

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and don't necessarily reflect those of blibli.com.

Blibli.com does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.



Machine Learning: an Introduction and cases

Hendri Karisma

hendri.karisma@gdn-commerce.com



- Sr. Research and Development Engineer at blibli.com (PT. Global Digital Niaga)
- R&D Team in AI Squad
- Working for Fraud Detection System, Customer Group for abuser detection, dynamic recommendation system project and Customer Segmentation.
- <https://about.me/hendriKarisma>

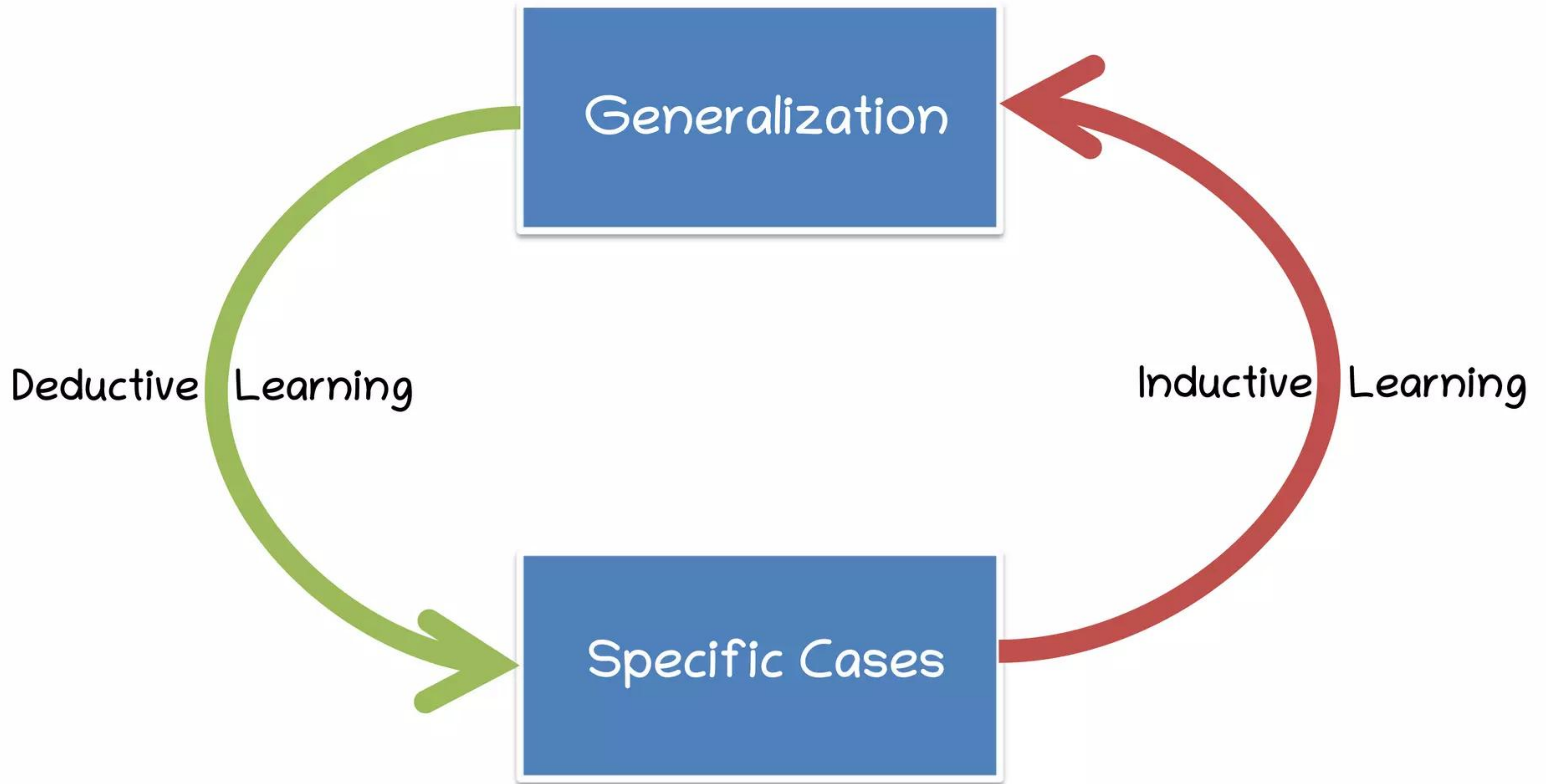
- Definition and background
- Methods
- Problems and Solutions
- Technologies
- Cases

“Automation of Information” –
Prof. Dr. Ing. Iping Supriana

- S. Rusel and P. Norvig, Artificial Intelligence in Modern Approach

<p>Thinking Humanly</p> <p>“The exciting new effort to make computers think . . . <i>machines with minds</i>, in the full and literal sense.” (Haugeland, 1985)</p> <p>“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” (Bellman, 1978)</p>	<p>Thinking Rationally</p> <p>“The study of mental faculties through the use of computational models.” (Charniak and McDermott, 1985)</p> <p>“The study of the computations that make it possible to perceive, reason, and act.” (Winston, 1992)</p>
<p>Acting Humanly</p> <p>“The art of creating machines that perform functions that require intelligence when performed by people.” (Kurzweil, 1990)</p> <p>“The study of how to make computers do things at which, at the moment, people are better.” (Rich and Knight, 1991)</p>	<p>Acting Rationally</p> <p>“Computational Intelligence is the study of the design of intelligent agents.” (Poole <i>et al.</i>, 1998)</p> <p>“AI . . . is concerned with intelligent behavior in artifacts.” (Nilsson, 1998)</p>

- Searching for solution
- Knowledge Base and Planning
- Reasoning
- Learning



Data ????

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .” - Prof. Tom Mitchel

- Analytical (Exact)

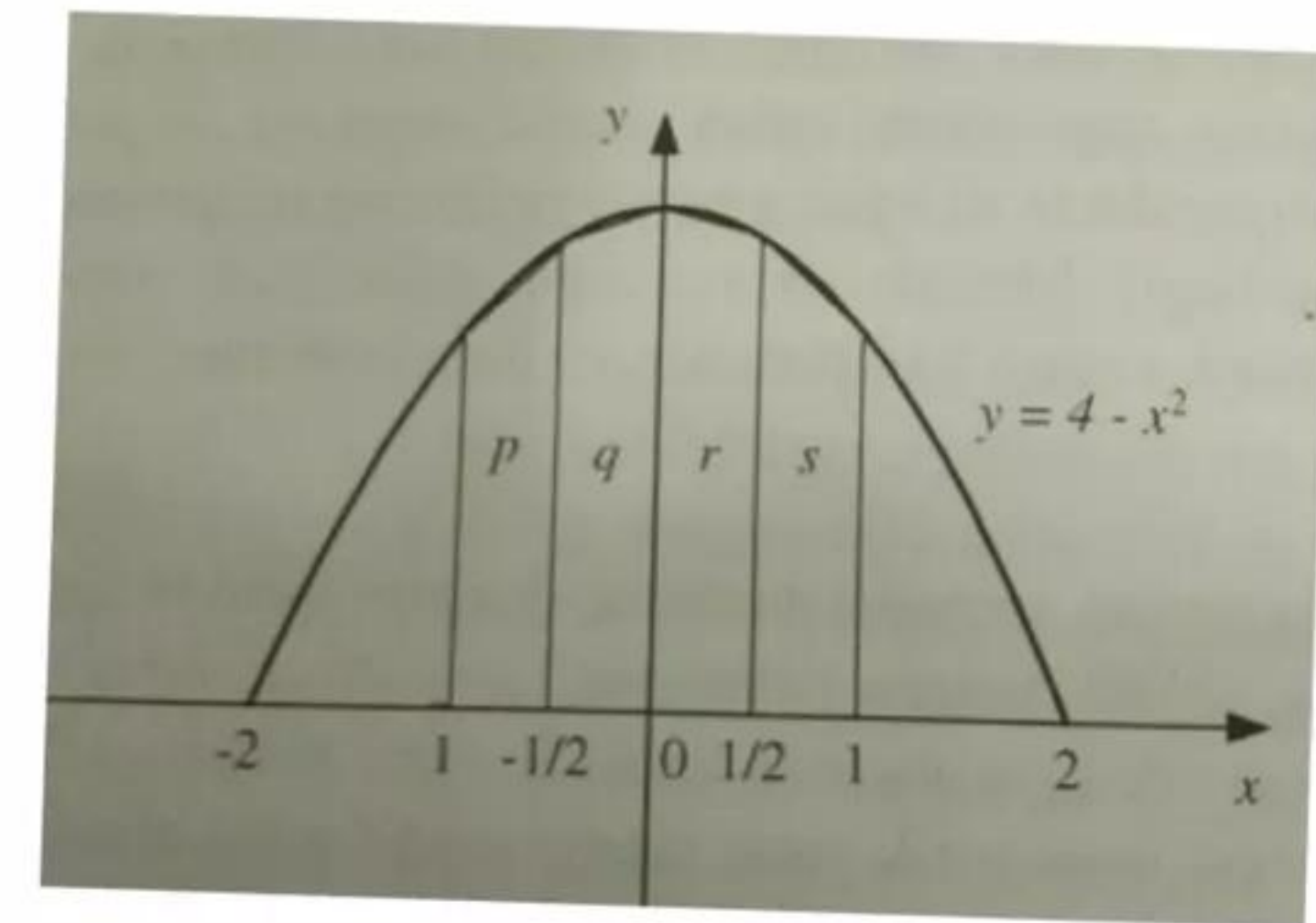
Example :

- analytics solution :

$$I = \int_{-1}^1 (4 - x^2) dx = [4x - x^3/3]_{x=-1}^{x=1} = \{4(1) - (1)/3\} - \{4(-1) - (-1)/3\} = 22/3$$

- Numerical solution

$$\begin{aligned} I &\approx p + q + r + s \\ &\approx \{[f(-1) + f(-1/2)] \times 0.5/2\} + \{[f(-1/2) + f(0)] \times 0.5/2\} + \\ &\quad \{[f(0) + f(1/2)] \times 0.5/2\} + \{[f(1/2) + f(1)] \times 0.5/2\} \\ &\approx 0.5/2 \{f(-1) + 2f(-1/2) + 2f(0) + 2f(1/2) + f(1)\} \\ &\approx 0.5/2 \{3 + 7.5 + 8 + 7.5 + 3\} \\ &\approx 7.25 \end{aligned}$$

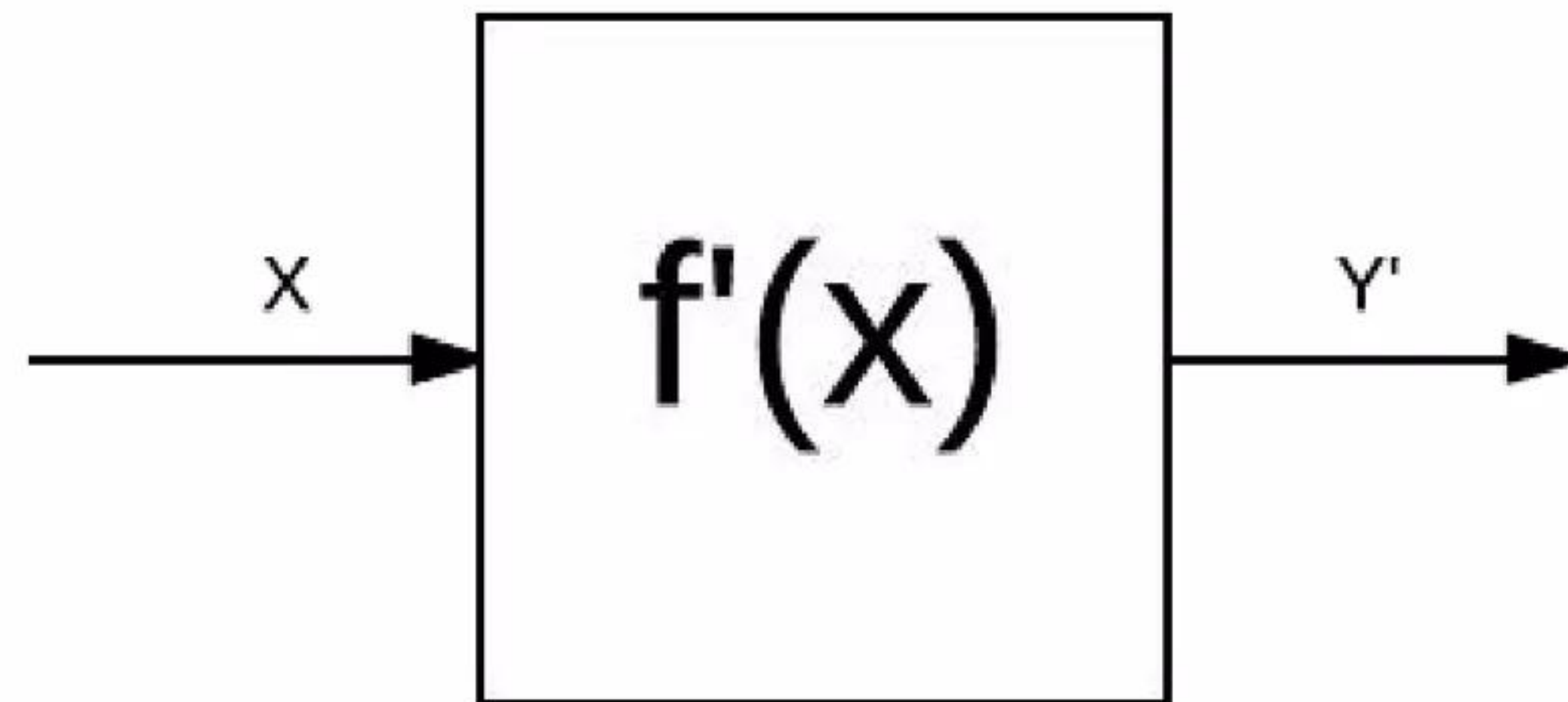
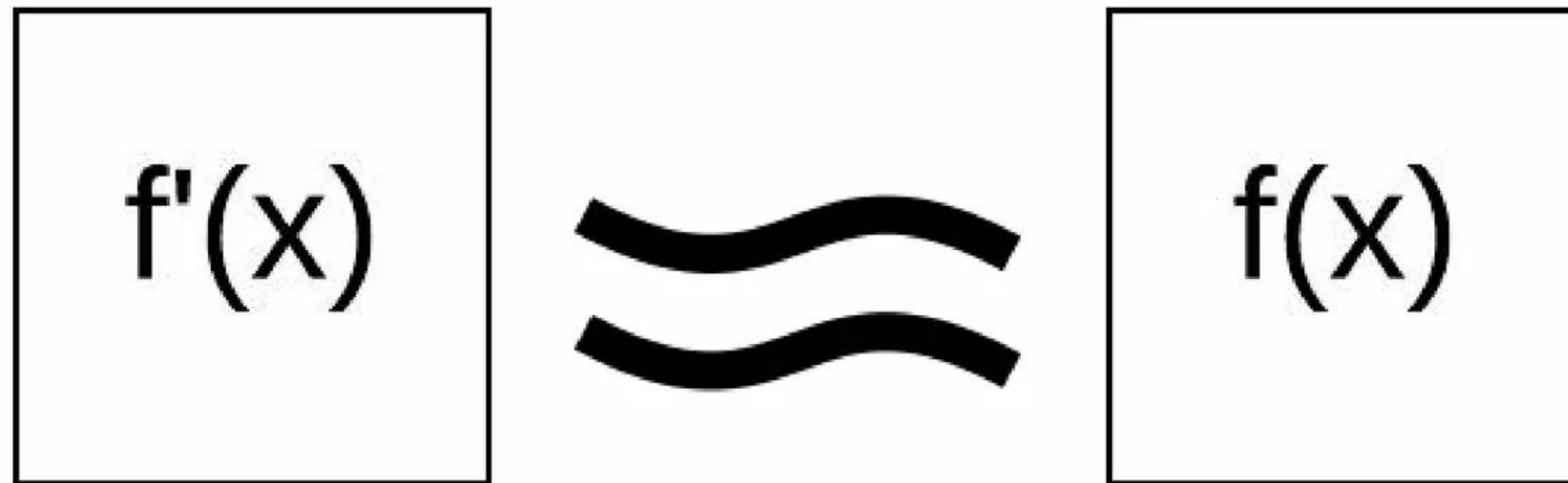


- Error = $|7.25 - 22/3| = |7.25 - 7.33| = 0.08333$

- Numerical (Aprox)

- Is numerical methods just about ML method that we know in the book?

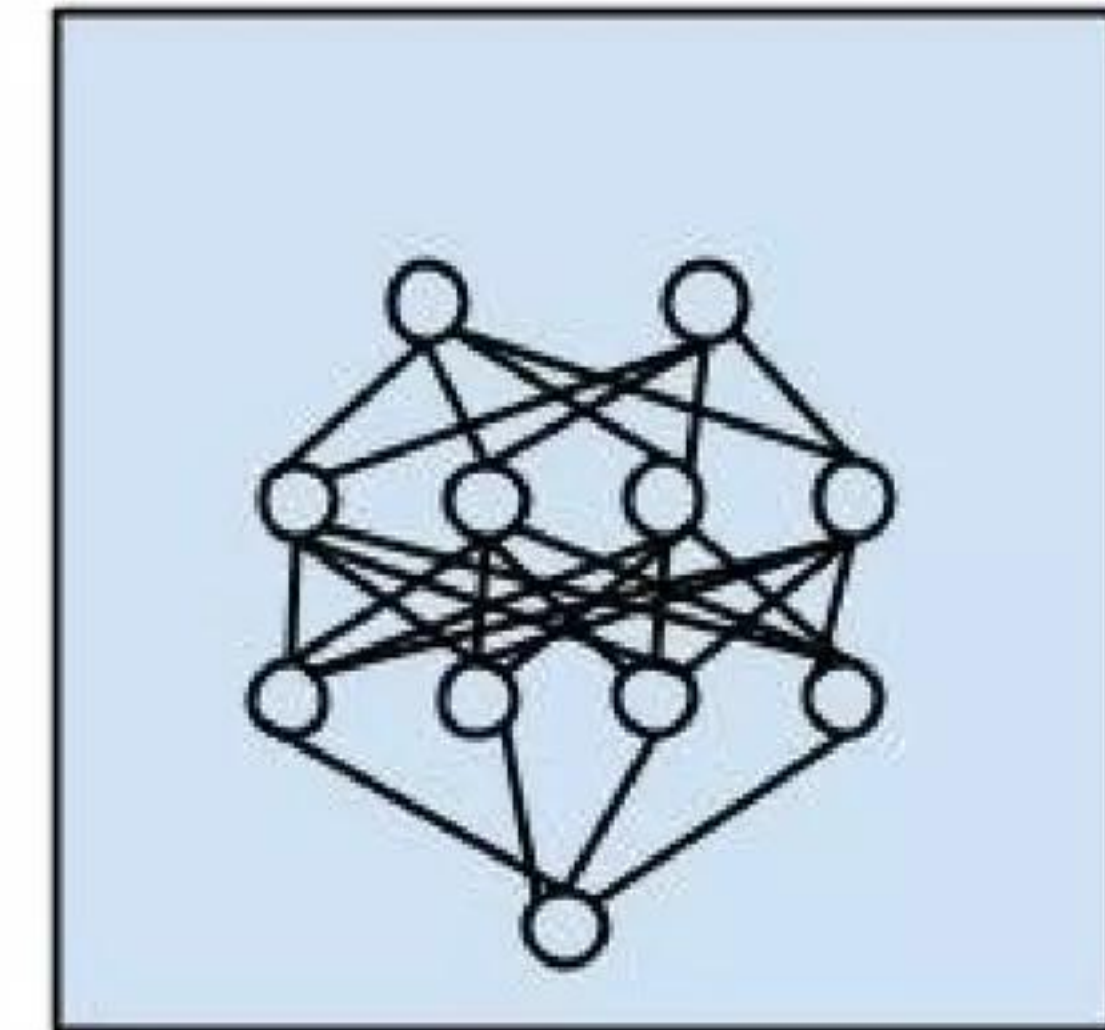
- Newton raphson, Gauss Elimination, Gauss-Jordan, Jacobi method, Gauss-Seidel, Lagrange, Newton Gregory, Richardson Interpolation, etc.



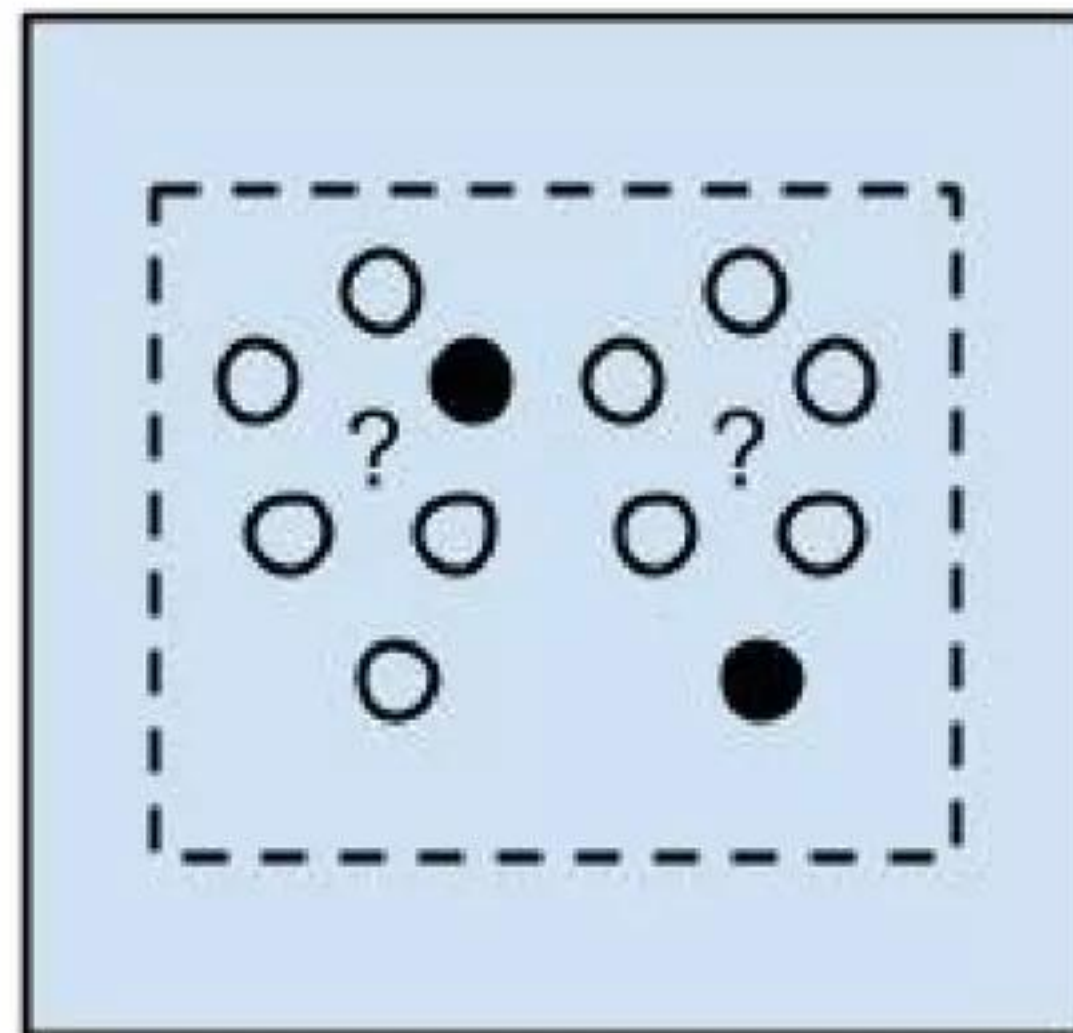
Count the error ($y - y'$)
Then minimize the error
or
Maximize the likelihood

- Information Theory (Decision Tree : ID-Tree, C4.5, etc)
- Probability (Bayesian : Naive Bayes, Belief Network, etc)
- Graphical Model (Belief network, HMM, CRF, Neural Network, etc)
- Numerical Method / Regression (Stochastic Gradient Descent: Linear Regression, Multiple Linear Regression, Neural Network, Stochastic Gradient Ascent : E-M Algorithm)

