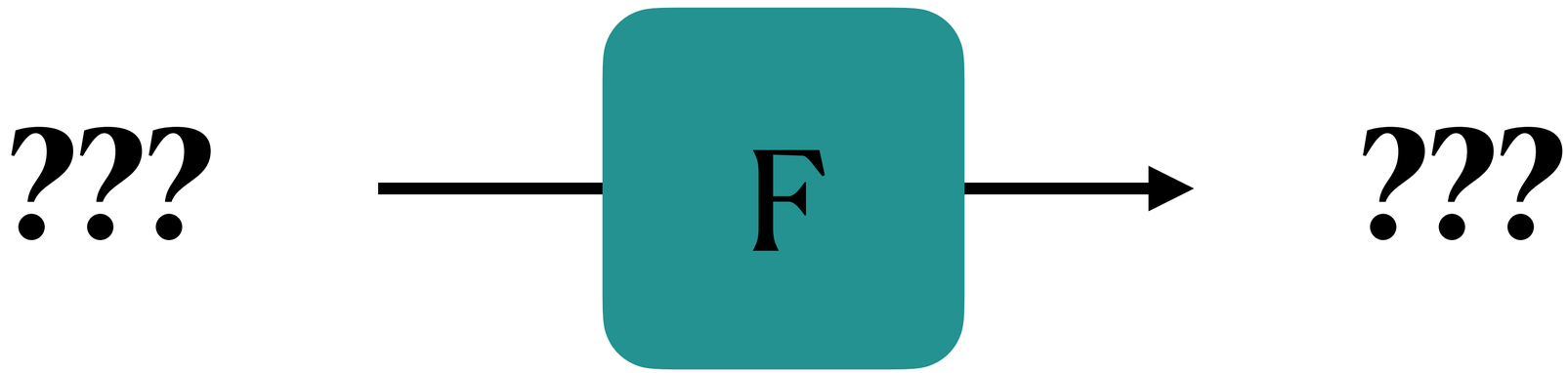
FRIDAY, FEBRUARY 02, 2024

Posted by Rajat Sen and Yichen Zhou, Google Research



Despite DL-based forecasters largely **outperforming** traditional methods and progress being made in **reducing training and inference costs**, they face challenges: most DL architectures require **long and involved training and validation cycles** before a customer can test the model on a new time-series. A foundation model for time-series forecasting, in contrast, can provide decent out-of-the-box forecasts on unseen time-series data with no additional training, enabling users to focus on refining forecasts for the actual downstream task like **retail demand planning**.

What is your input-output of interest?



+ Constraints

???

Laws are linear

Pascal's law (1653)

$$\Delta p = \rho g \Delta h$$

Hooke's law (1678)

$$F = -kx$$

Newton's law of viscosity (1701)

$$\tau = \mu \frac{du}{dy}$$

Ohm's law (1781)

$$I = V/R$$

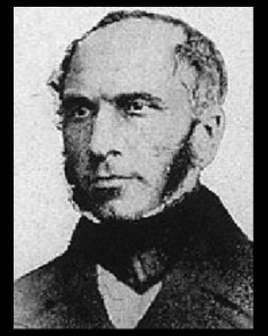
Fourier's law (1822)

$$q = -k \frac{dT}{dx}$$

Fick's law (1855)

$$J = -D \frac{dC}{dx}$$

Darcy's law (1856)

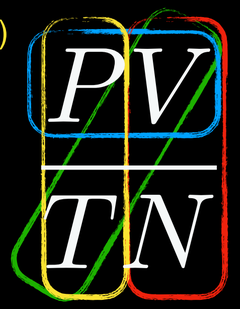


$$Q = \frac{kA}{\mu L} \Delta p$$

Ideal gas law (1834)

Amonton's law (1808)

Boyle's law (1662)