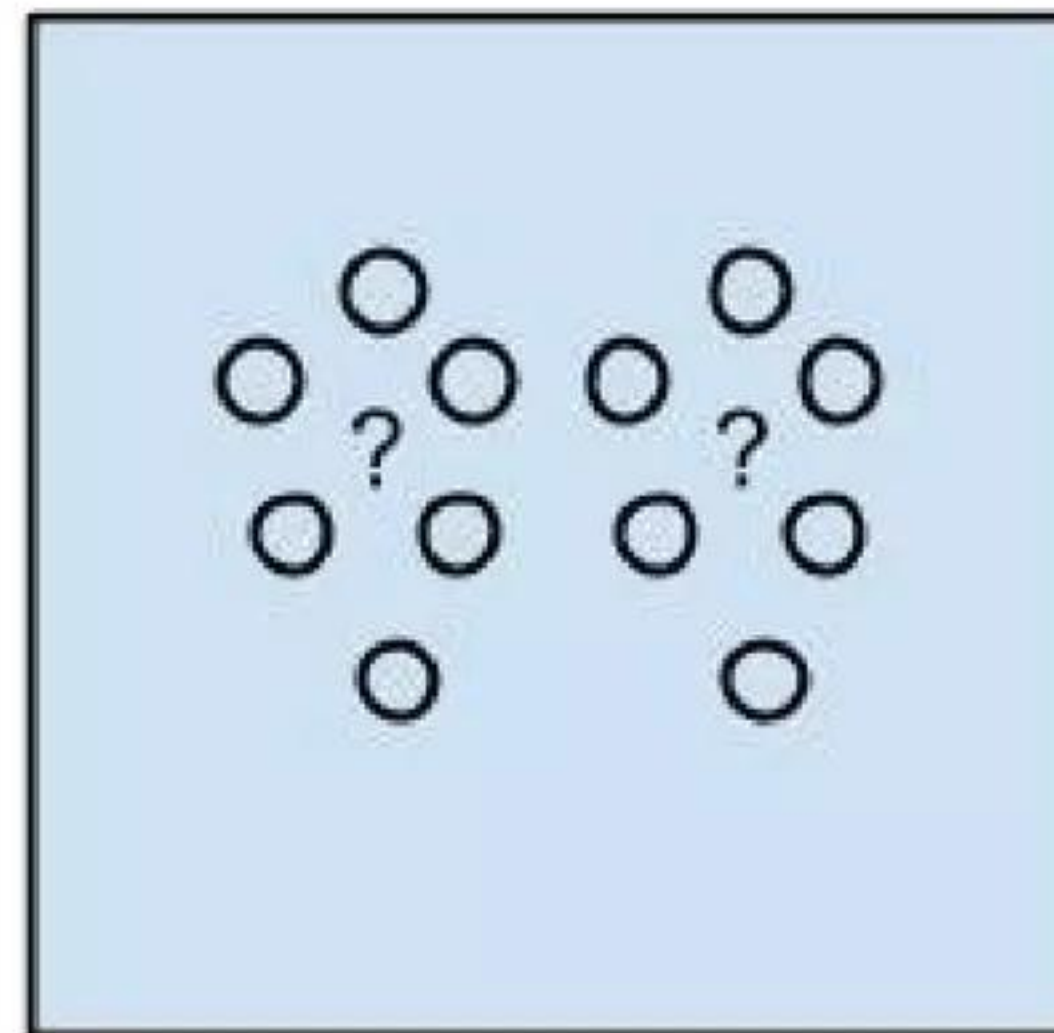
- Supervised
- Unsupervised
- Reinforcement Learning
- Semi-Supervised
- Deep Learning



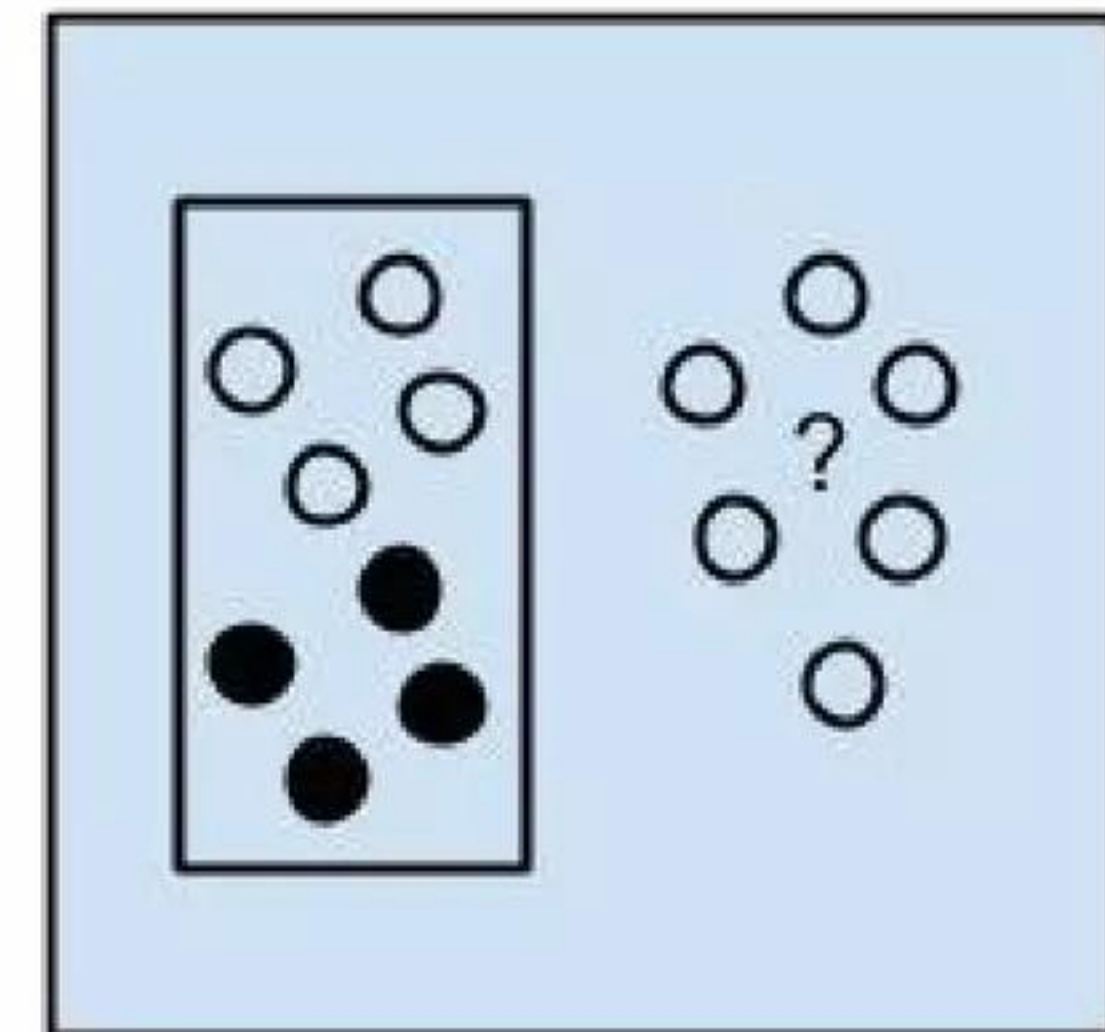
Deep Learning Algorithms



Semi-supervised Learning Algorithms

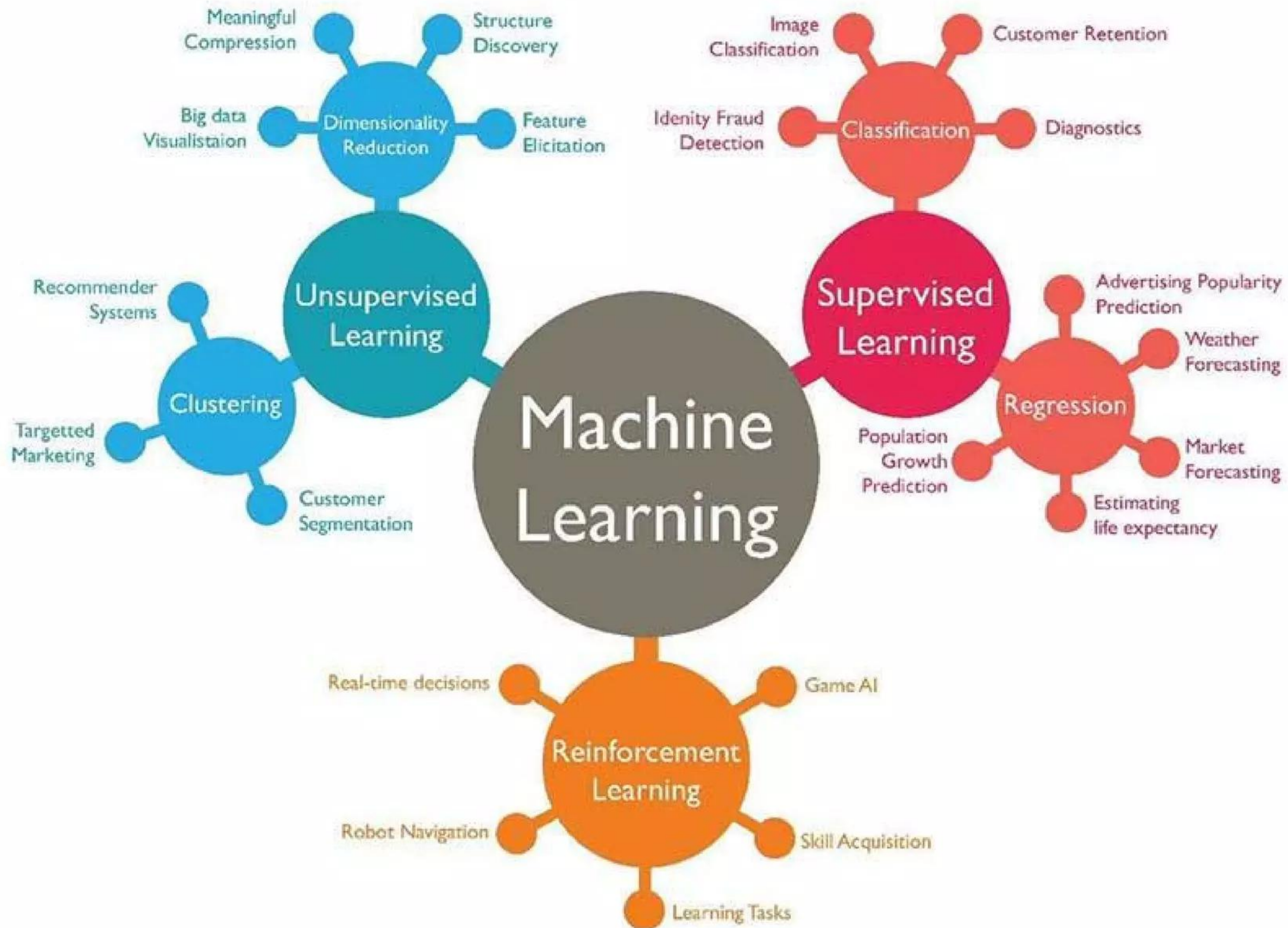


Unsupervised Learning Algorithms



Supervised Learning Algorithms

Machine Learning Taxonomy #2



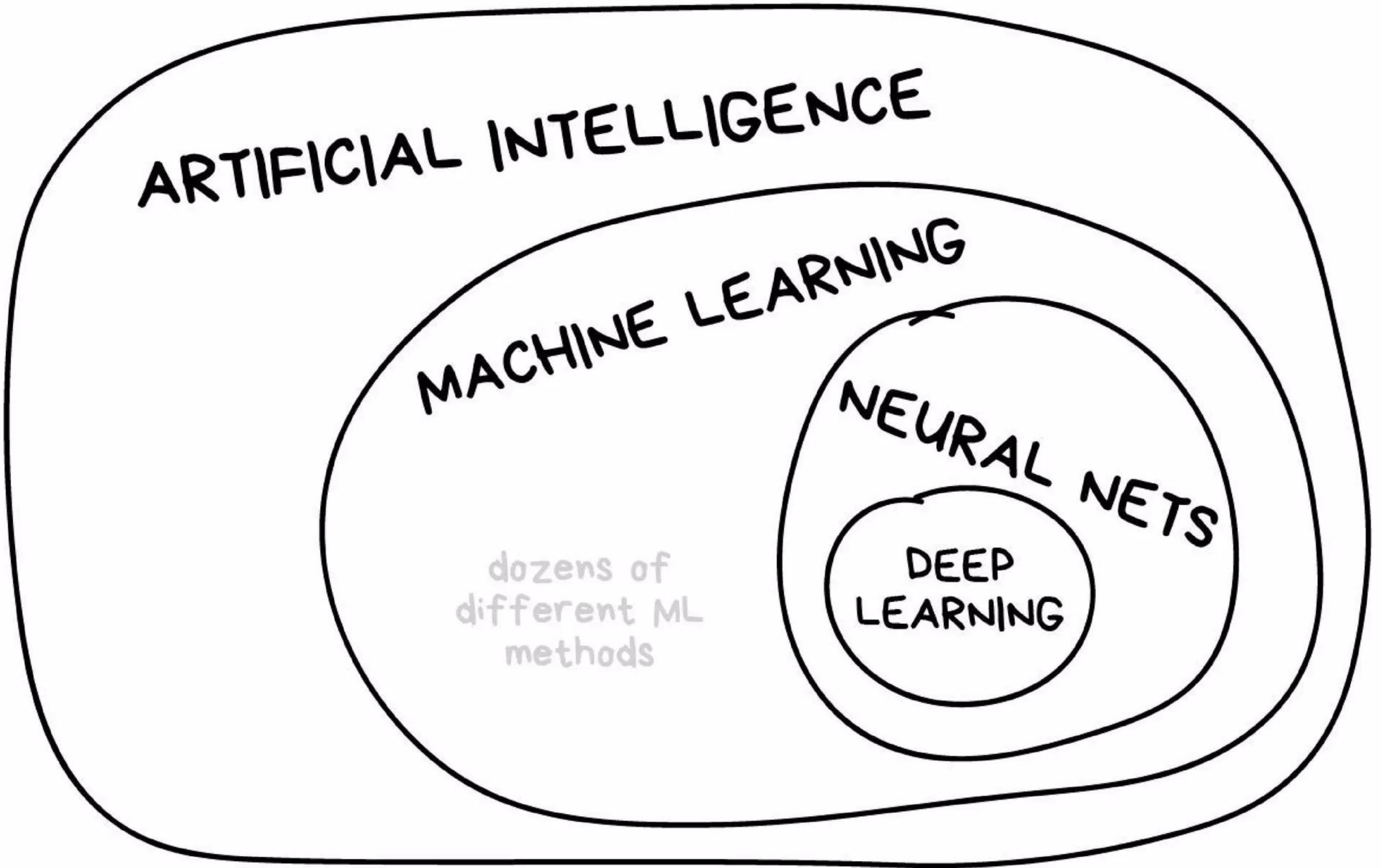
ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

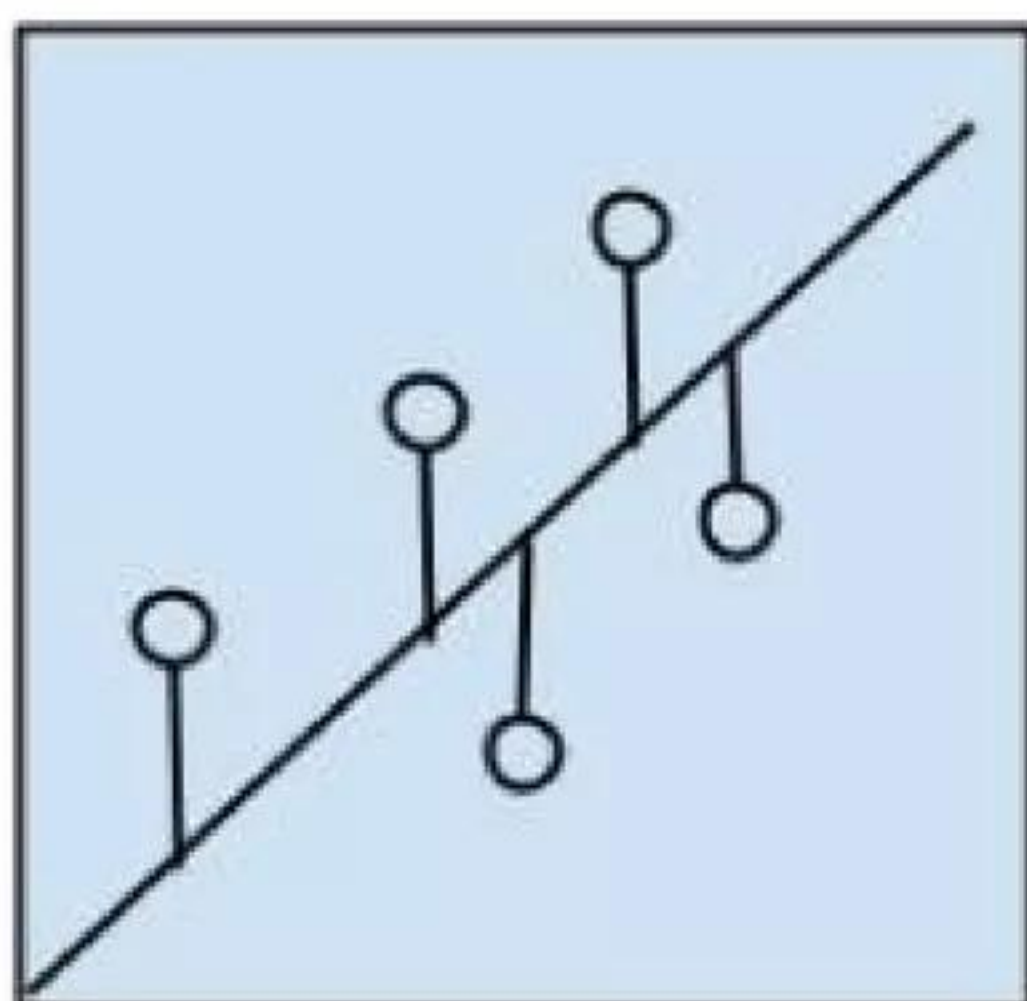
NEURAL NETS

DEEP
LEARNING

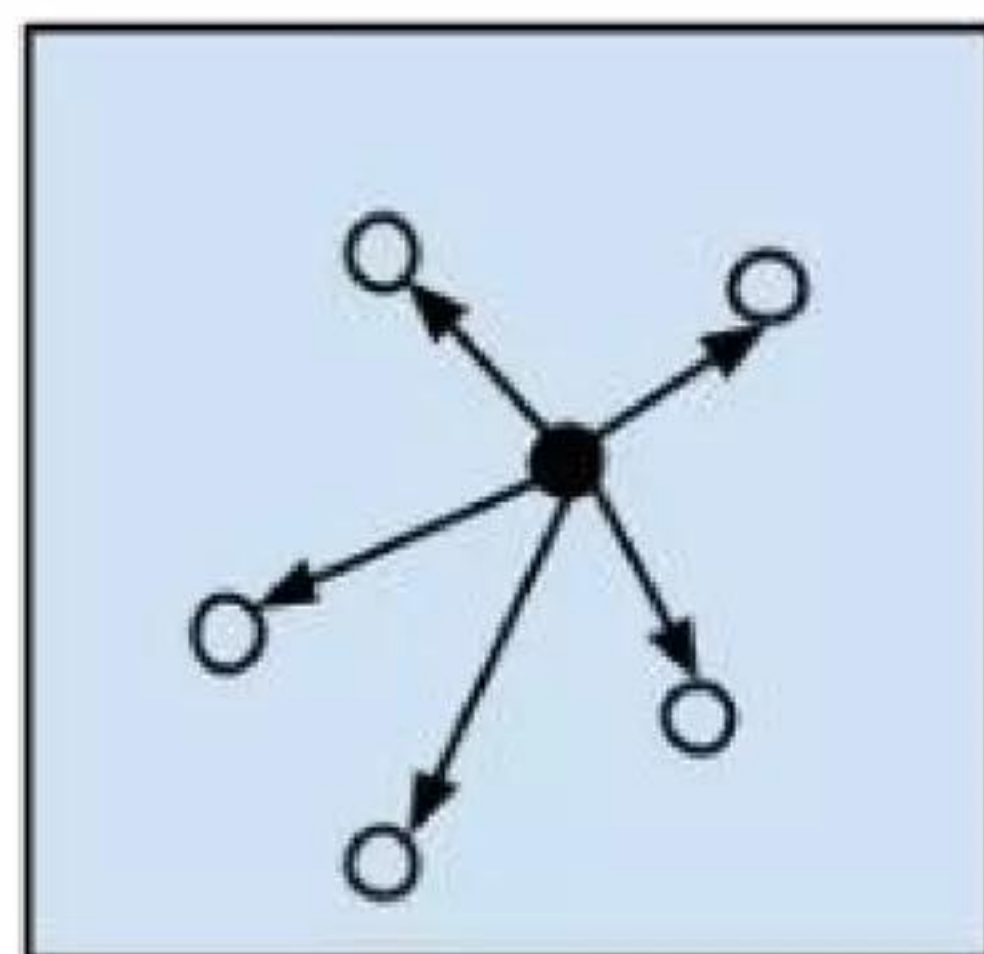
dozens of
different ML
methods



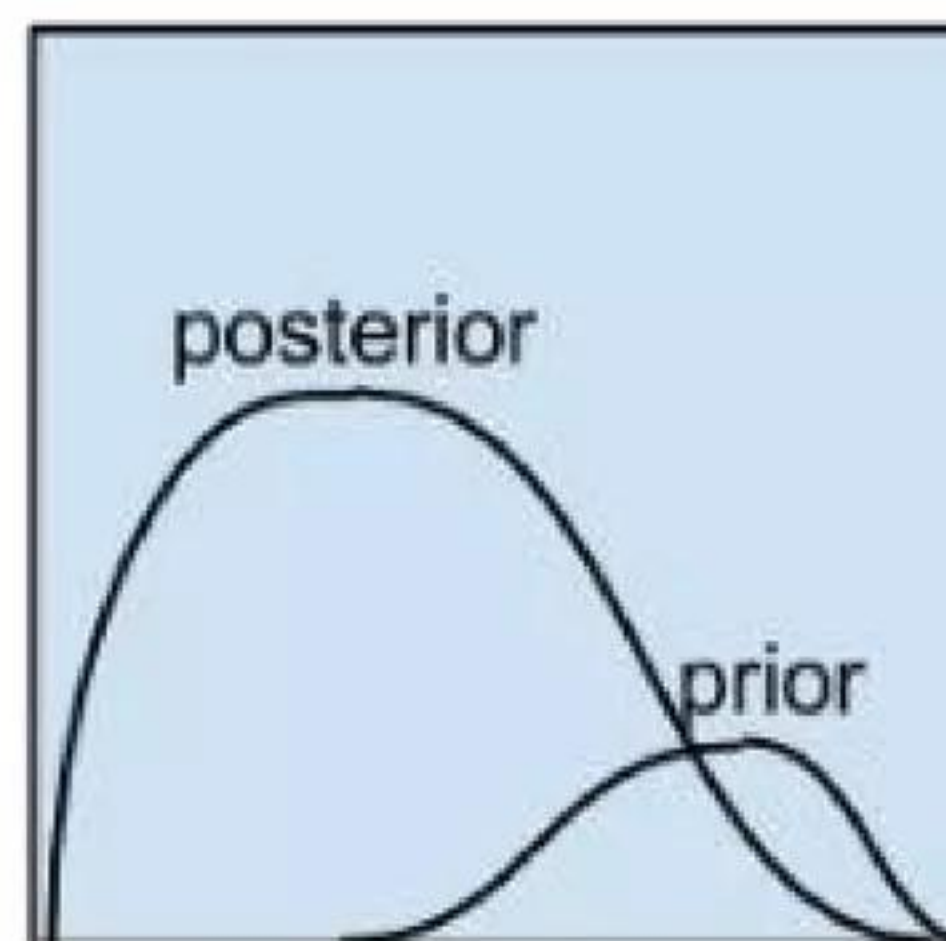
Machine Learning Taxonomy



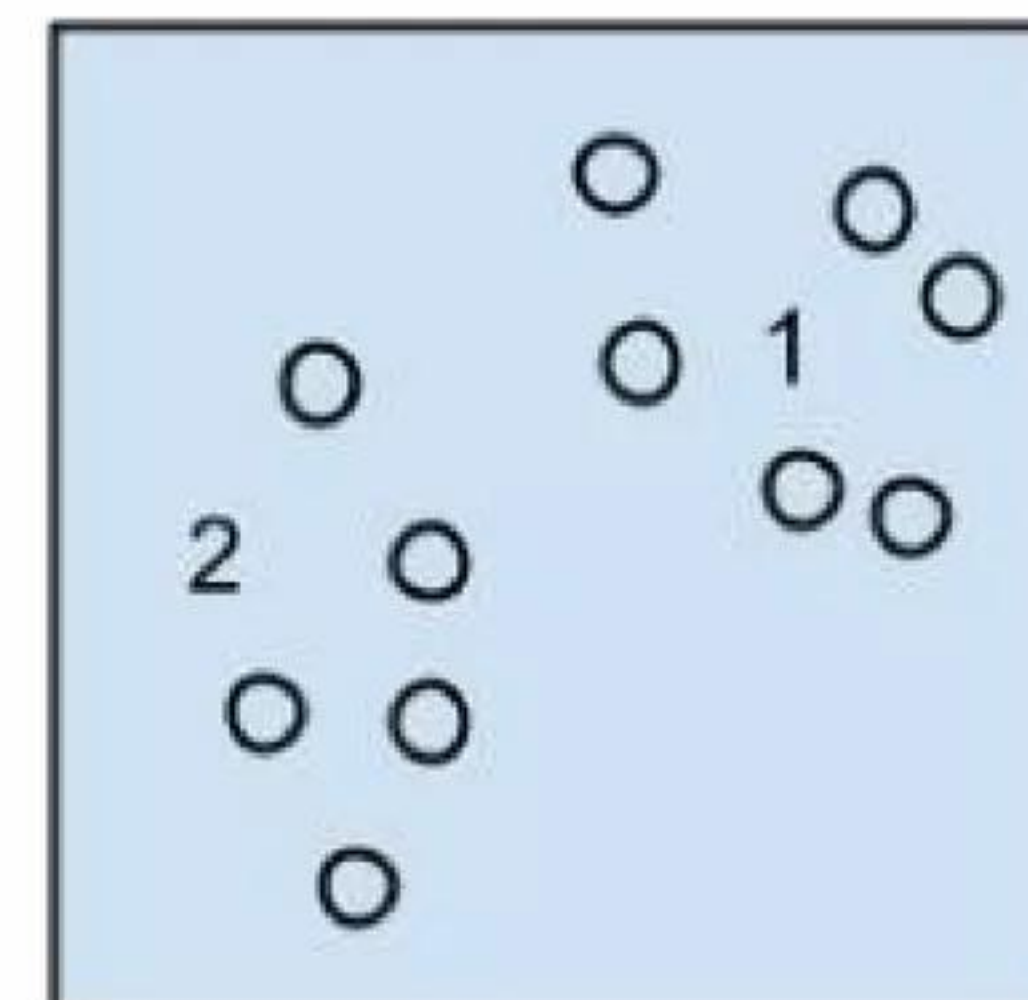
Regression Algorithms



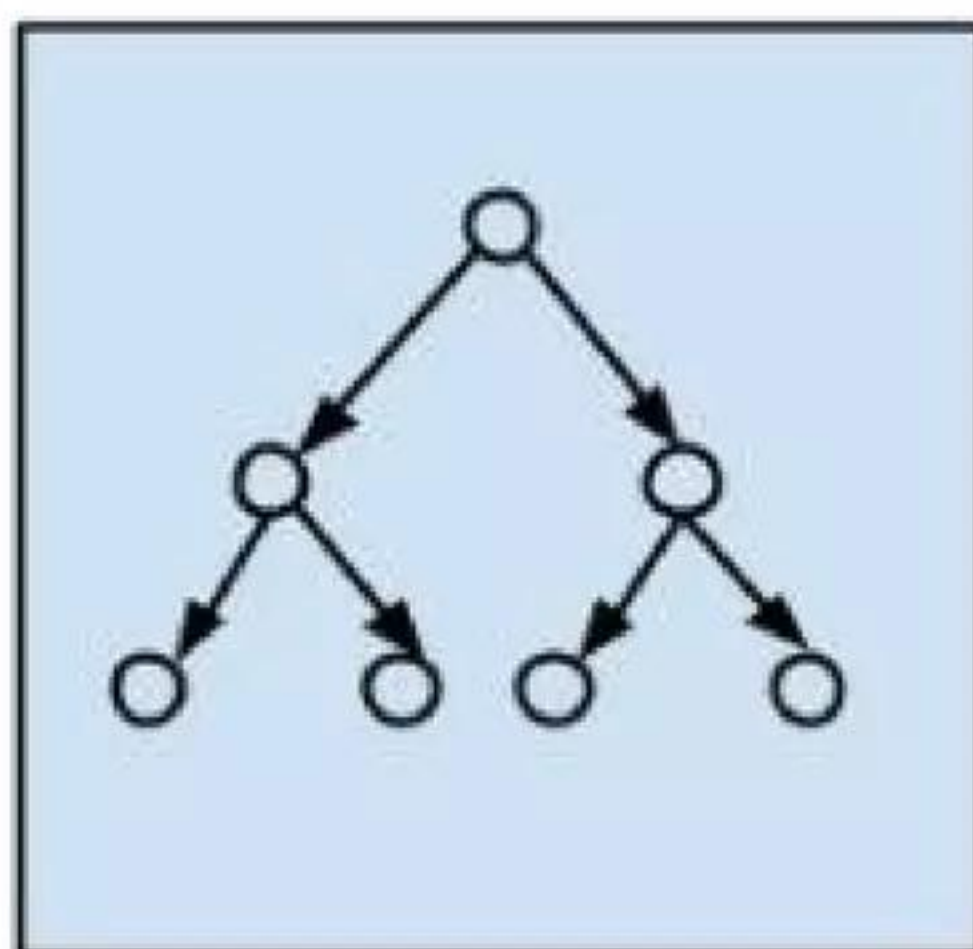
Instance-based Algorithms



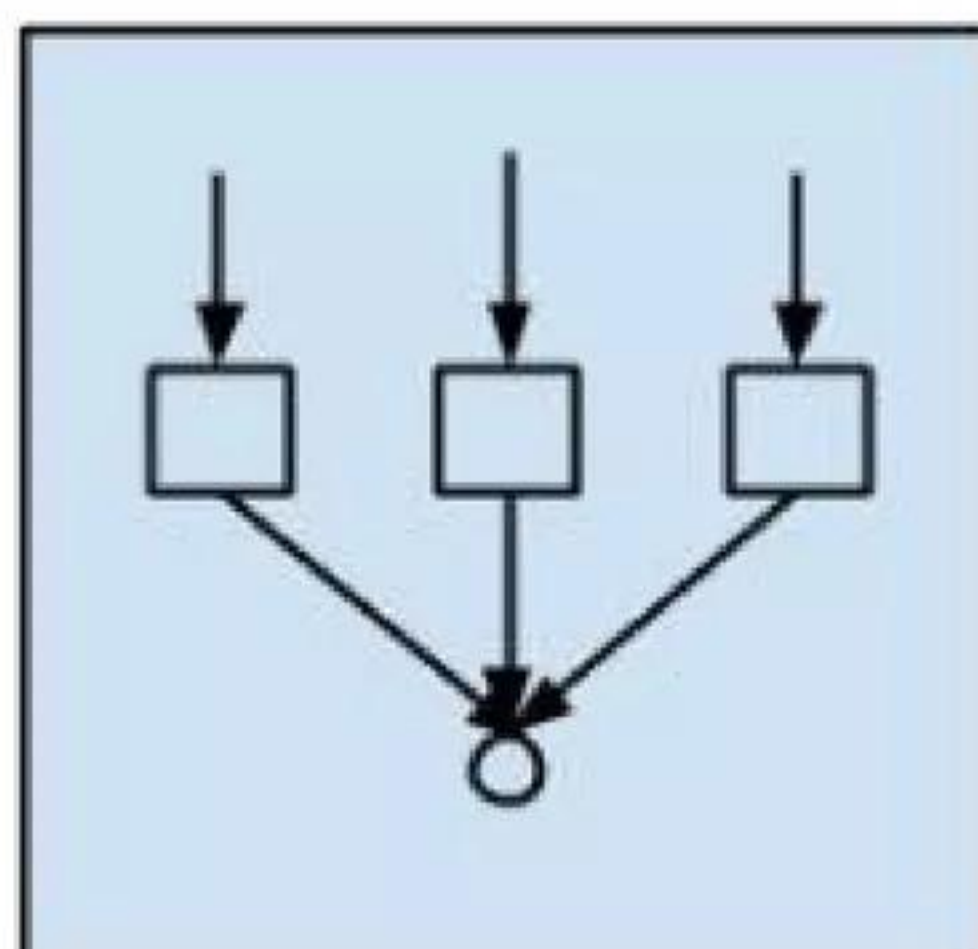
Bayesian Algorithms



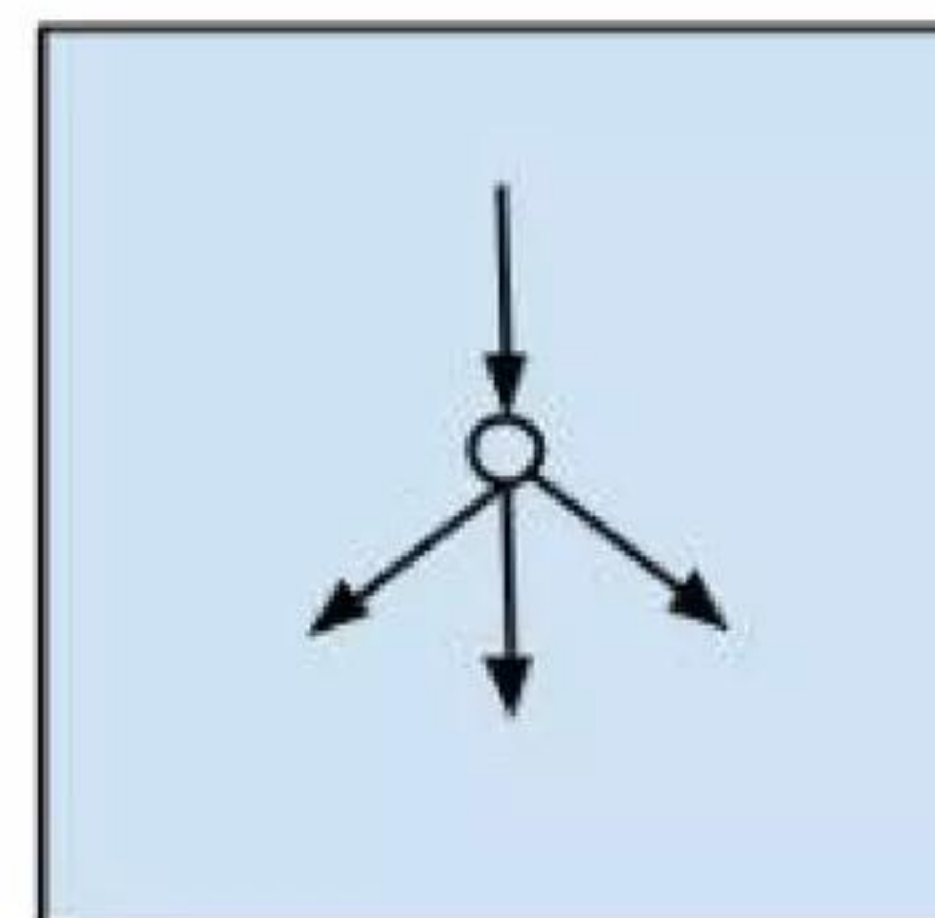
Clustering Algorithms



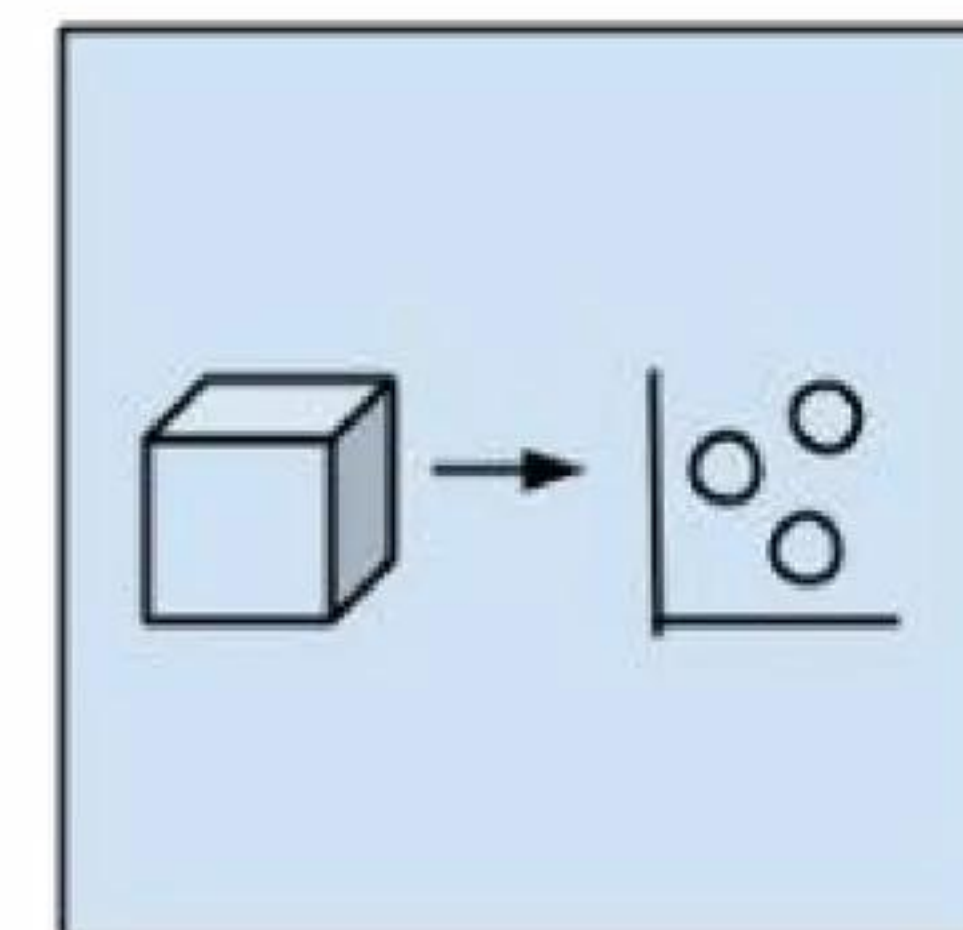
Decision Tree Algorithms



Ensemble Algorithms



Artificial Neural Network Algorithms



Dimensional Reduction Algorithms

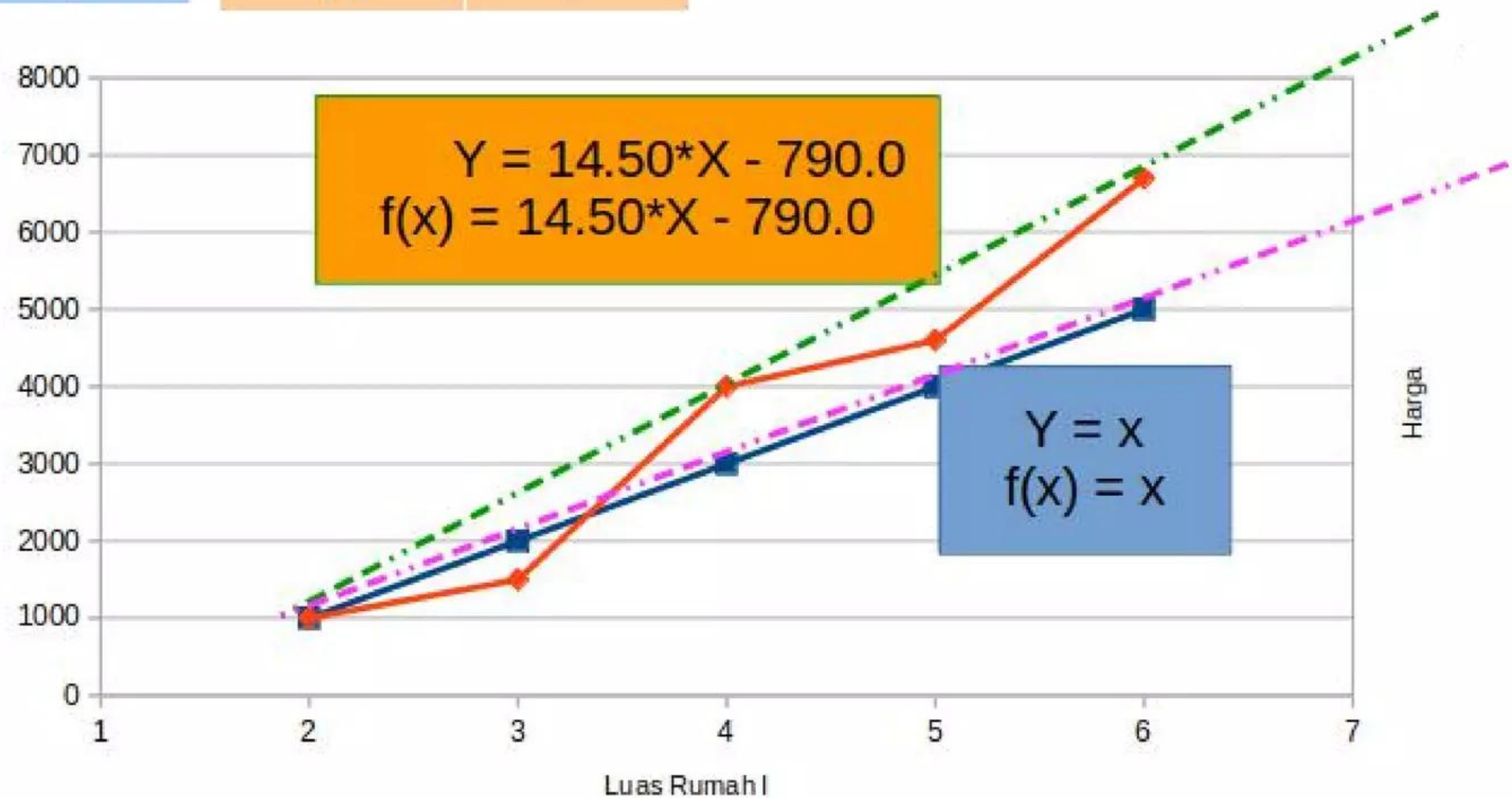
Regression

Contoh

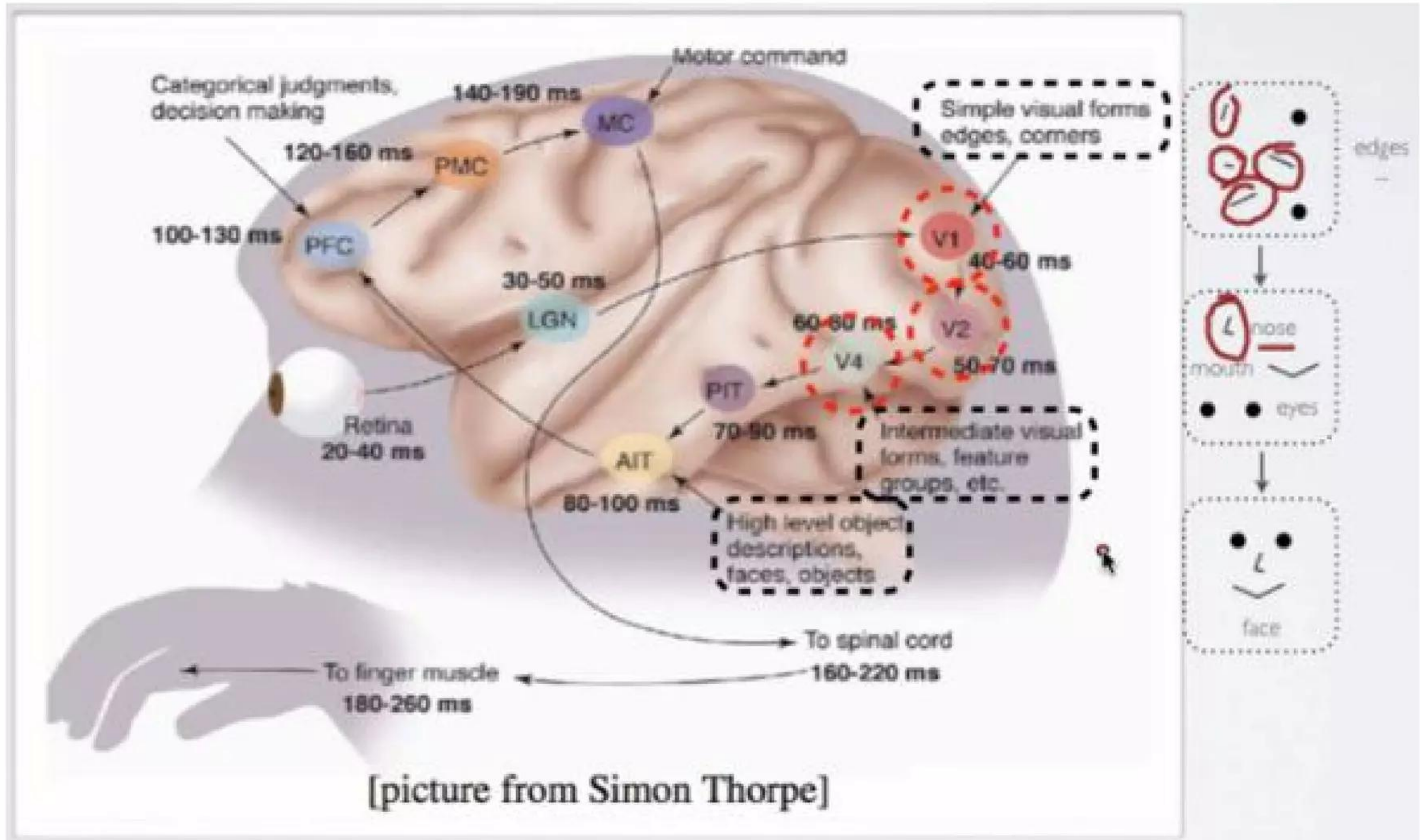
Luas Rumah (m ²)	Harga (Juta)	Luas Rumah (m ²)	Harga (Juta)
100	1000	100	1000
200	2000	200	1500
300	3000	300	4000
400	4000	400	4600
500	5000	500	6700
600	?	600	?