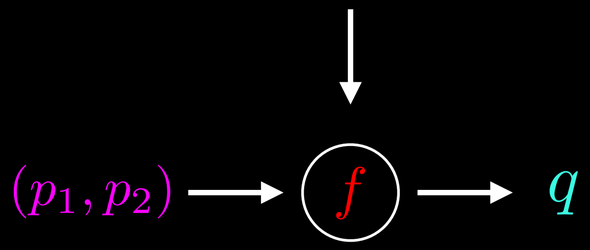
$$\frac{PV}{TN} = k_B$$

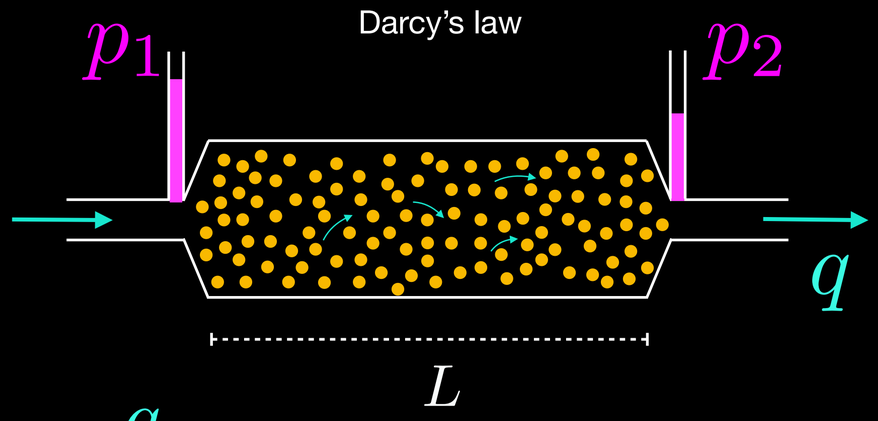
Charles's law (1787)

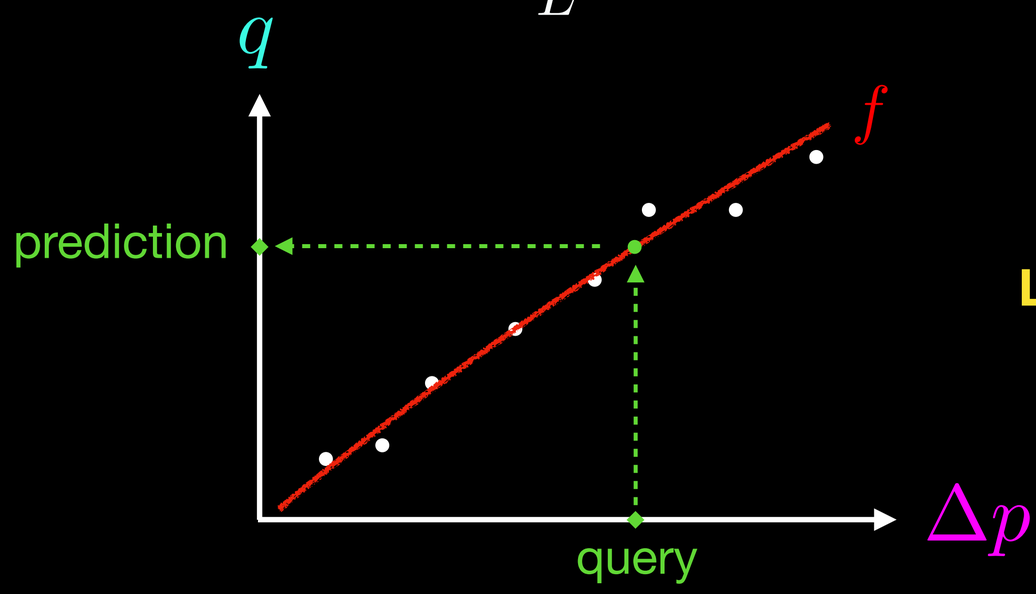
Avogadro's law (1811)

From experiment to Law

p_1	p_2	q
1.3	1.0	22
1.6	1.5	23
3.4	2.4	46
4.8	3.5	67
6.7	4.5	83
…		
2.3	**1.4**	**?**







Linear regression

Machine or human learning

Training data: $\mathcal{D}_{\text{train}}$

p_1	p_2	q
1.3	1.0	22
1.6	1.5	23
3.4	2.4	46
4.8	3.5	67
6.7	4.5	83
...		