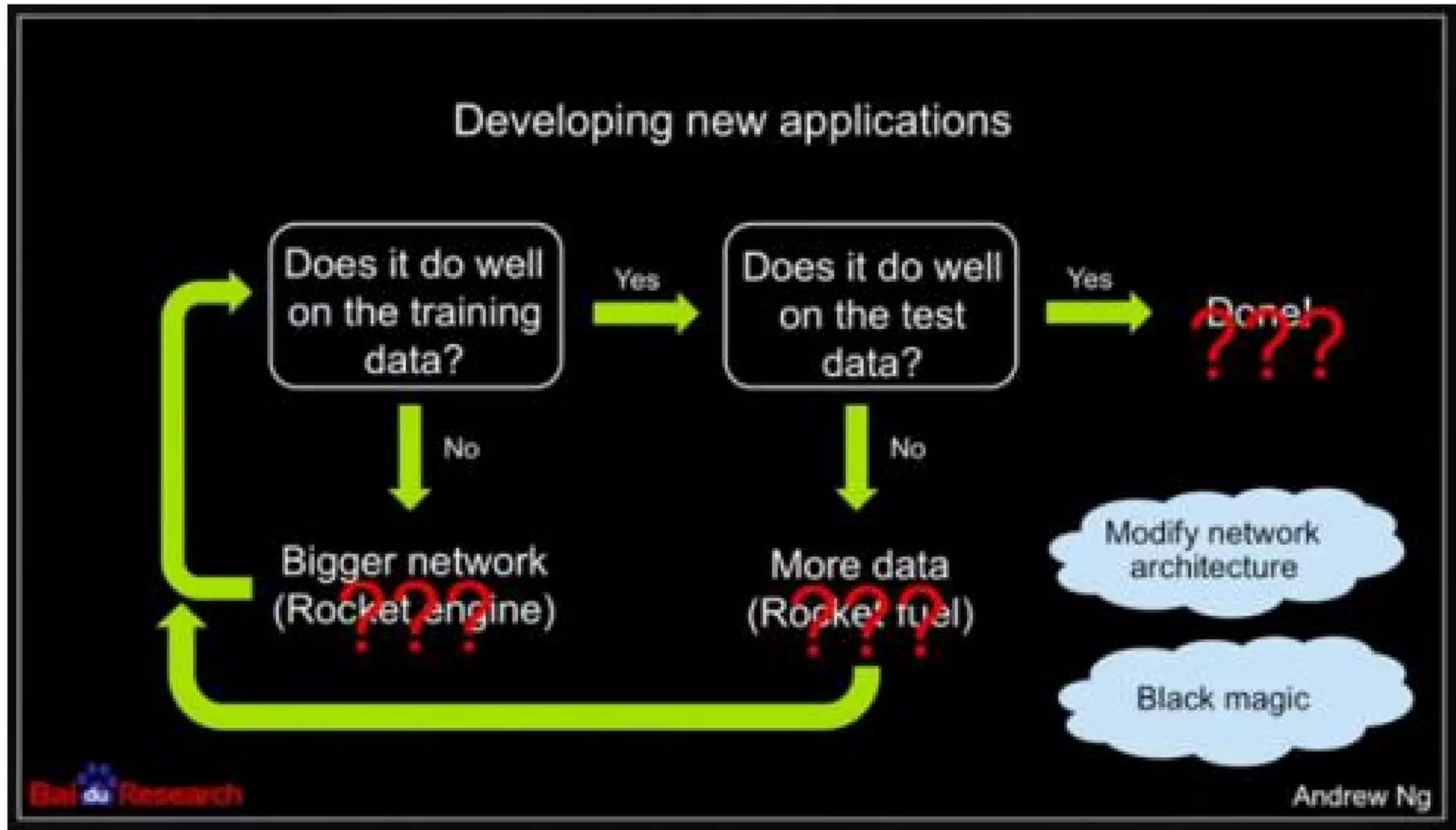
X	Y
1	1
2	1,5
3	4
4	4,5
5	6,7

$$Y = 1.500 * X - 0.5000$$

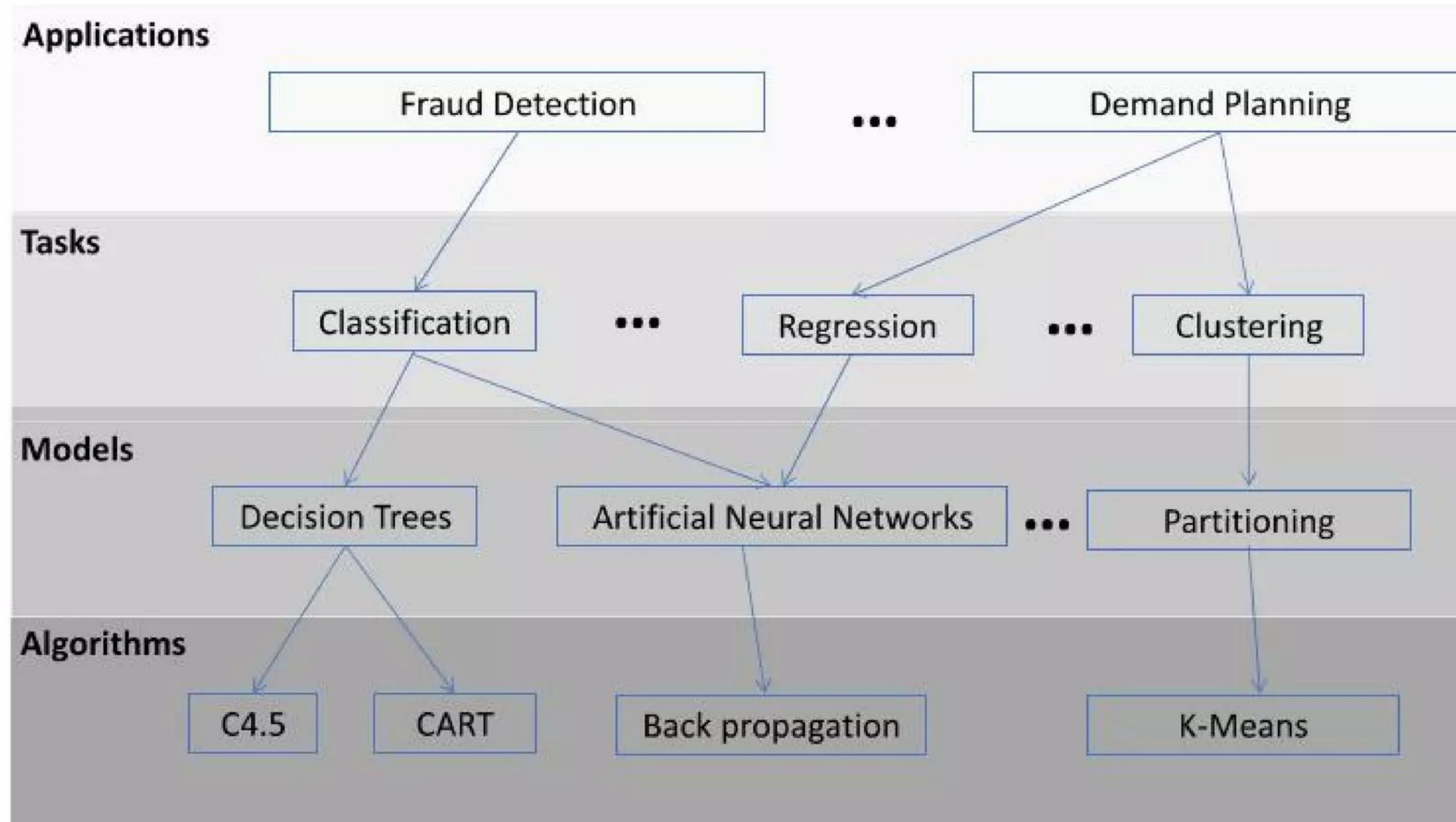


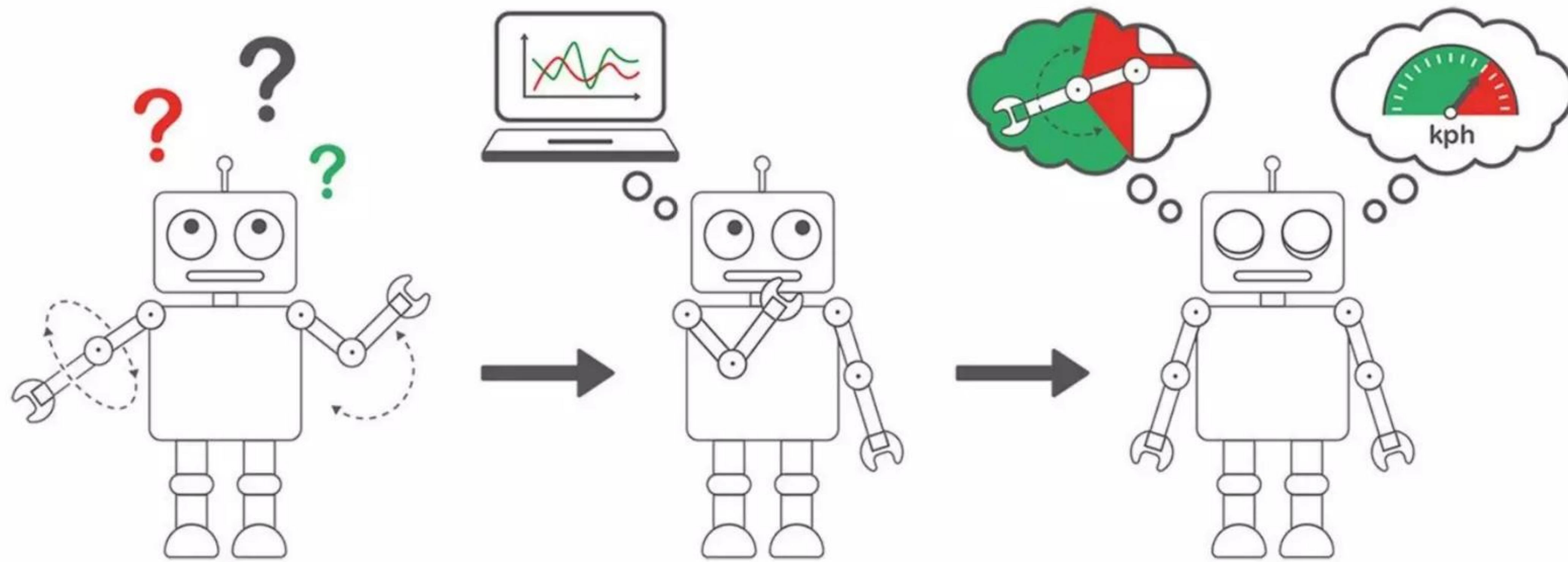
Deep Learning





The four layer of Datamining

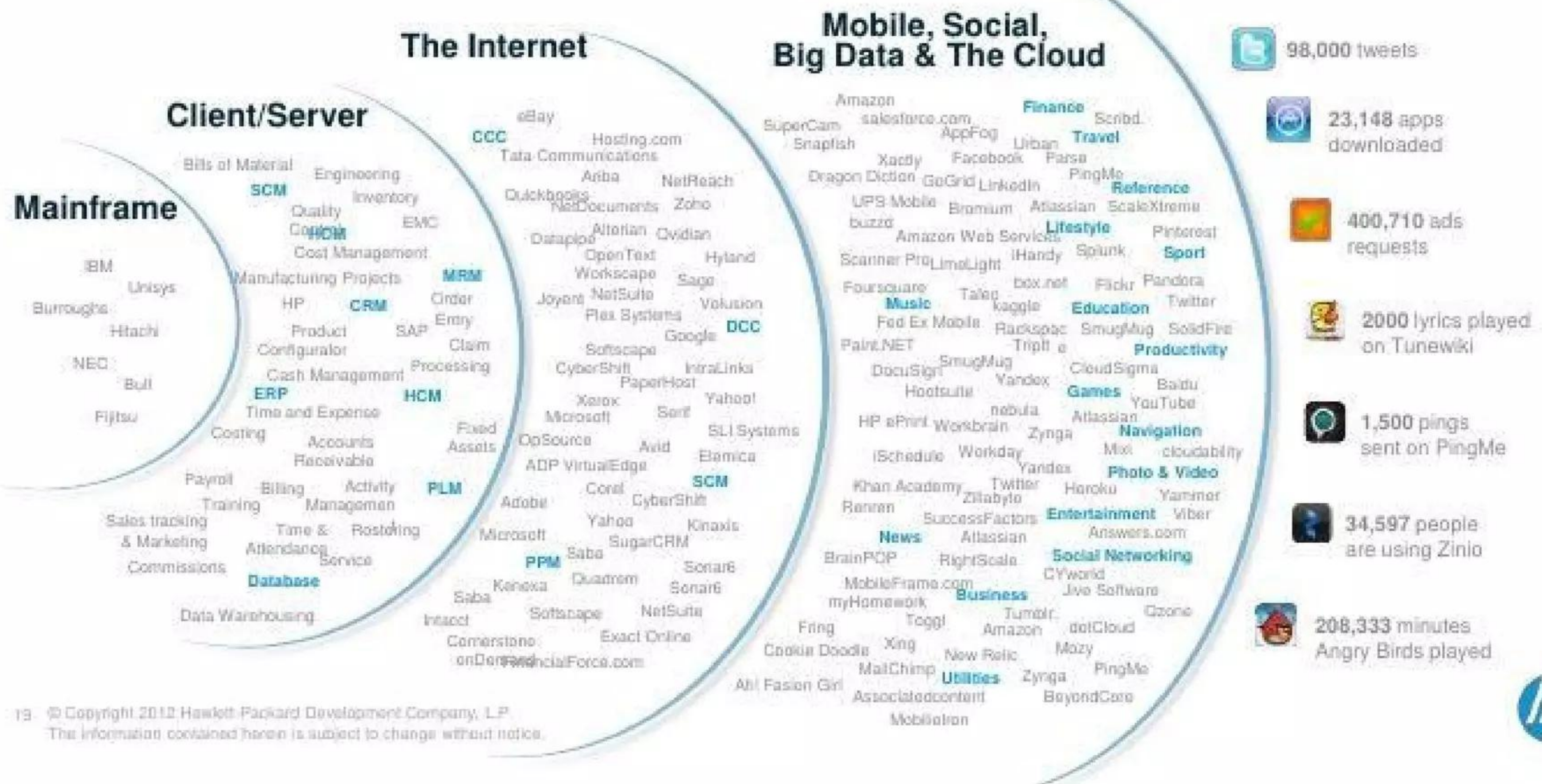




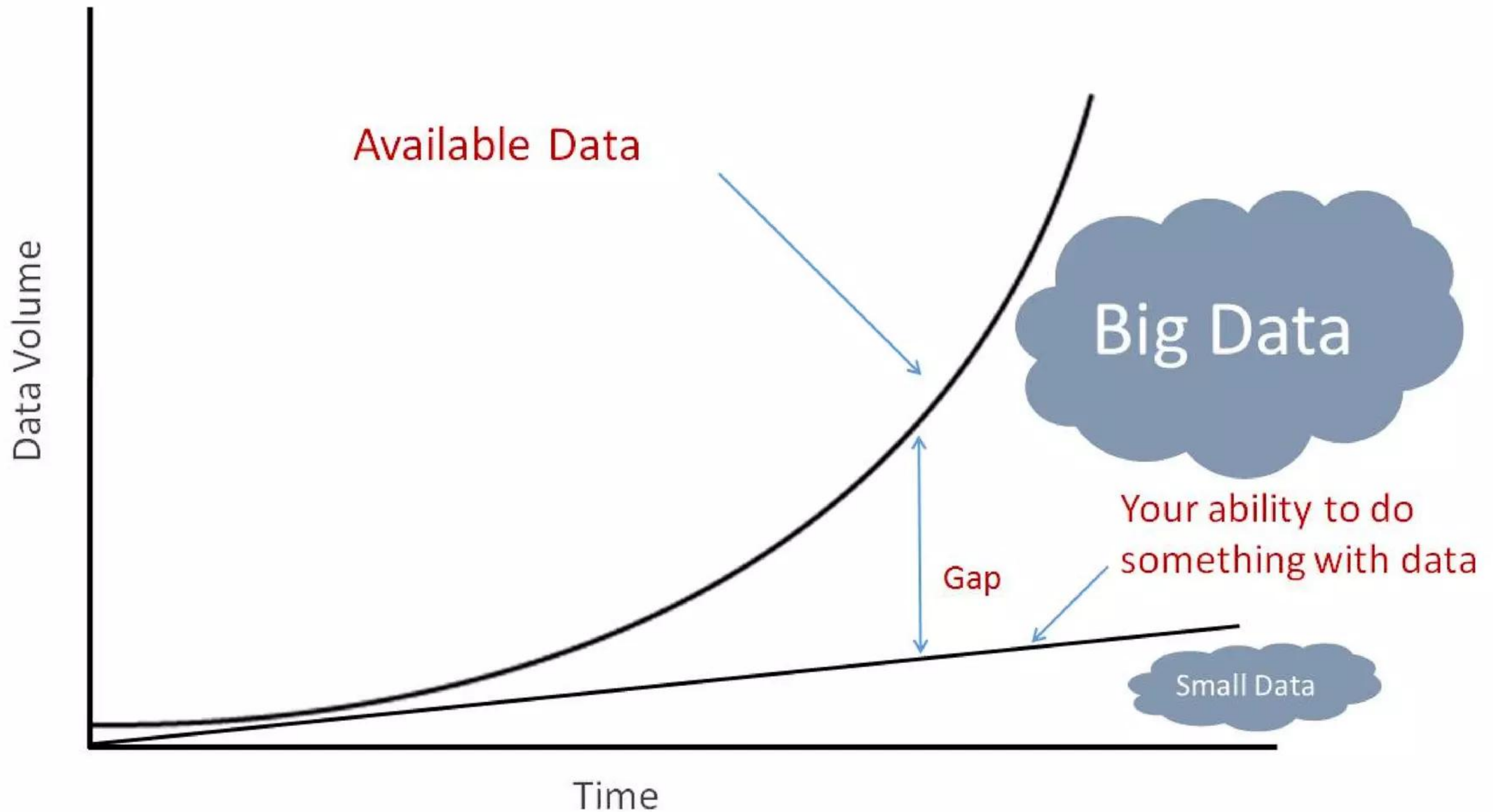
ML ALGORITHMS

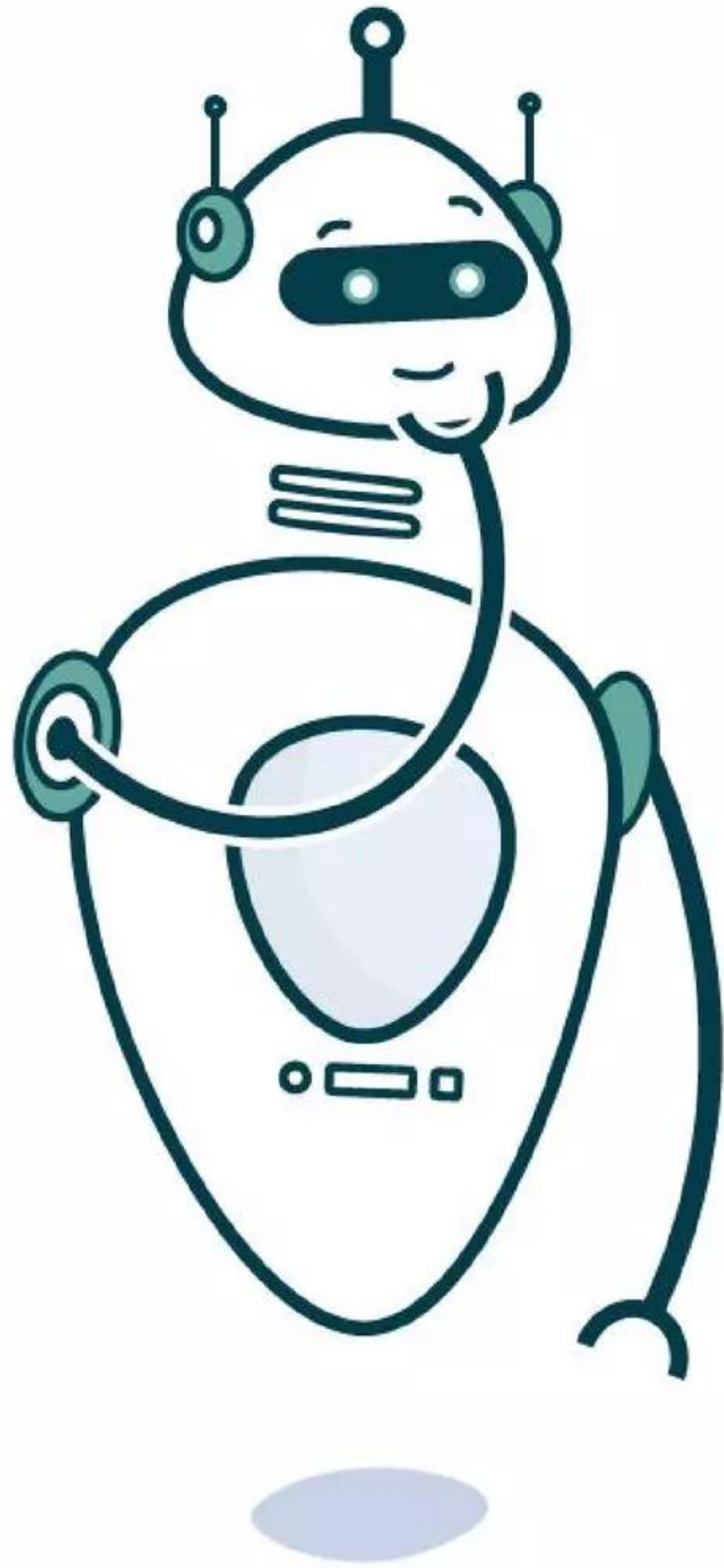
TECHGRABYTE

Accelerating Innovation & Change – every 60 seconds



Data Explosion



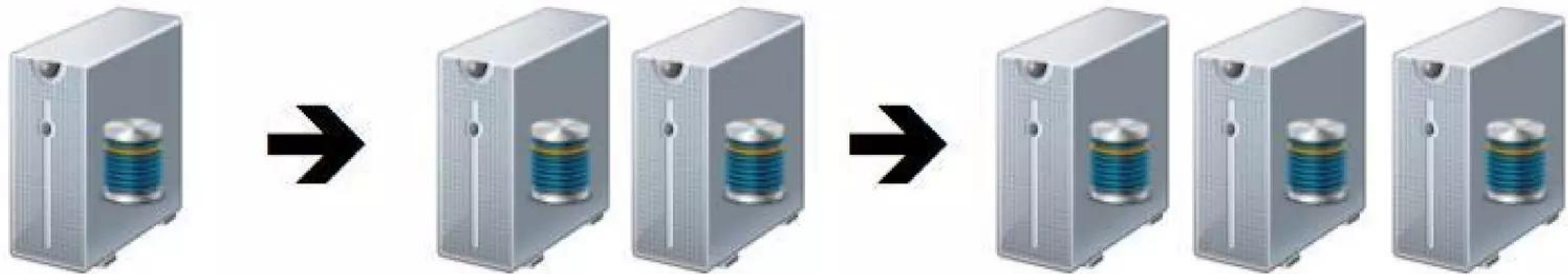


Solutions #1

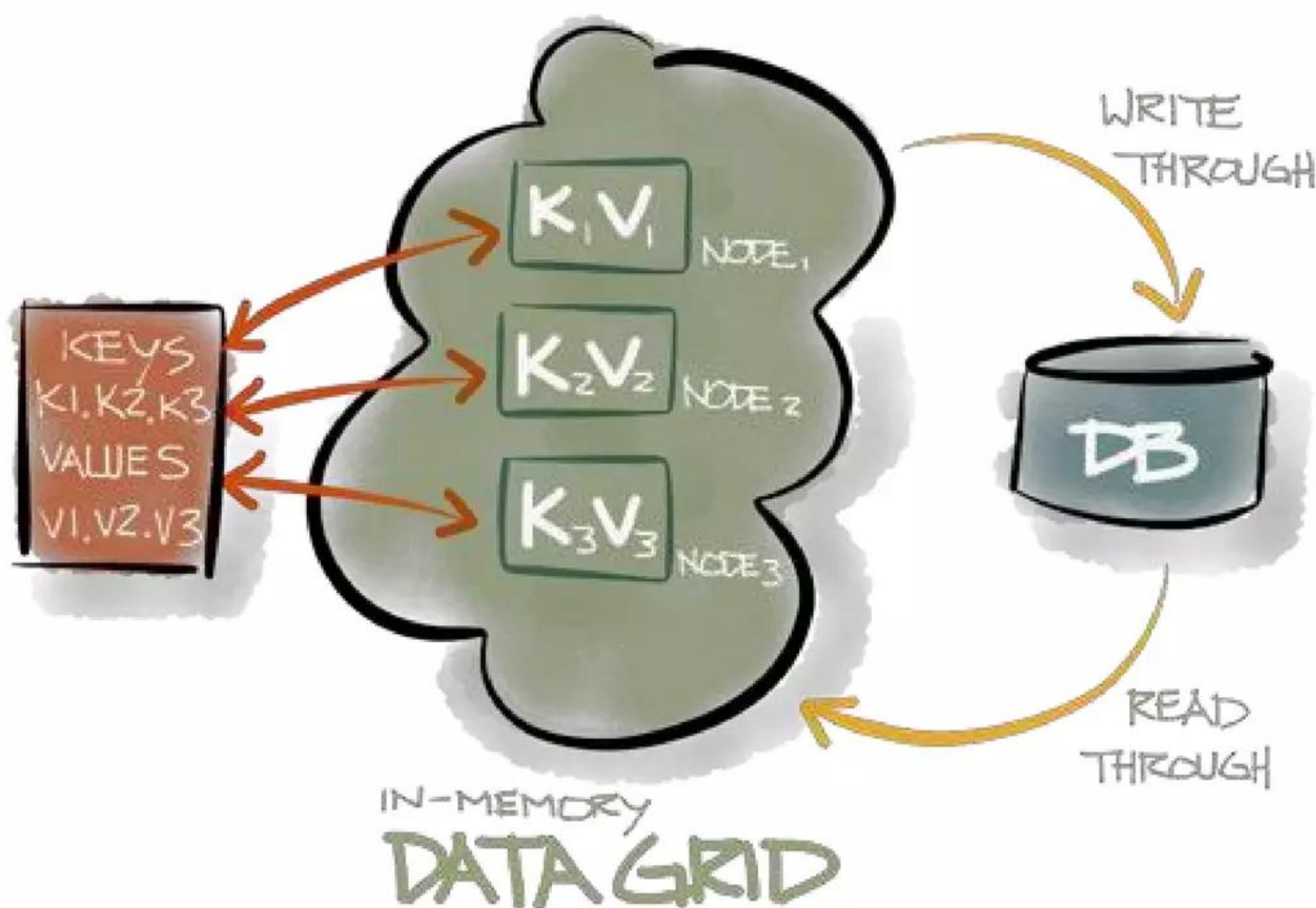
Scale-Up



Scale-Out



- In-memory data fabric: provides low-latency access and processing of large quantities of data by distributing data across the dynamic random access memory (DRAM), Flash, or SSD of a distributed computer system



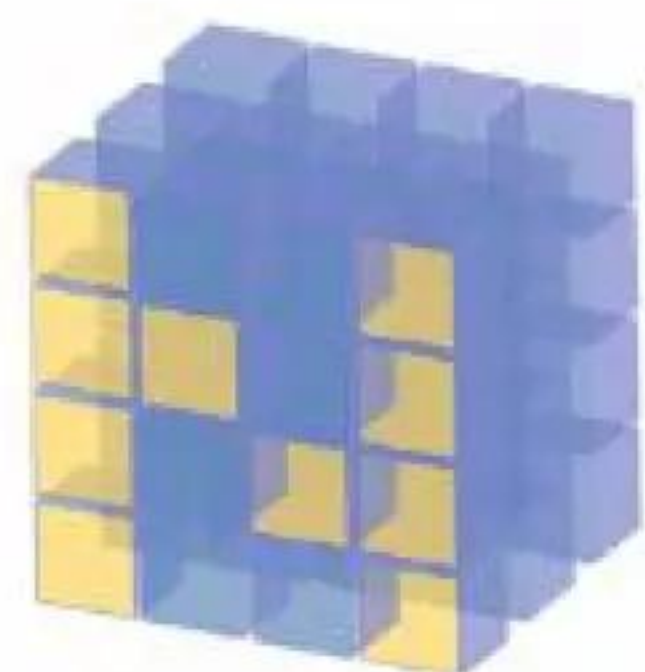
- Cluster machine
- GPU Machine (OpenCL and nVidia CUDA)



- Numpy
 - Scipy
 - Pandas
 - Scikit-learn
 - Matplotlib
 - Seaborn
 - Tensorflow

 - *pydata.org
 - *anaconda
- Other Tech (to support ML):
 - Apache Kafka
 - Apache Spark
 - Db: mongo, postgre
 - elasticsearch
 - CUDA/OpenCL

- Weka
- Deeplearn4j (working with spark and gpu)
- H2O (working with spark and GPU, support tensor, mxnet, and caffe)
- JcuDNN (JNI wrapping nvidia cuDNN)
- Mahout (hadoop)
- Mllib Spark



NumPy

Seaborn



TensorFlow™

Pandas



elasticsearch

APACHE
Spark™



OpenCL™

APACHE
kafka®

Stack and Services

MACHINE INTELLIGENCE 3.0

ENTERPRISE INTELLIGENCE

VISUAL Orbital Insight planet clarifai DEEP VISION cortica Iqonic SPACE_KNOW Captricity netra deepomatic	AUDIO Gridspace TalkIQ nexidia twilio CAPIO Expect Labs Clover Mobvoi Curious.AI popUP archive	SENSOR PREDIX IoT MAANA Sentenai PLANET OS UPTAKE IMUBIT Preferred Networks thingworx KONUX Alluvium	INTERNAL DATA PRIMER IBM WATSON Cyrcorp Palantir ARIMO Alation Sapho Outlier Digital Reasoning	MARKET mattermark Quid DataFox PREMISE Bottlenose MOTIVA enigma CB INSIGHTS Tracxn predata
--	--	---	---	--

ENTERPRISE FUNCTIONS

CUSTOMER SUPPORT DigitalGenius Kasisto ELOQUENT Wiseio ACTIONIQ zendesk Preact CLARABRIDGE	SALES collective[i] sense fuse machines AVISO salesforce INSIDE SALES.COM Zensight clari	MARKETING MINTIGO Lattice RADIUS LiftIgniter PERSADO brightLumel retention MOTIVA COGNICOR AIRPR msg ai	SECURITY CYLANCE DARKTRACE ZIMPERIUM deepinstinct Sentinel DEMISTO graphistry drawbridge SignalSense AppZen	RECRUITING textio entelo Wade & Wendy hiQ unilive SpringRole GIGSTER HireVue
---	---	--	---	---

AUTONOMOUS SYSTEMS

GROUND NAVIGATION drive.ai AdasWorks ZOOX MOBILEYE UBER Google TESLA nuTonomy Auro Robotics	AERIAL SKYDIO SHIELD AI Airware DJI LILY DroneDeploy pilot.ai SKYCATCH	INDUSTRIAL JAYBRIDGE OSARO CLEARPATH fetch KINDRED rethink robotics HARVEST	PERSONAL amazon alexa Cortana Allo facebook Siri Replika	PROFESSIONAL butter.ai pogo SKIPFLAG clara x.ai slack talla Zoom sudo
--	---	--	---	---

INDUSTRIES

AGRICULTURE BLUE RIVER MAVIX tule TRACE Pivotal Bio TerraVision AGRI-DATA Descartes Labs udio abundant robotics	EDUCATION KNEWTON volley gradescope CTI coursera UUDACITY alt school	INVESTMENT Bloomberg sentient iSENTIUM KENSHO alpha.sense Dataminr CEREBELLUM CAPITAL Quandl	LEGAL blueJ BEAGLE Everlaw RAVEL seal ROSS LEGAL ROBOT	LOGISTICS NAUTO Acerta PRETECKT clearmetal Routific MARBLE PITSTOP
--	---	---	---	--

INDUSTRIES CONT'D

MATERIALS zymergen Citrine Eigen Innovations SIGHT MACHINE GINKGO BIOWORKS nanotronics CALCULARIO	RETAIL FINANCE TALA zest finance Lendo earnest affirm MIRADOR wealthfront Betterment	PATIENT PULSE CareSkore ZEPHYR HEALTH Watson Health Oncoda SENTRIAN Atomwise Numerate	IMAGE BUTTERFLY 3SCAN ARTERYS enlitic BAYLABS imagia Google DeepMind	BIOLOGICAL iCarbonX color GRAIL deep genomics RECURSION LUMINIST Numerate Atomwise verily WHOLE BIOME
---	---	--	---	--

TECHNOLOGY STACK

AGENT ENABLERS
 OCTANE.AI howdy Maluuba KITT.AI
 OpenAI Gym Kasisto AUTOMAT
 semanticmachines

DATA SCIENCE
 DOMINO SPARKBEYOND rapidminer
 kaggle DataRobot yhat AYASDI
 data iku seldon yseop bigml

MACHINE LEARNING
 CognitiveScale GoogleML context relevant
 Cyrcorp HyperScience nora logics minds.ai H2O.ai
 SCALED INFERENCE sparkcognition loop GEOMETRIC INTELLIGENCE
 deepsense.io reactive skymind bonsai

NATURAL LANGUAGE
 agolo FYLIEN LEXALYTICS
 Narrative Science loop spaCy LUMINOSO
 cortical.io MonkeyLearn

DEVELOPMENT
 SIGOPT HyperOpt fuzzyio okite
 rainforest lobe Anodot
 Signifai LAYER 6 bonsai

DATA CAPTURE
 CrowdFlower diffbot CrowdAI import io
 Paxata DATASIFT amazon mechanicalturk enigma
 WorkFusion DATALOGUE TRIFACTA parsehub

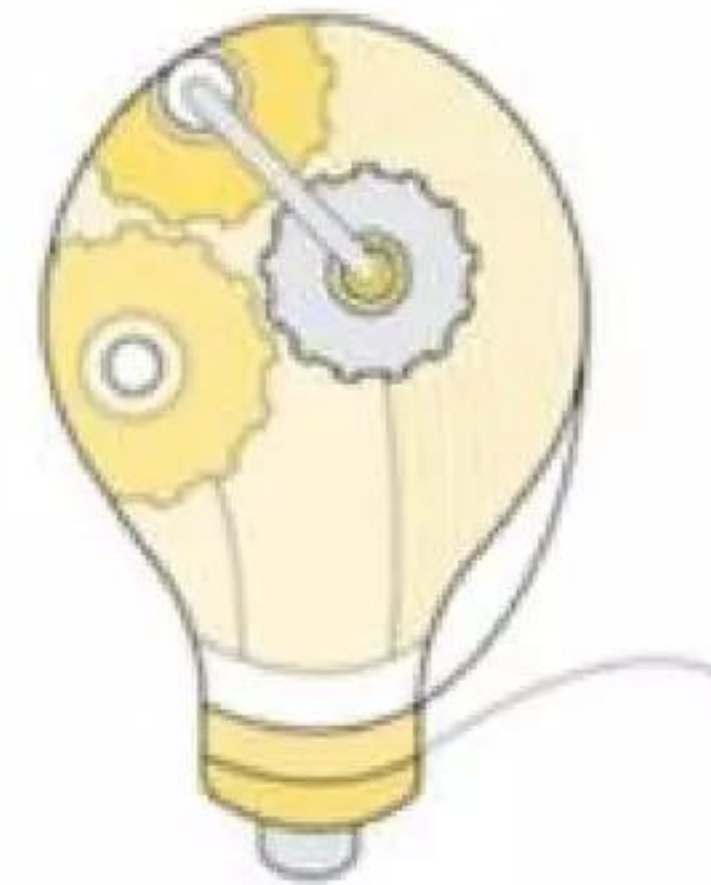
OPEN SOURCE LIBRARIES
 Keras Chainer CNTK TensorFlow Caffe
 H2O DEEPLARNING4J theano torch
 DSSTNE Scikit-learn AzureML neon
 MXNet DMTK Spark PaddlePaddle WEKA

HARDWARE
 KNUPATH TENSTORRENT Cirascale
 NVIDIA intel nervana Movidius
 tensilica GoogleTPU 10²⁶ Labs Qualcomm
 Cerebras Isosemi

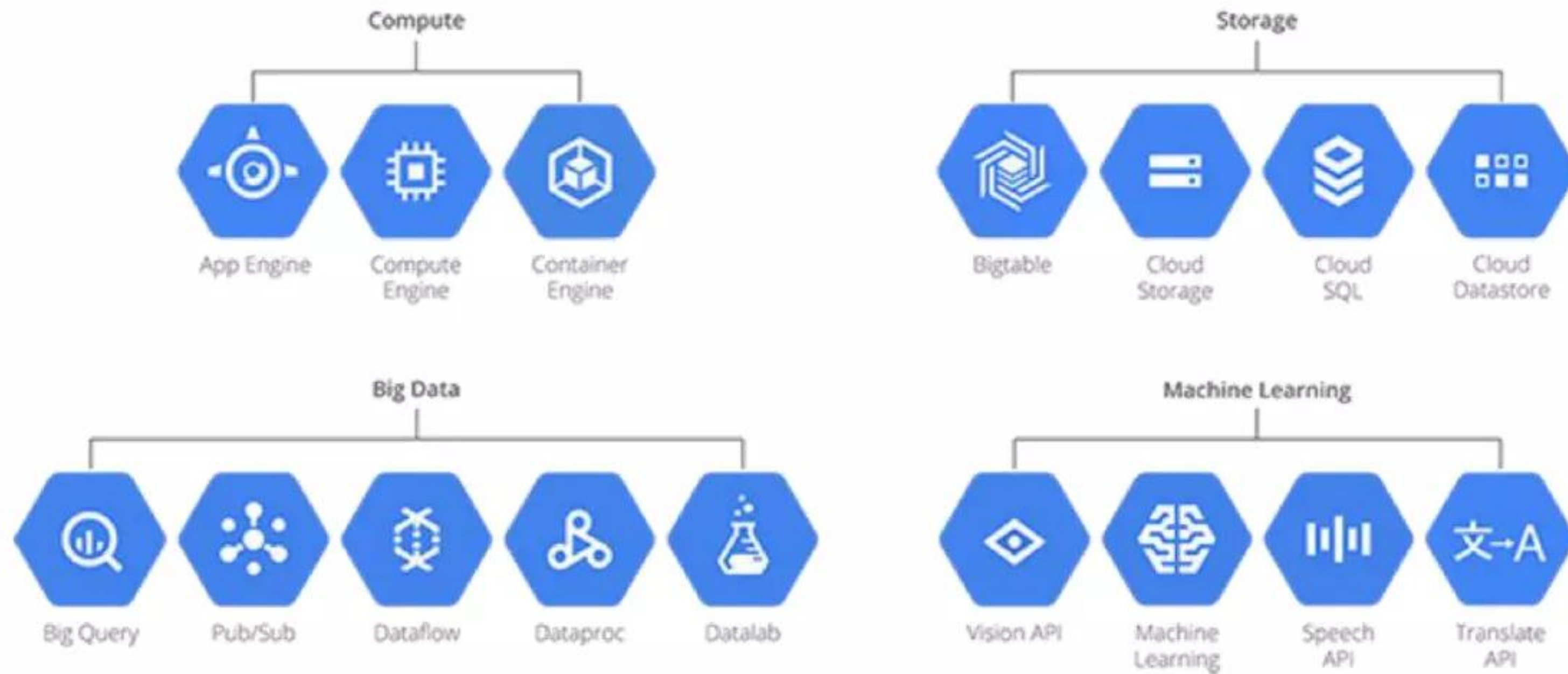
RESEARCH
 OpenAI nraisense ELEMENT^{AI} vicarious
 KNOGGIN Numenta Kimera Systems Cogitai



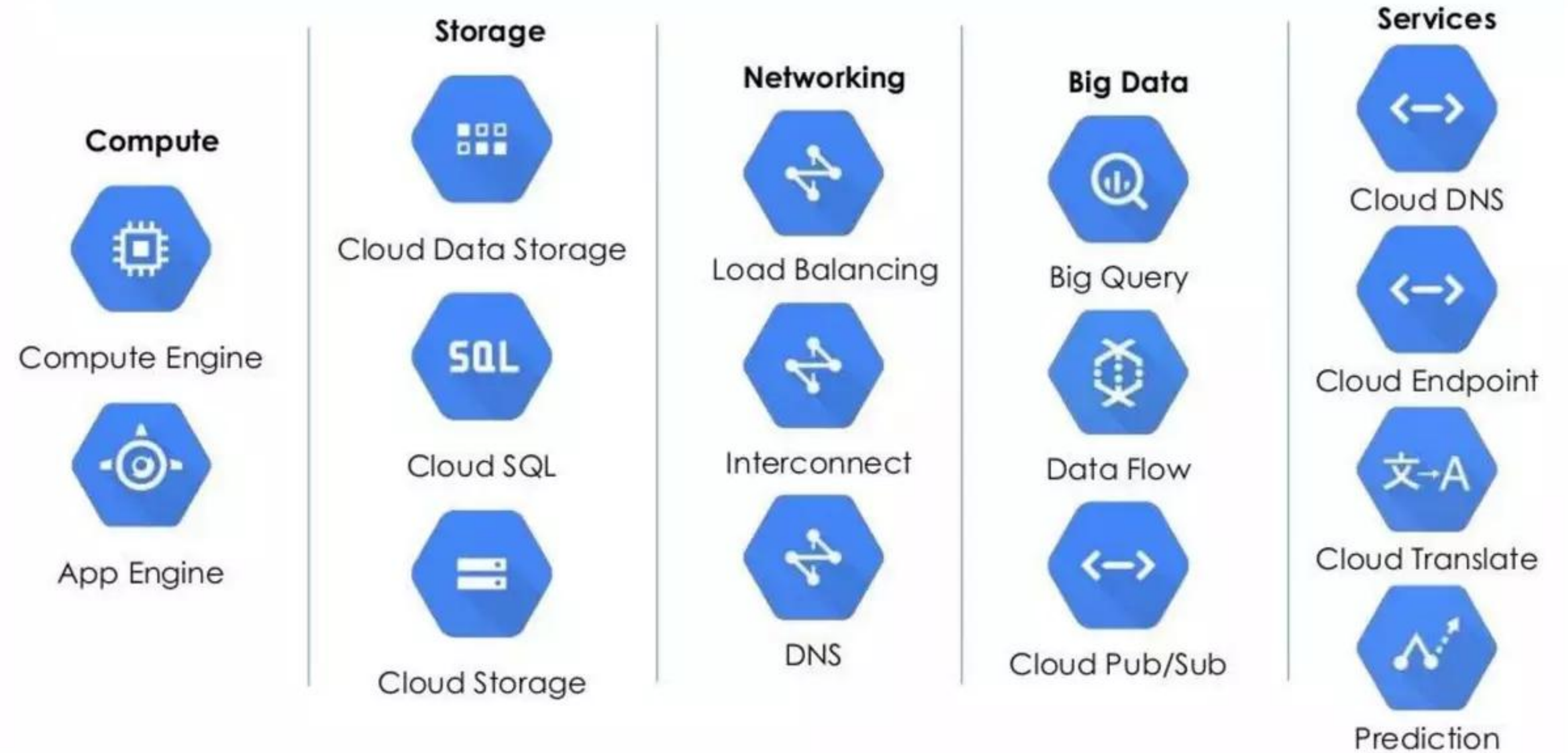
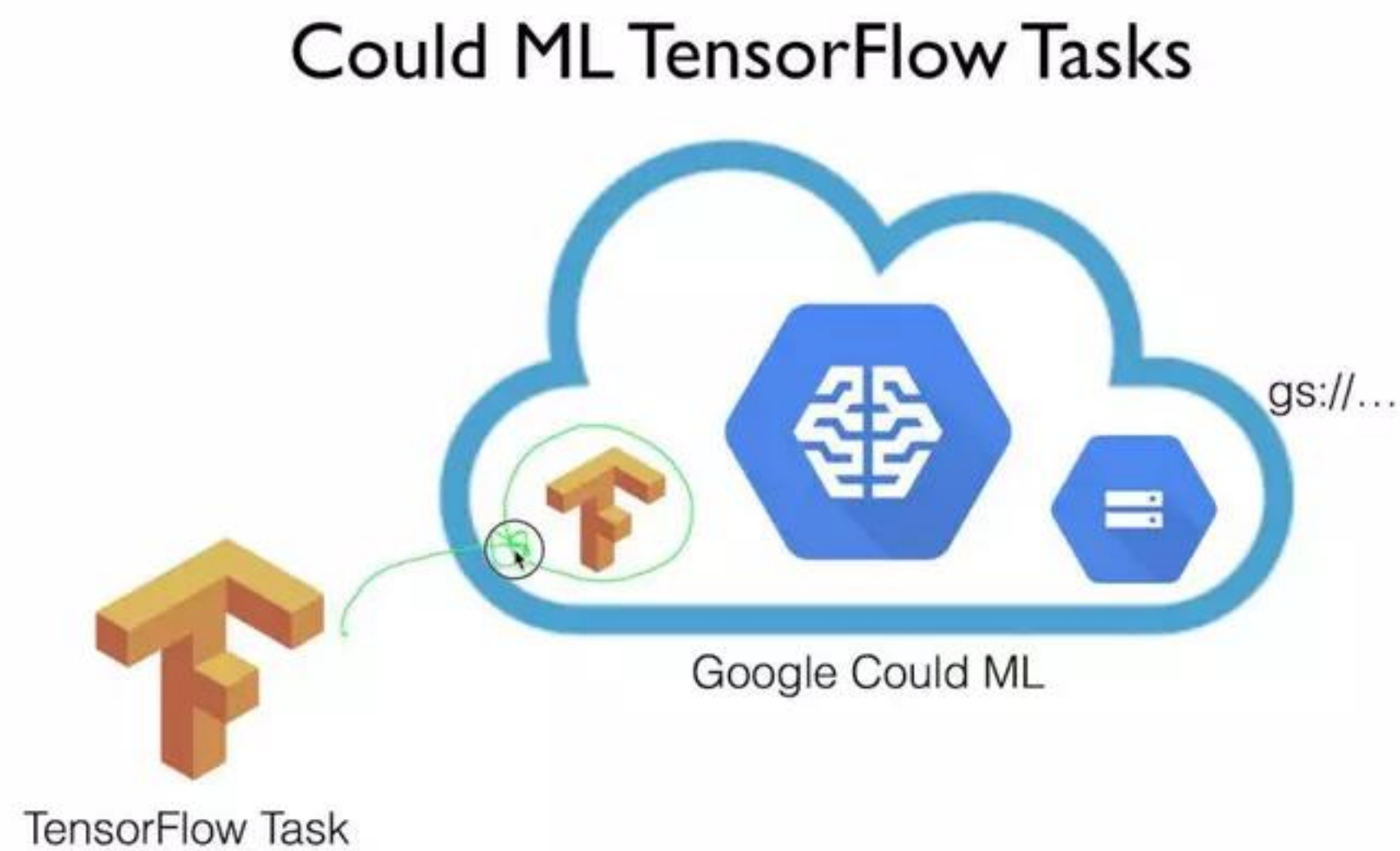
Google Cloud Platform