Example →

Training

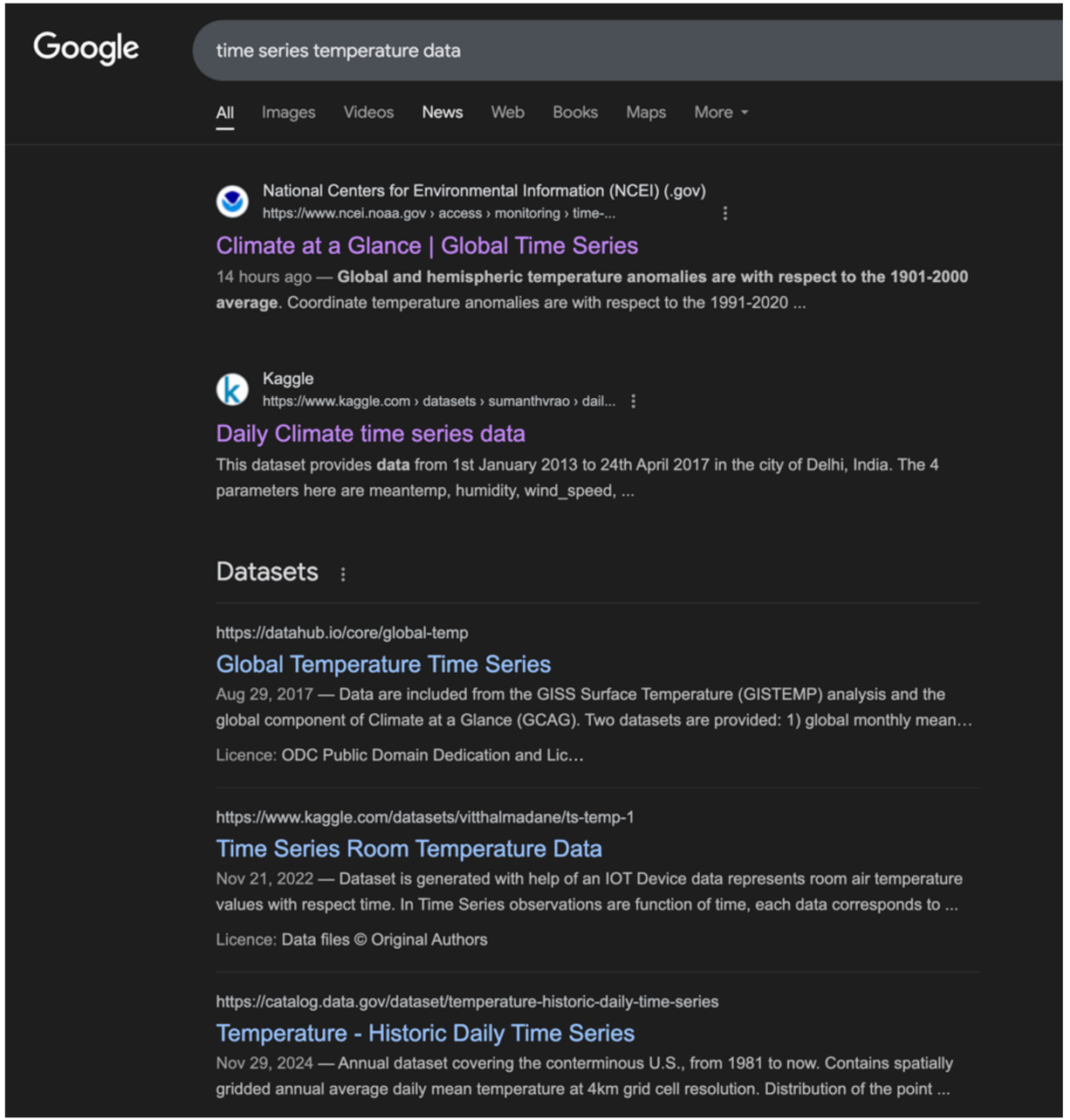
Feature extractor

$(p_1, p_2) \rightarrow \phi \rightarrow \Delta p \rightarrow f \rightarrow q$

Predictor *Labels*

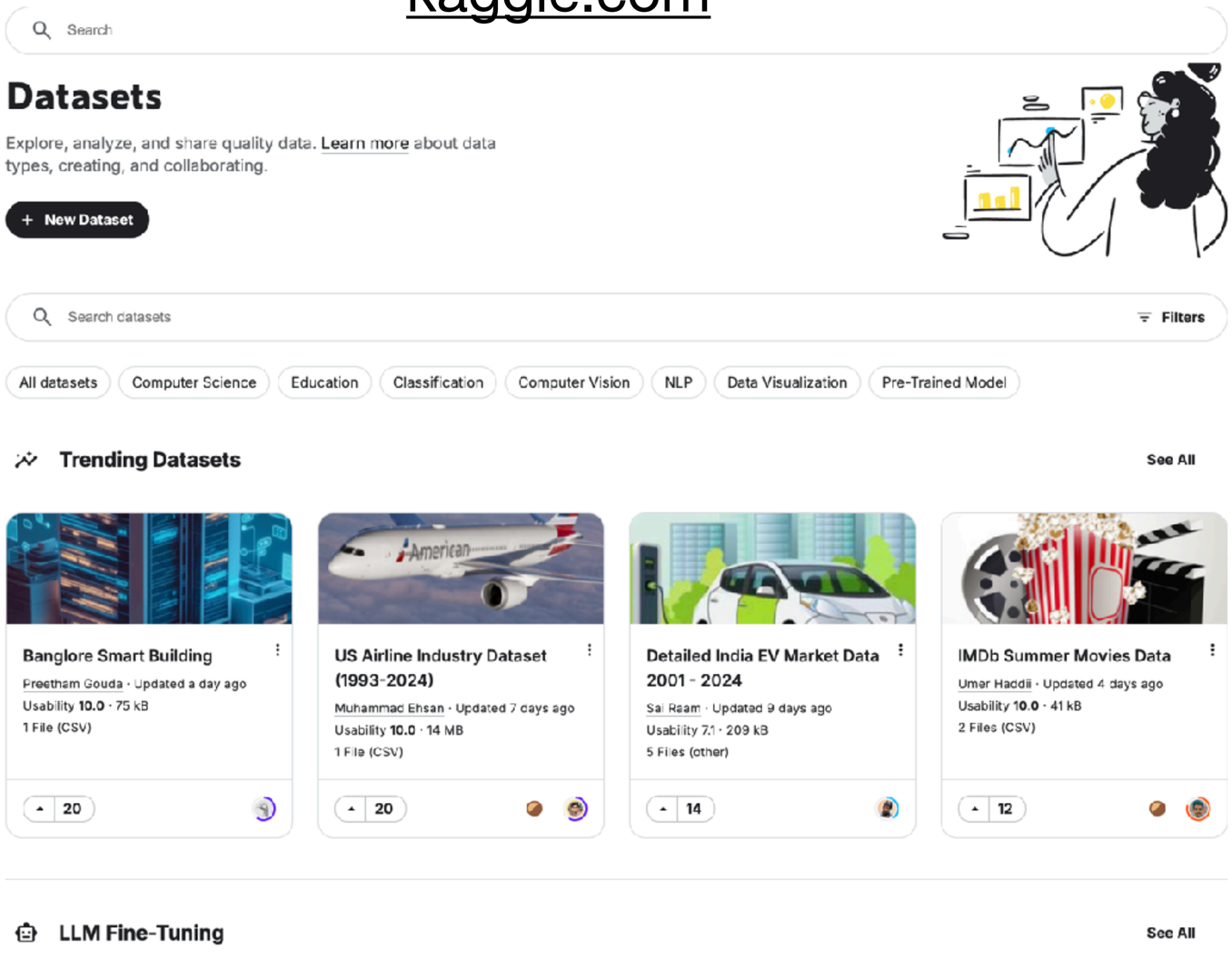
- Which predictors are possible?
- How good is the predictor?
- How can we find the best predictor?

Looking for data



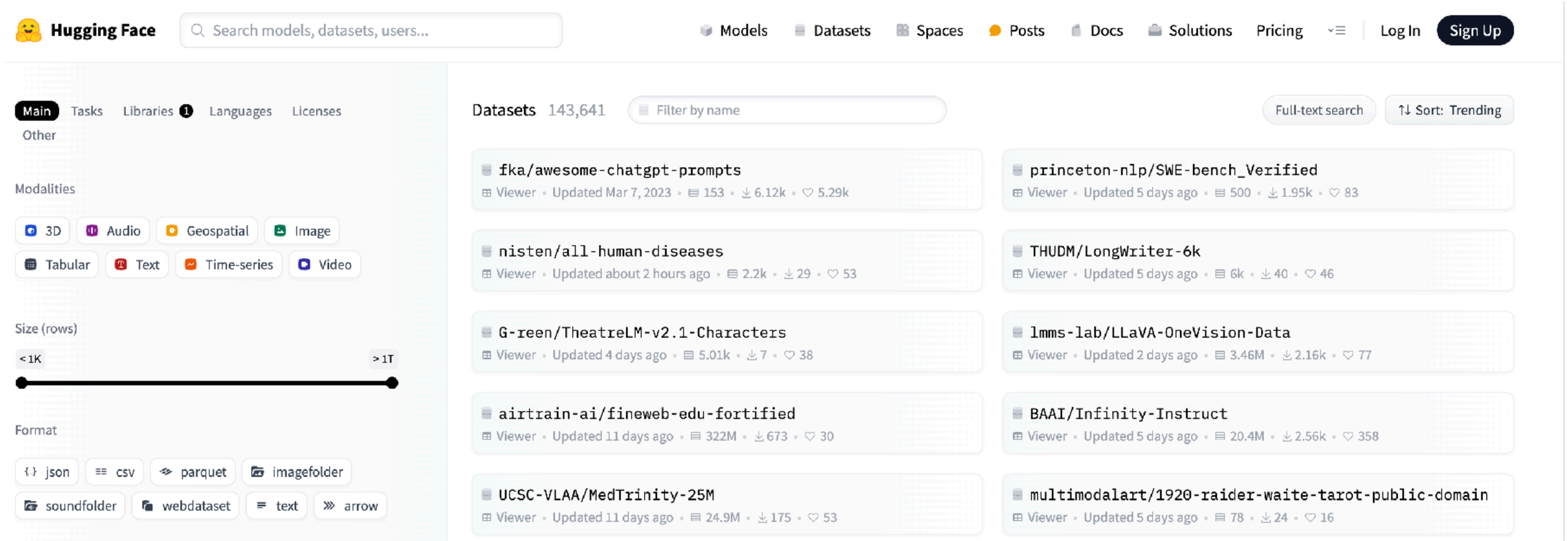
Where to look for data?

kaggle.com



Where to look for data?

huggingface.co



Homework

Collect data from your phone and do something with it