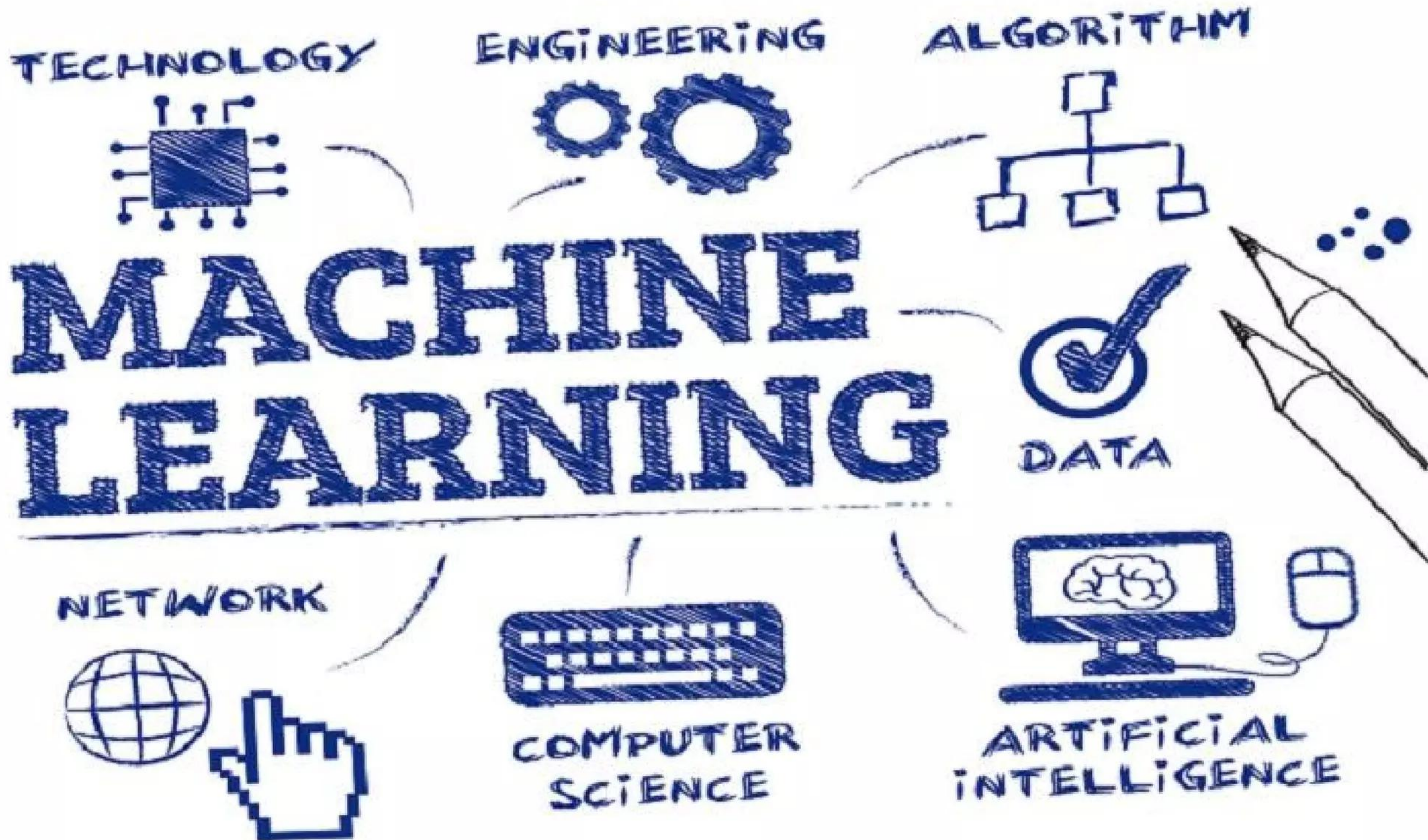
Google Cloud Platform



Google Cloud Platform

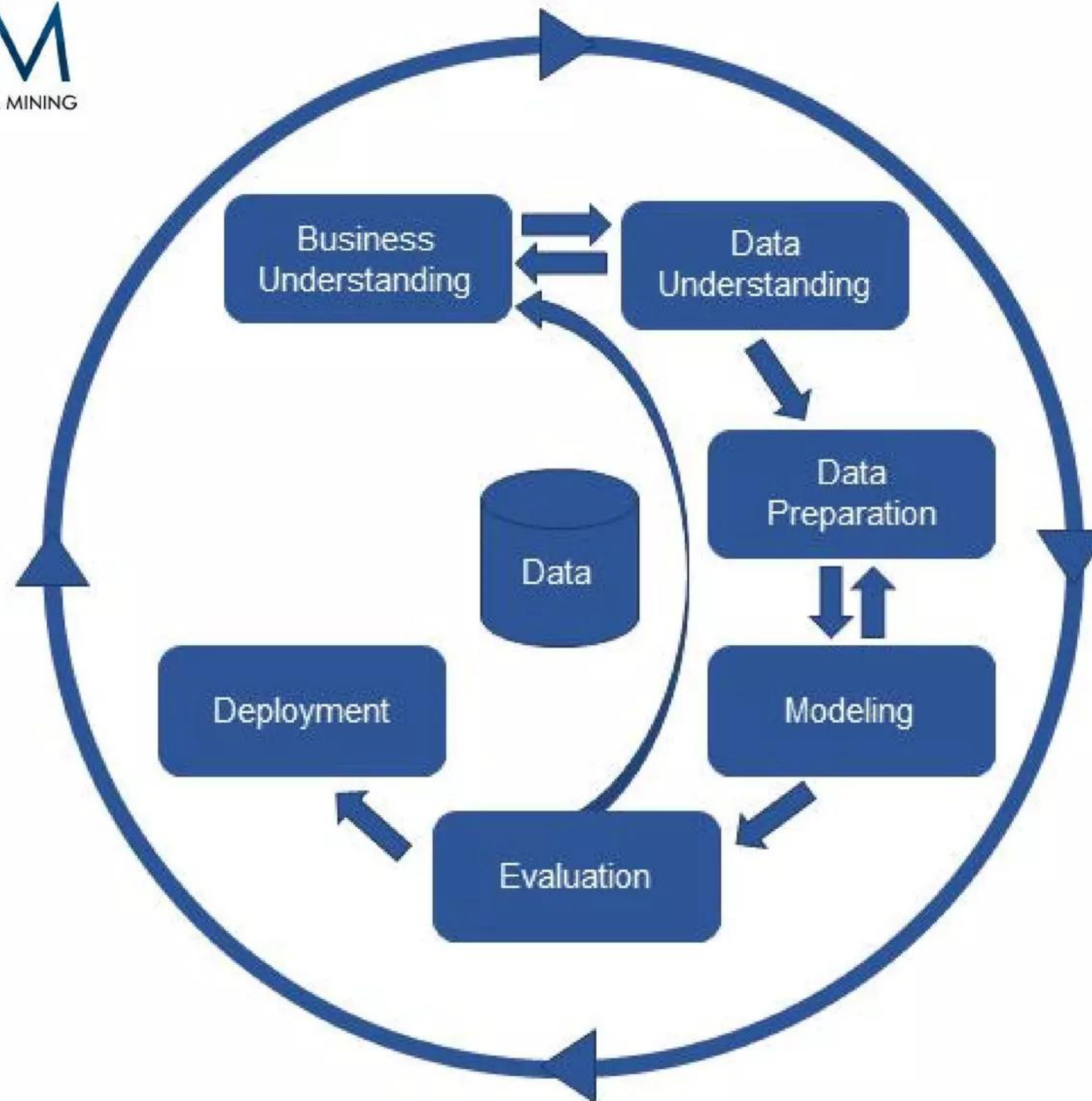


Compile all components

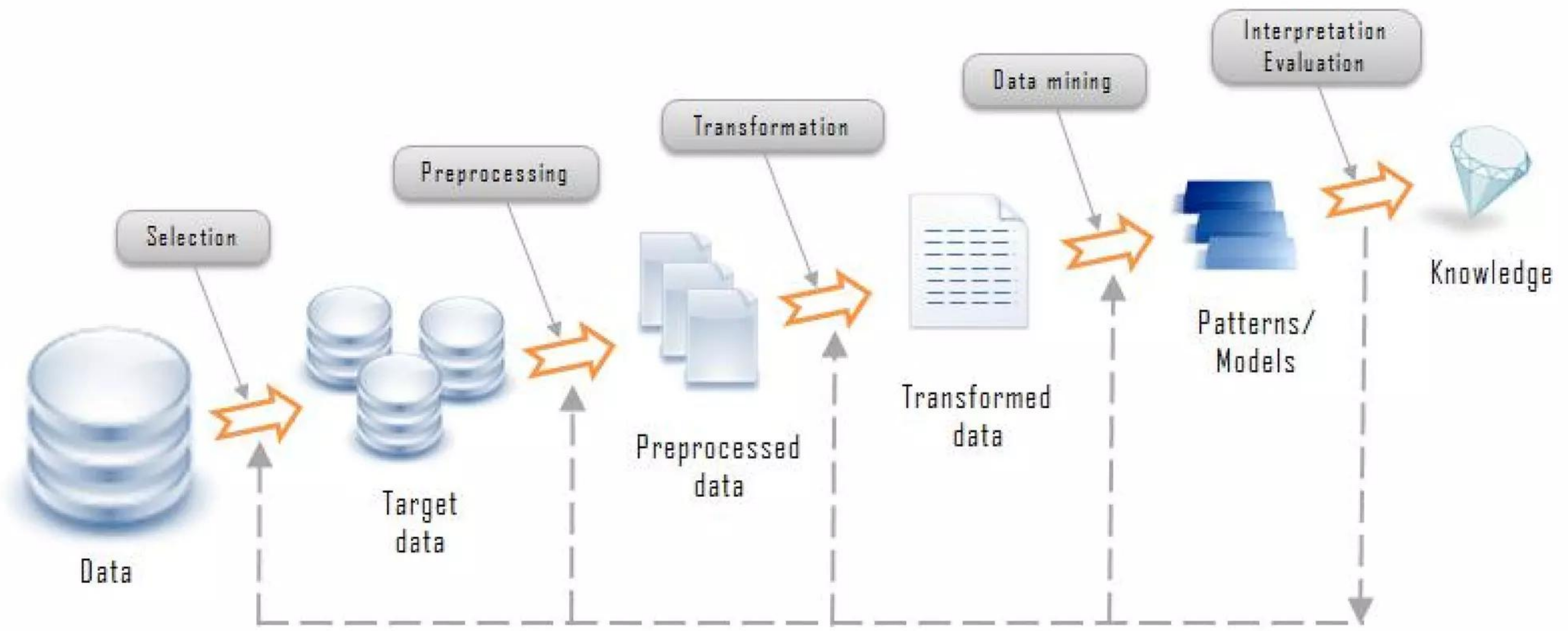


How we applied Machine Learning

CRISP-DM
CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING

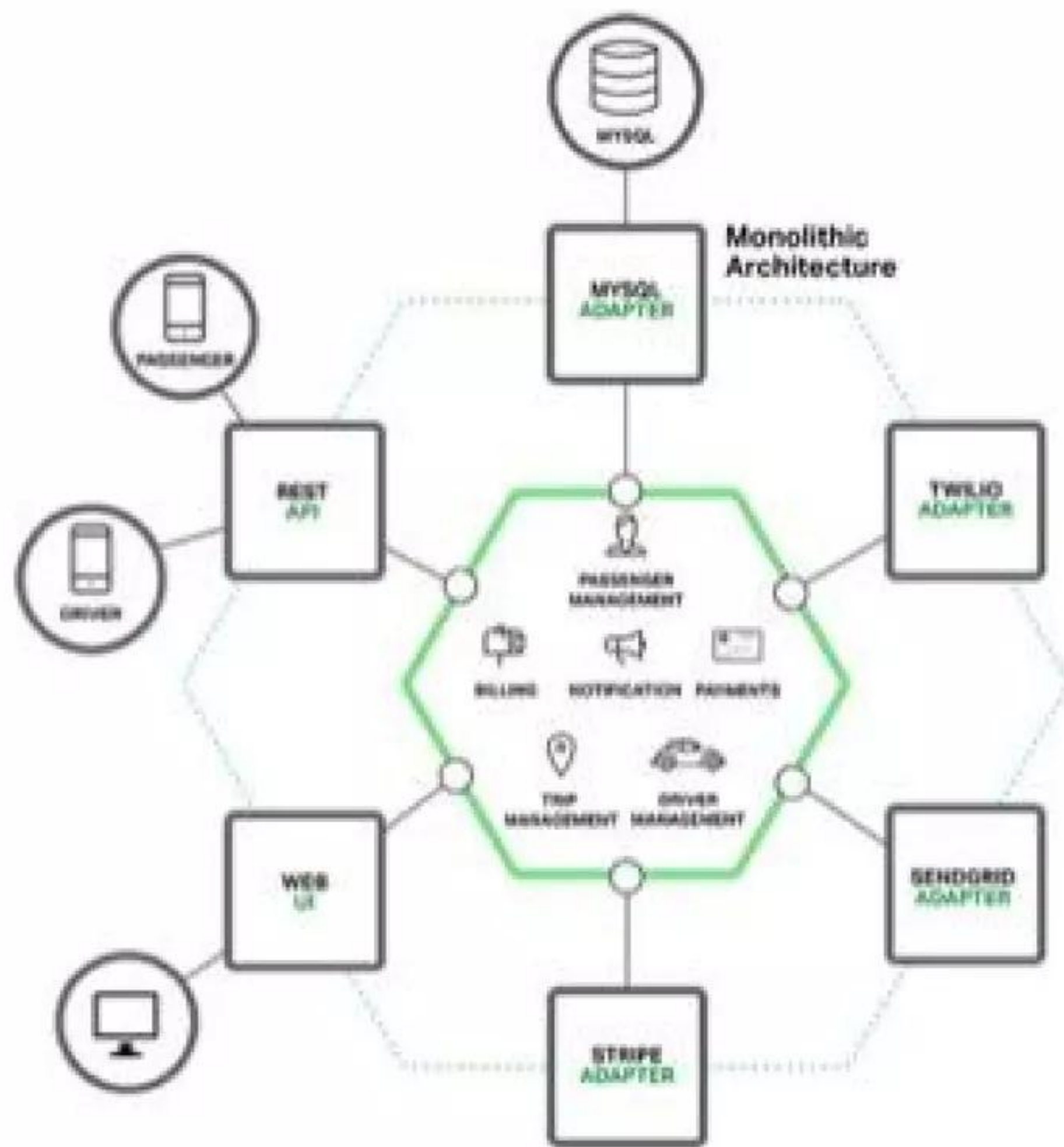


Process mining

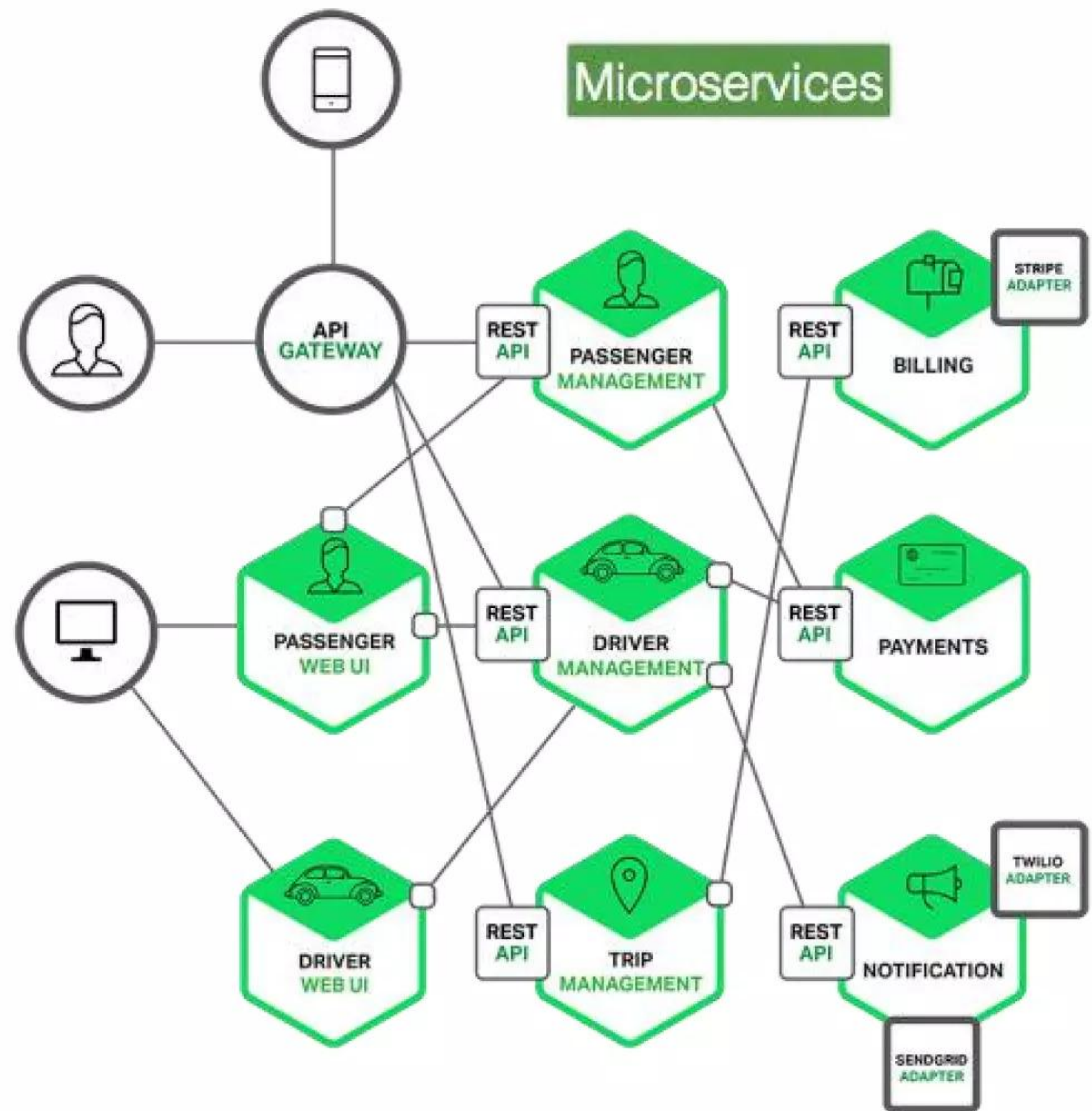


Another problem

- We are using micro services



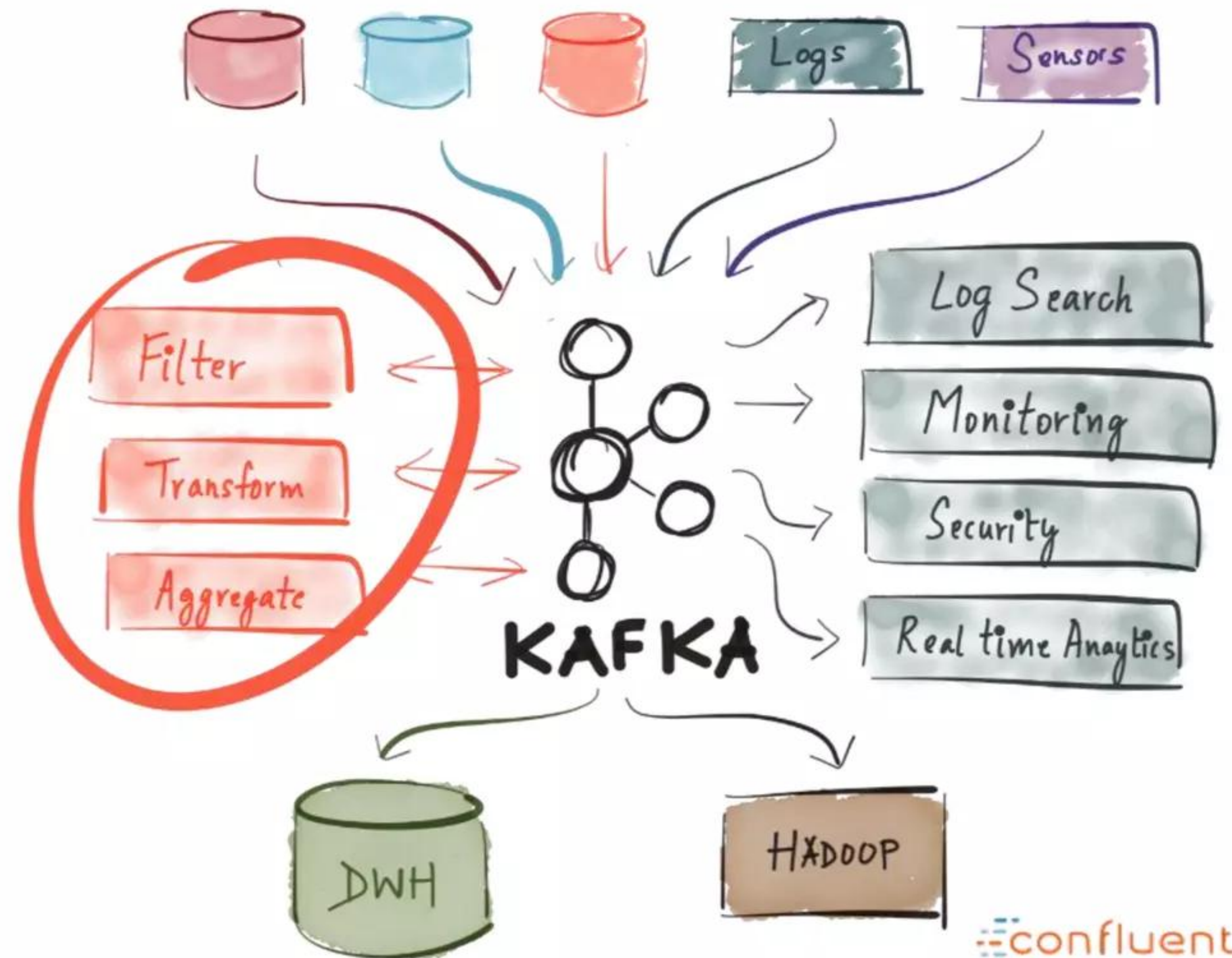
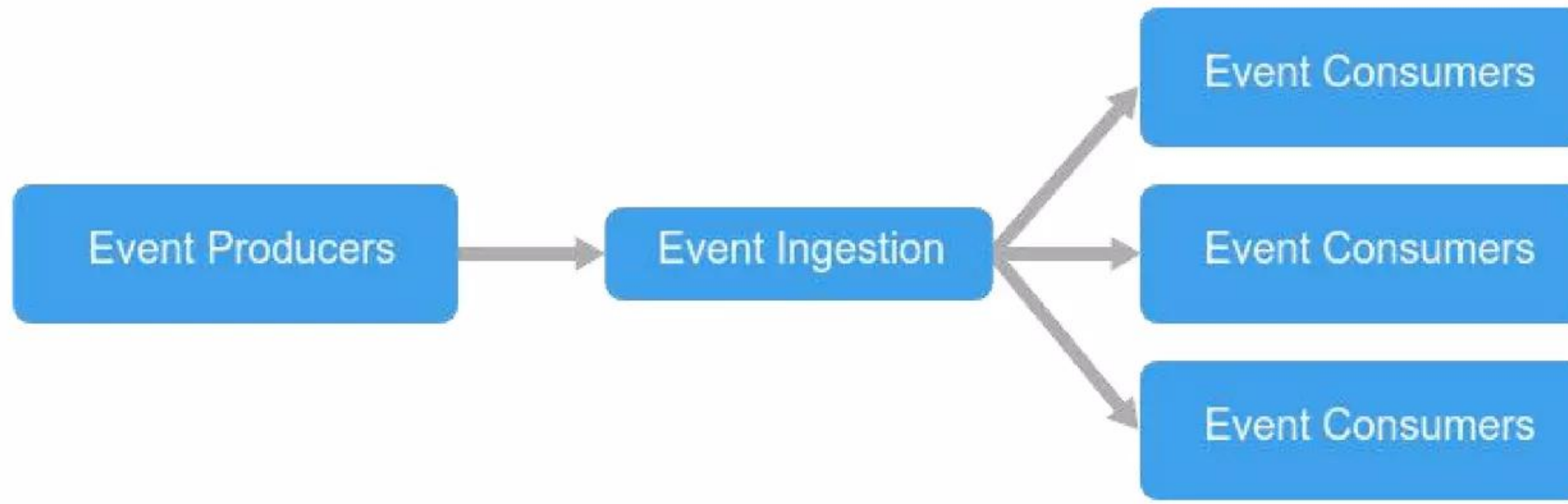
Monolithic Application



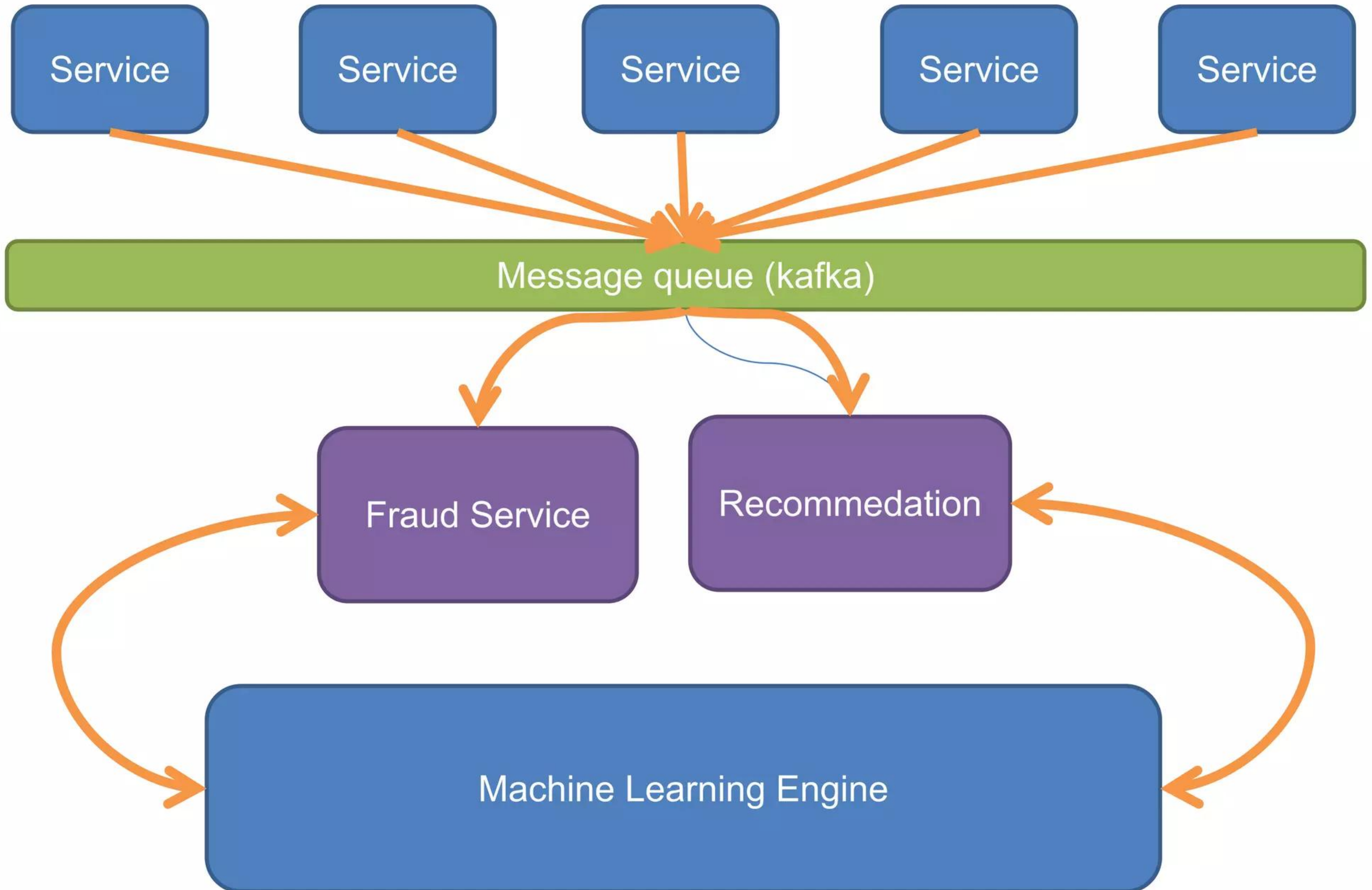
How to Collect



Event Drive Architecture

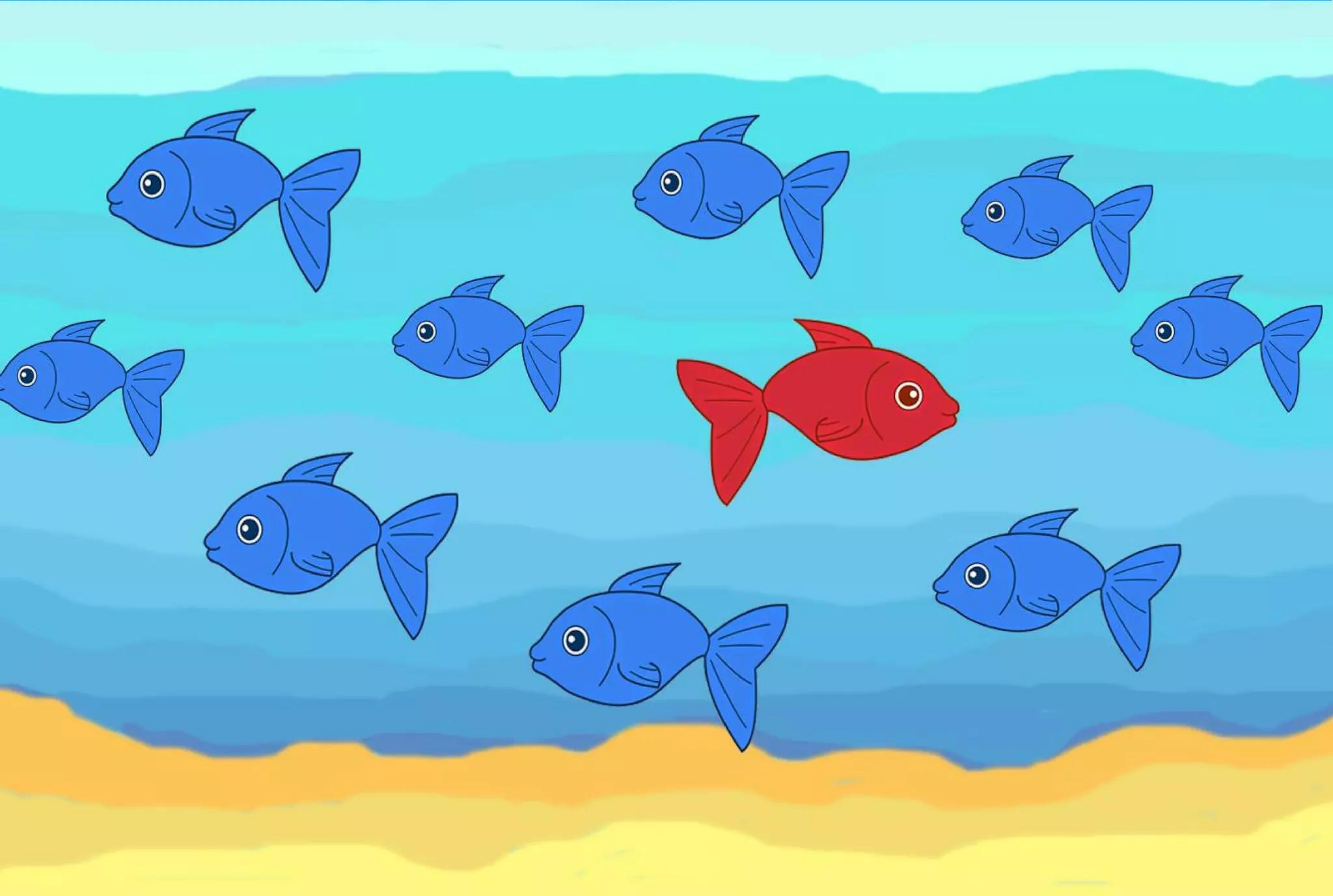


Event Drive Architecture





Anomaly Detection



- Anomalies are patterns in data that do not conform to a well defined notion of normal behavior
- These nonconforming patterns : outliers, discordant observations, exceptions, aberrations, surprises, peculiarities, or contaminants in different application domains

- Defining a normal region
- In many domains normal behavior keeps evolving
- Availability of labeled data for training/validation of models used by anomaly detection techniques is usually a major issue
- Often the data contains noise that tends to be similar to the actual anomalies

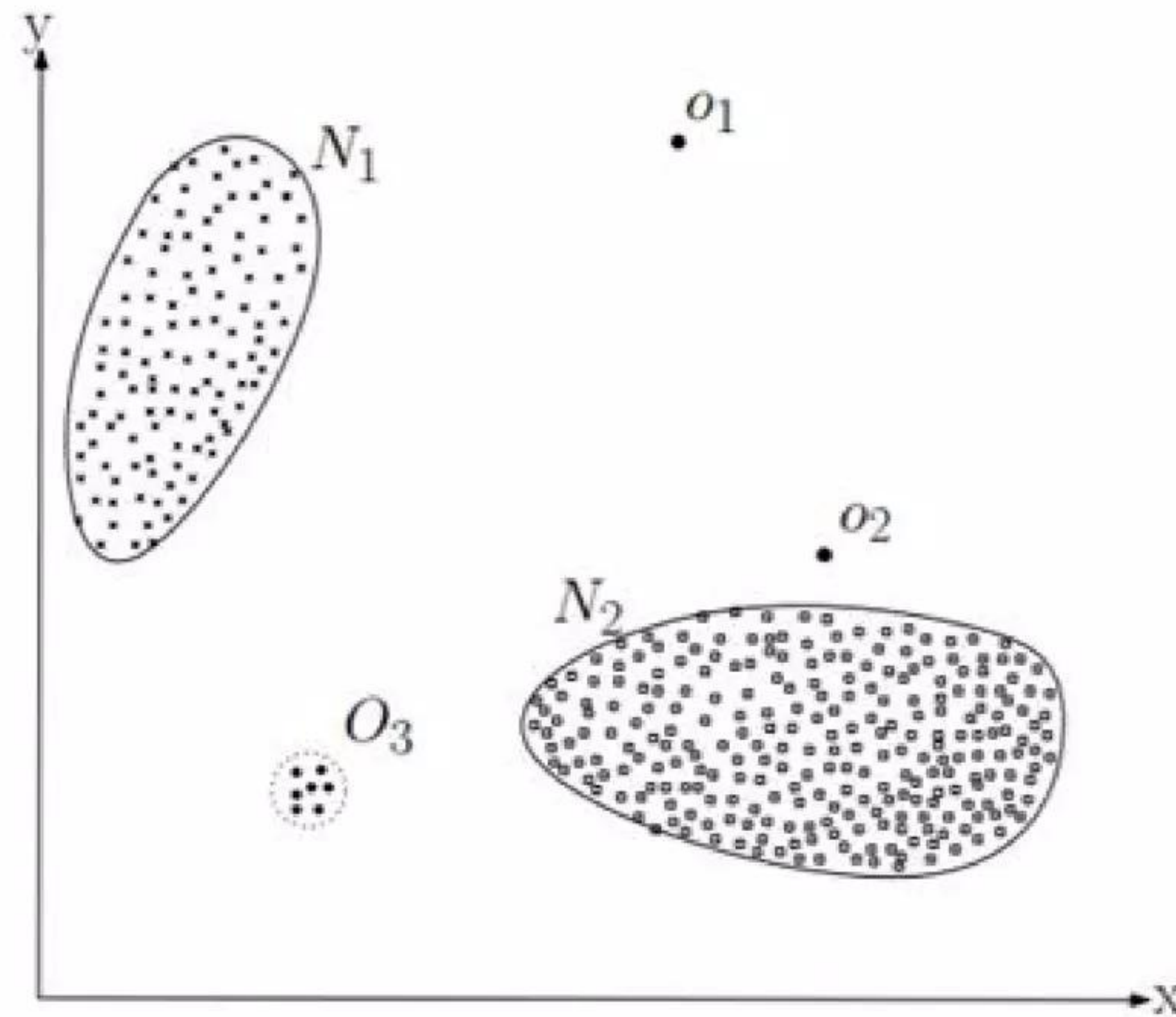


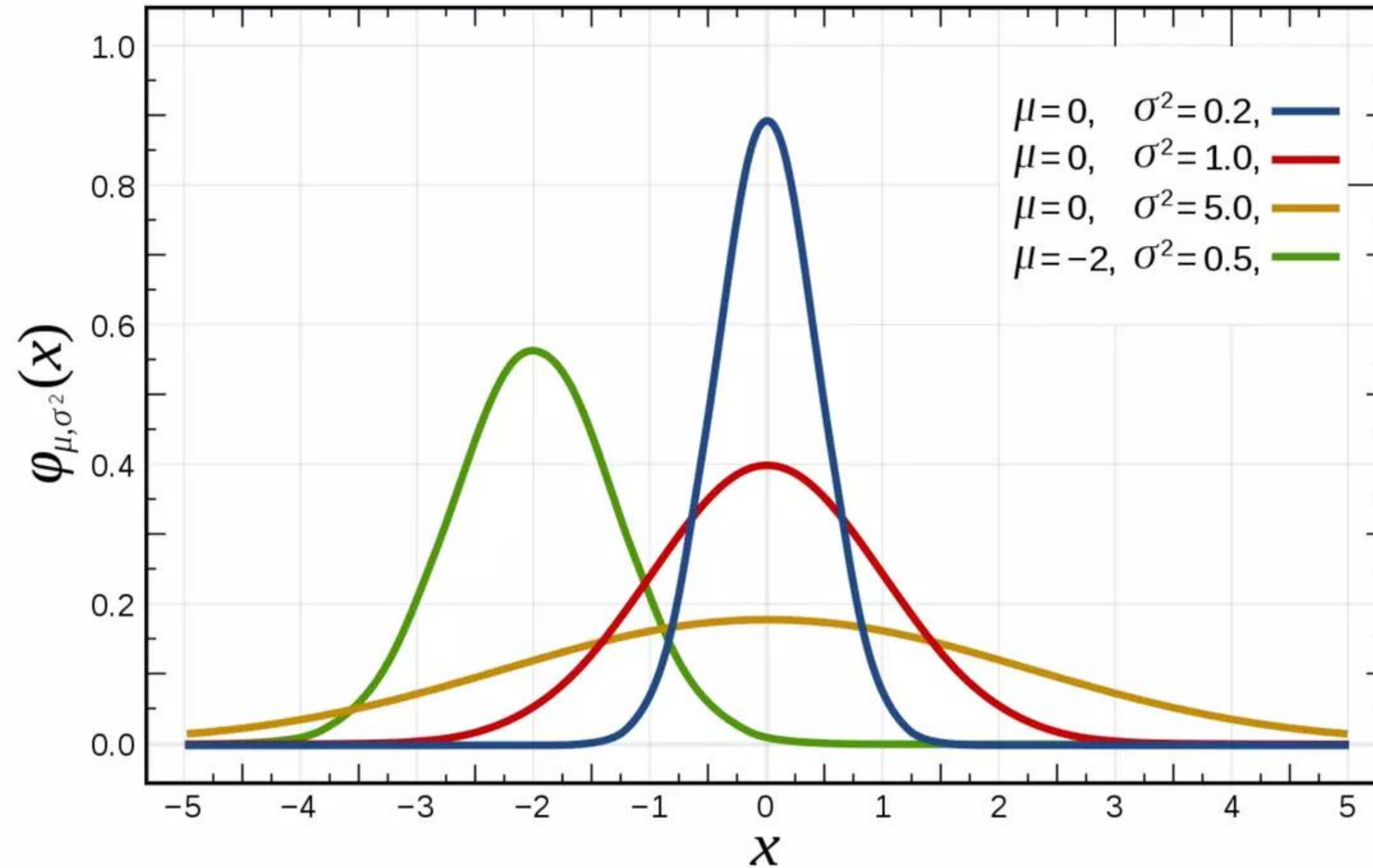
Fig. 1. A simple example of anomalies in a 2-dimensional data set.

Solution Method :

- Gaussian Mixture Model
- Fitted by EM - Algorithm

Gaussian Distribution

Before GMM, try to remember the gaussian distribution



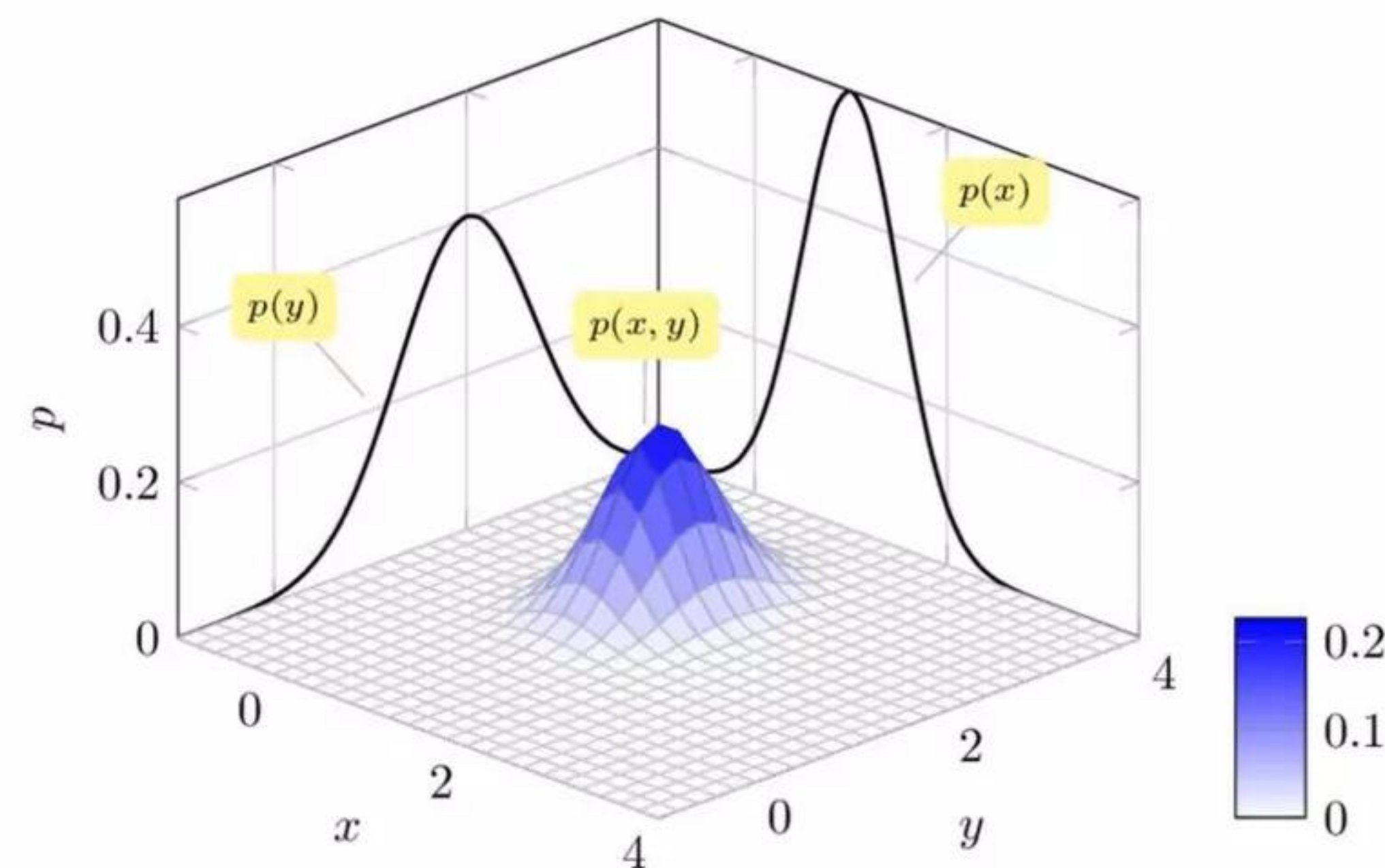
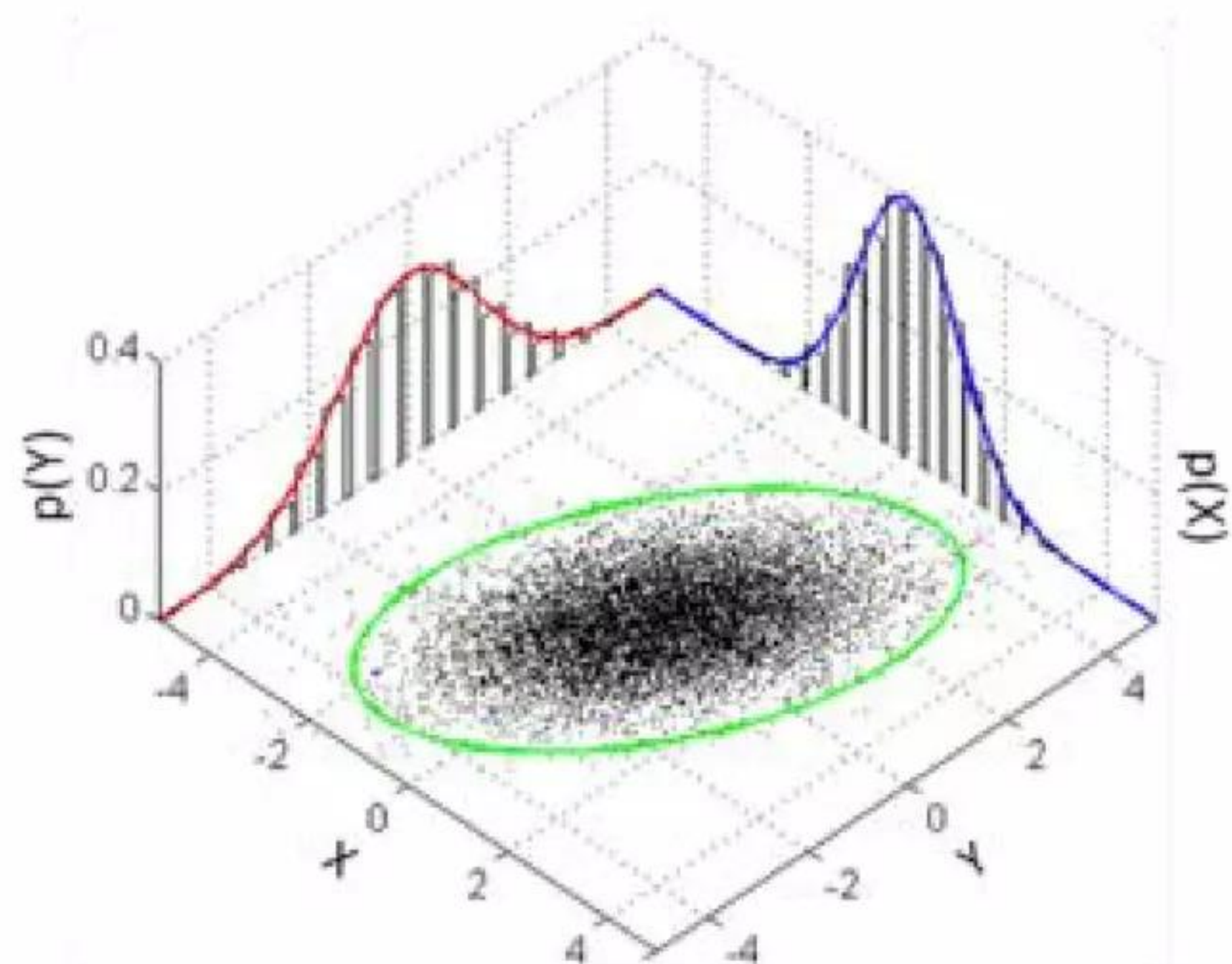
$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\phi(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

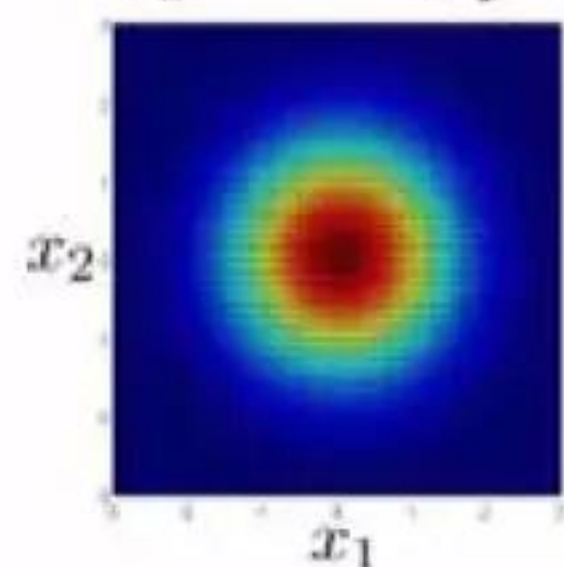
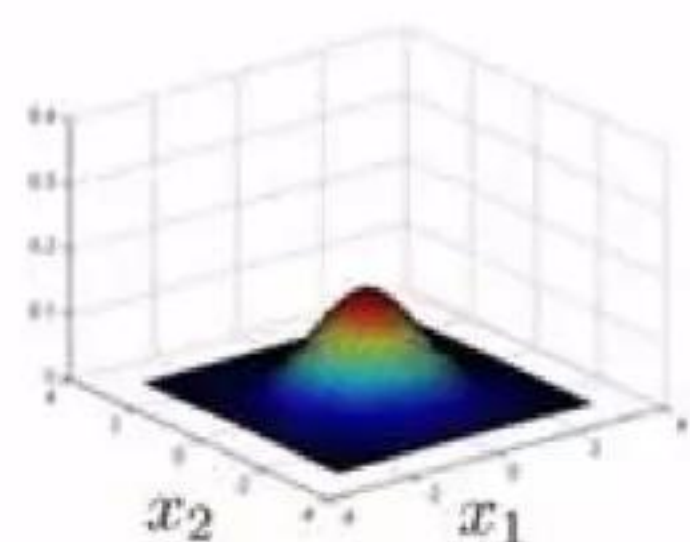
Gaussian Distribution Multivariate



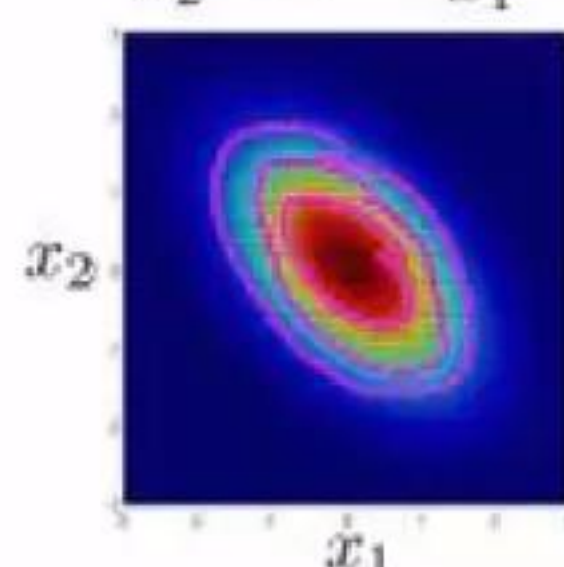
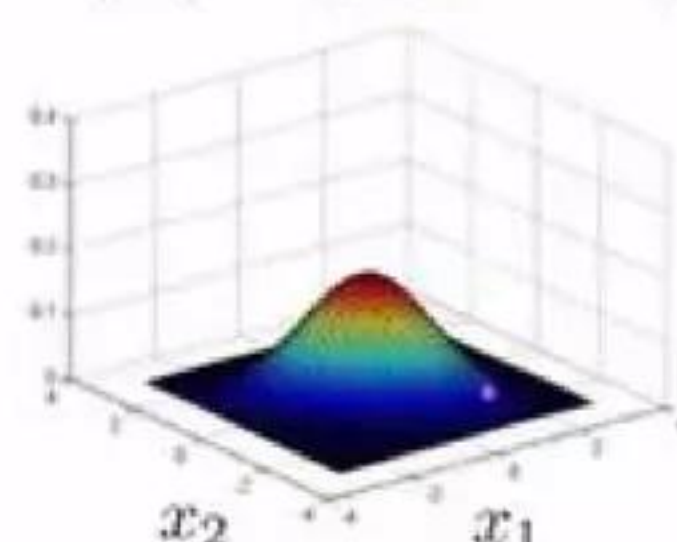
$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2} \|\boldsymbol{\Sigma}\|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]$$

Multivariate Gaussian (Normal) examples

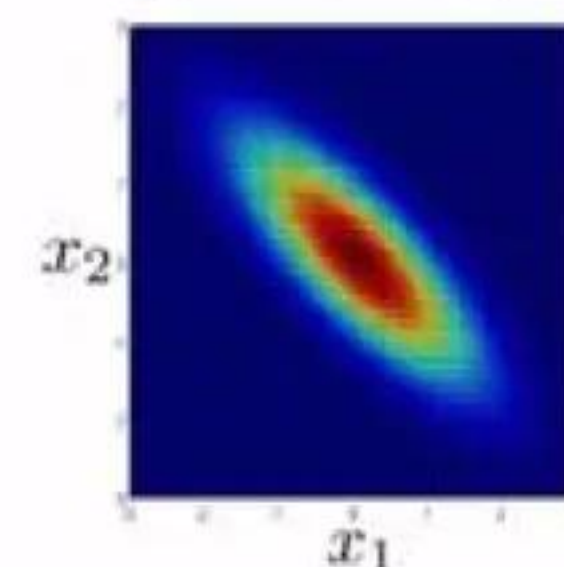
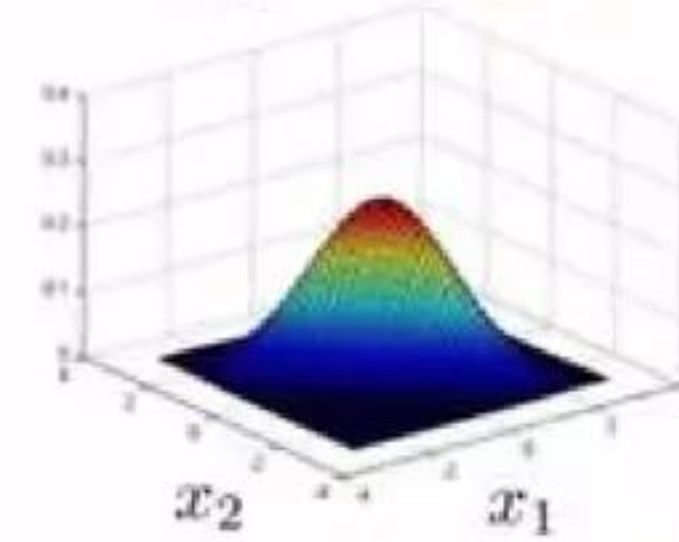
$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



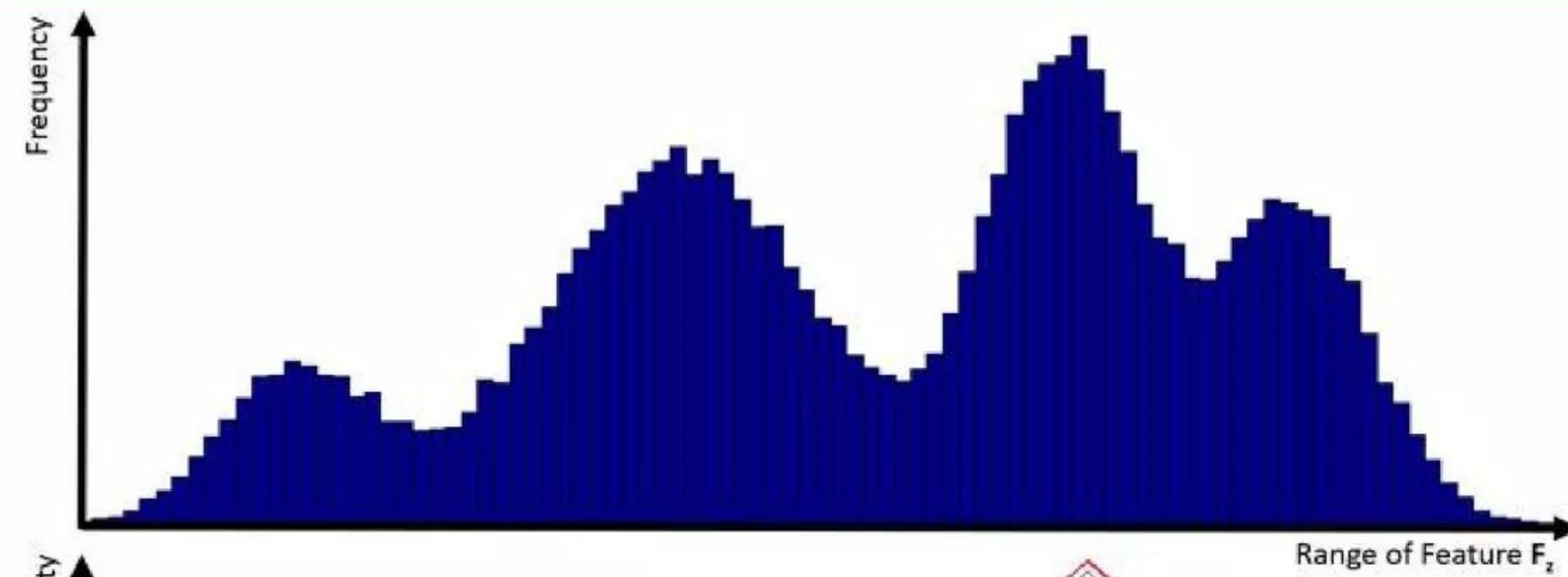
$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$



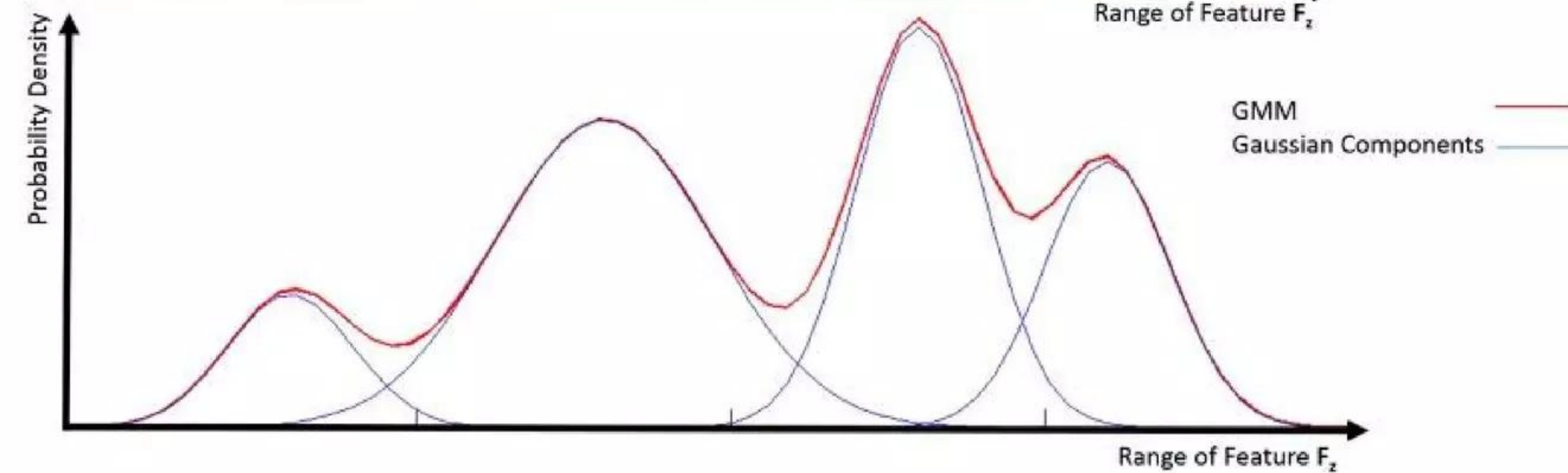
$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$



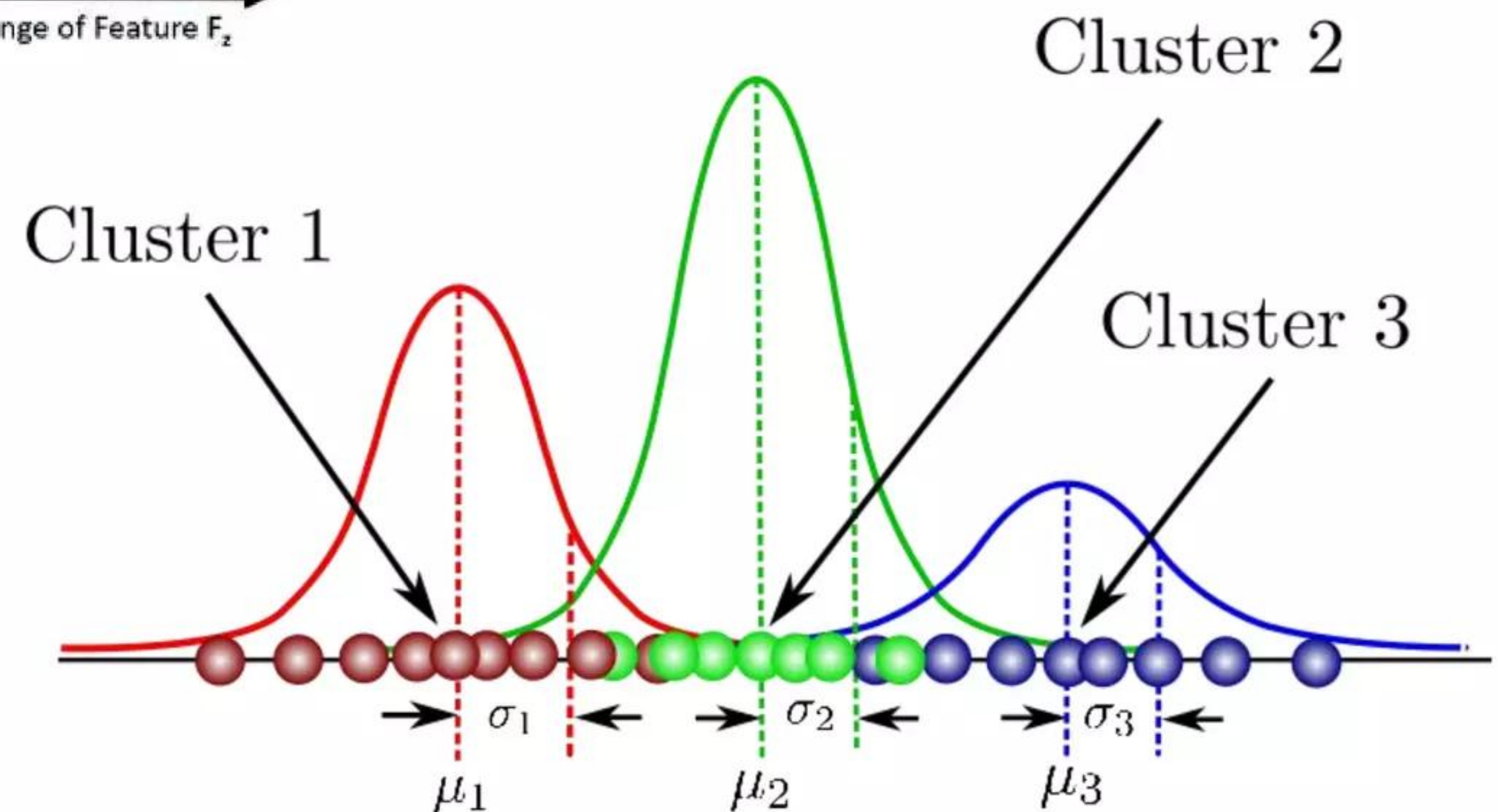
Gaussian Mixture



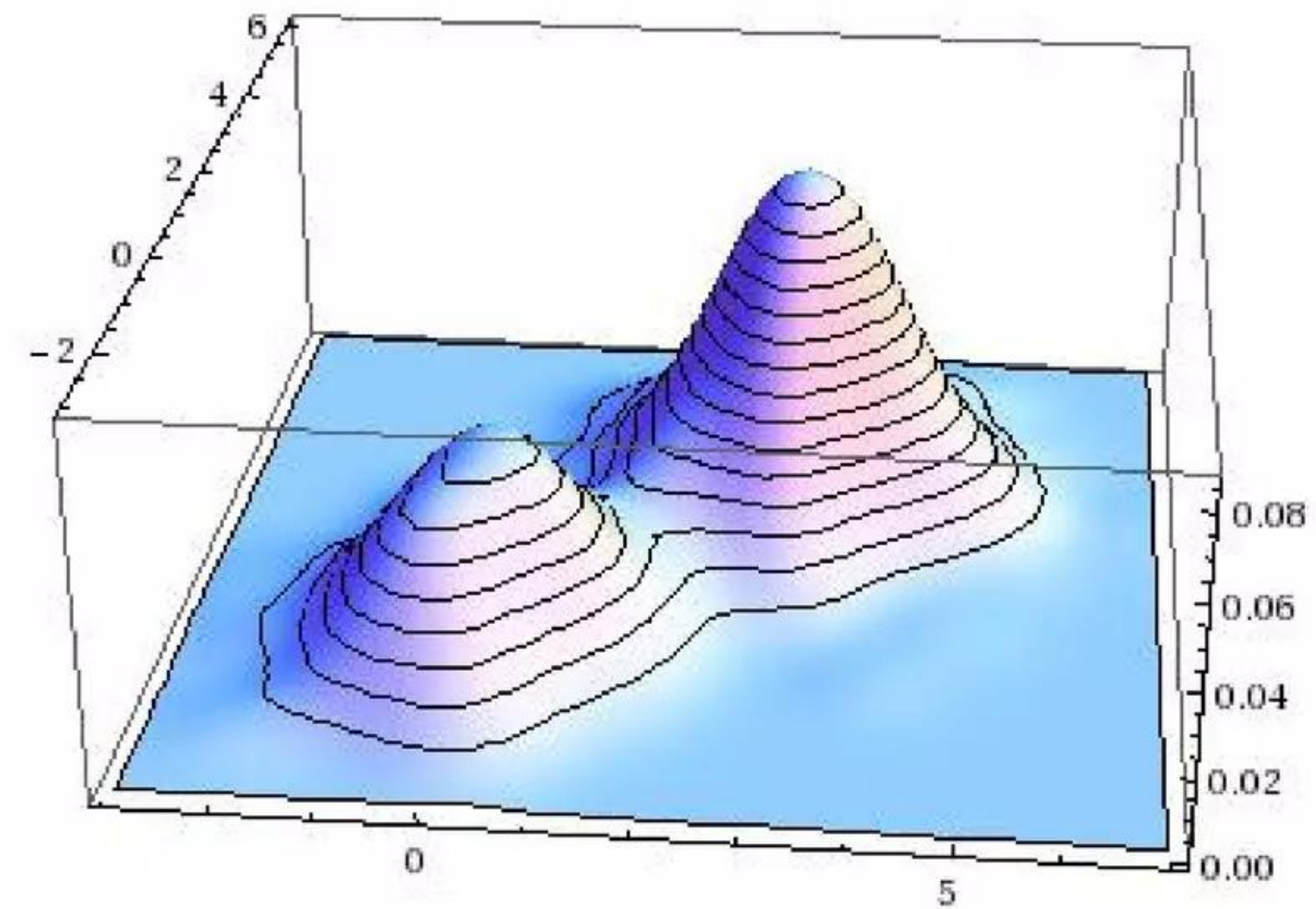
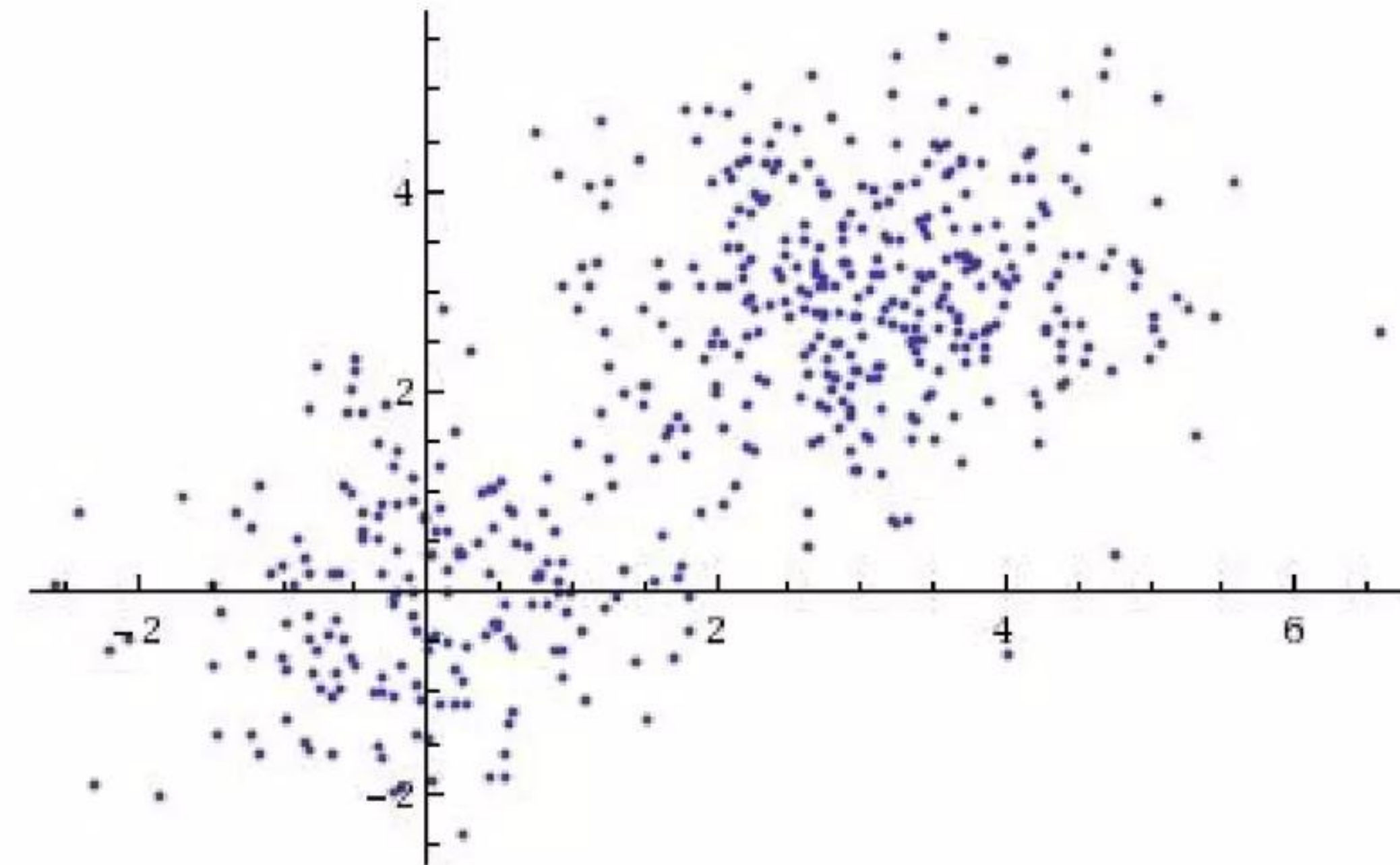
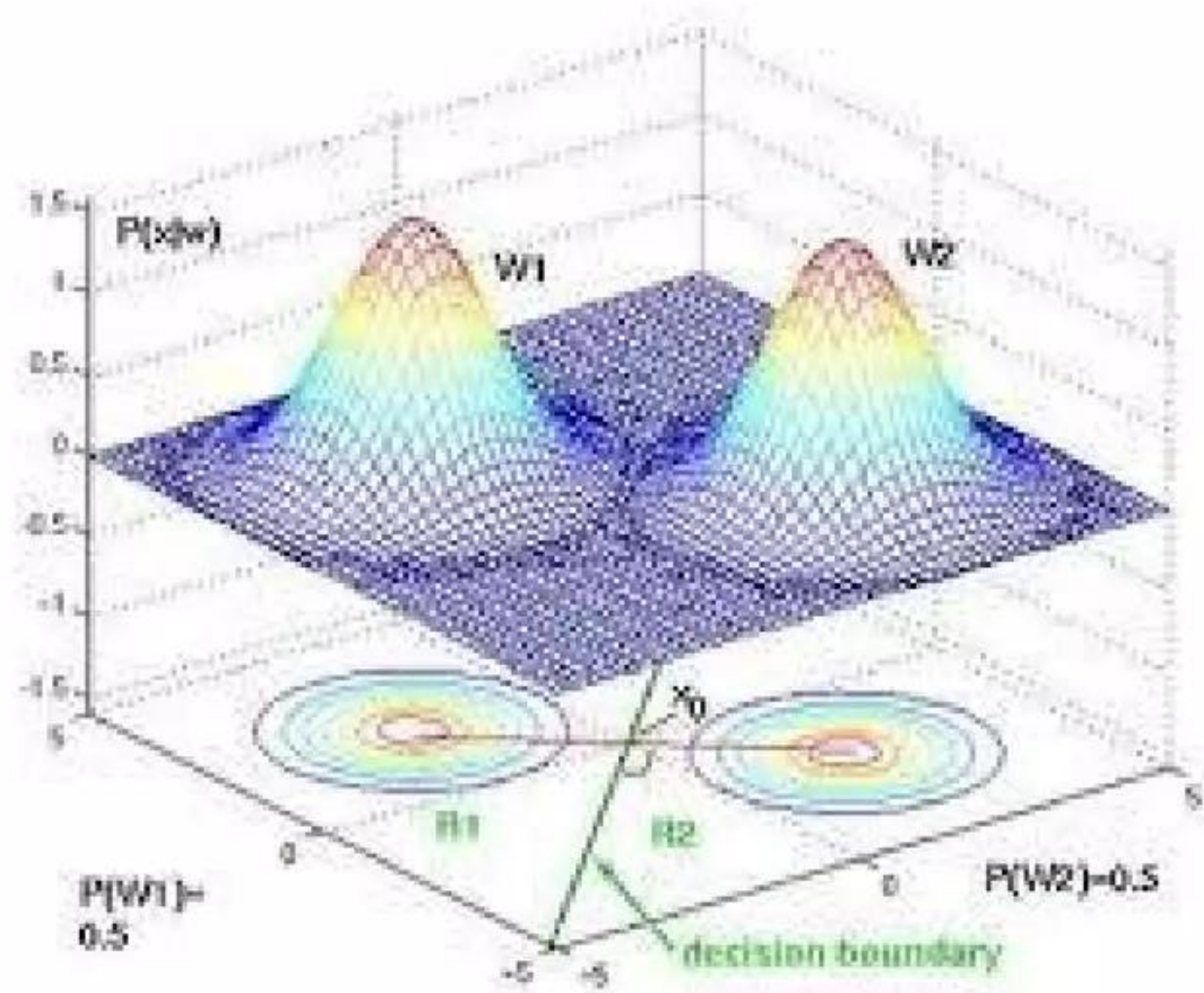
We have 4 gaussians mean 4 clusters on the left picture.



We have 3 gaussians mean 3 clusters on the right picture.



Gaussian Mixture (Multivariate)



We have more than one gaussian mixture mean we have more than one possible position for each data that we want to distribute to GMM.

for example we have data x then want to trying to distribute x to GMM, then we need to calculate the probability of x in first gaussian, then in second gaussian, until our last gaussian. It mean we have $p(x)$ given each gaussians parameters.

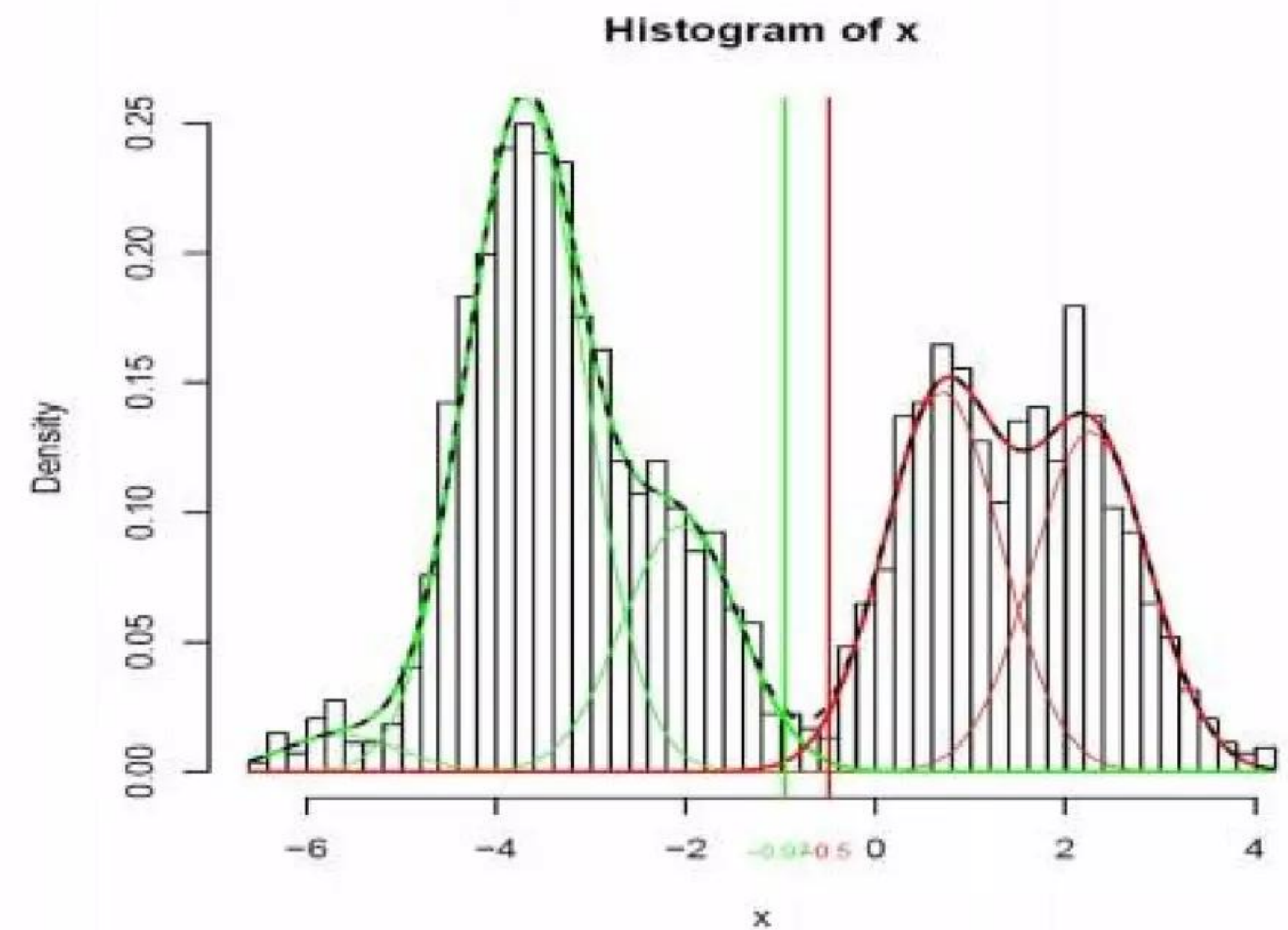
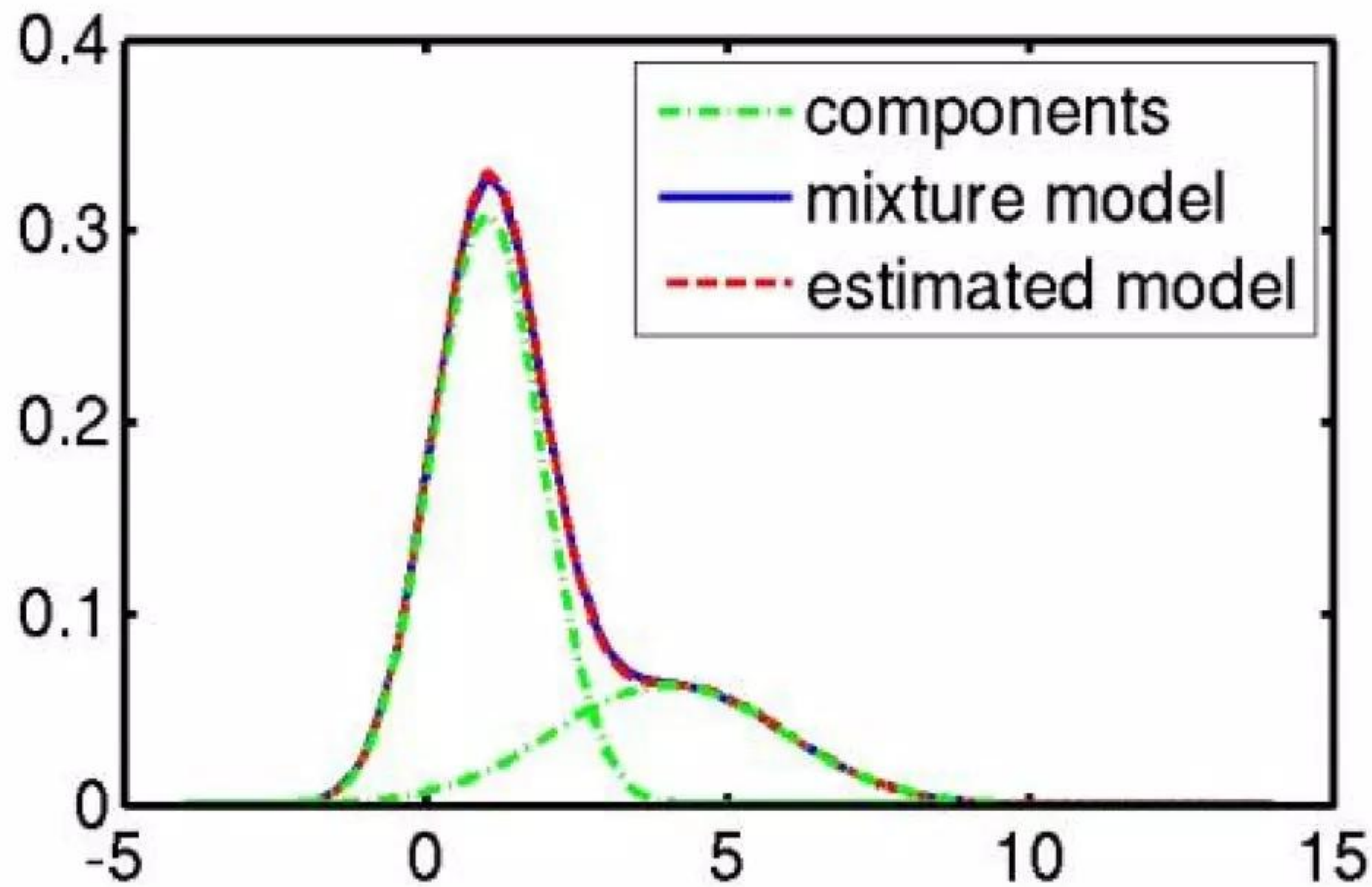
$$f(x) = \operatorname{argmax}\{p(x | \mu_1, \Sigma_1), p(x | \mu_2, \Sigma_2), p(x | \mu_3, \Sigma_3), \dots, p(x | \mu_{n+1}, \Sigma_{n+1})\}$$

where p is :

$$p(x | \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} \|\Sigma\|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

EM (Expectation Maximization) Algorithm

- *GMM Ref : <https://brilliant.org/wiki/gaussian-mixture-model/>*



EM (Expectation Maximization) Algorithm

- Expectation : $w_j^{(i)} := p(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma)$

$$p(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma) = \frac{p(x^{(i)} | z^{(i)} = j; \mu, \Sigma) p(z^{(i)} = j; \phi)}{\sum_{l=1}^k p(x^{(i)} | z^{(i)} = l; \mu, \Sigma) p(z^{(i)} = l; \phi)}$$

- Maximization :

$$\phi_j := \frac{1}{m} \sum_{i=1}^m w_j^{(i)},$$

$$\mu_j := \frac{\sum_{i=1}^m w_j^{(i)} x^{(i)}}{\sum_{i=1}^m w_j^{(i)}},$$

$$\Sigma_j := \frac{\sum_{i=1}^m w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^m w_j^{(i)}}$$

*Log likelihood :

$$\ell(\phi, \mu, \Sigma) = \sum_{i=1}^m \log p(x^{(i)} | z^{(i)}; \mu, \Sigma) + \log p(z^{(i)}; \phi).$$

, Dimana $P(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma)$.

$$= \sum_{i=1}^m \sum_{j=1}^k w_j^{(i)} \log \frac{\frac{1}{(2\pi)^{n/2} |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)} - \mu_j)^T \Sigma_j^{-1} (x^{(i)} - \mu_j)\right) \cdot \phi_j}{w_j^{(i)}}$$

$\theta_{i=1..K}$

= parameter of distribution of observation associated with component i

$\phi_{i=1..K}$

= mixture weight, i.e., prior probability of a particular component i

ϕ

= K -dimensional vector composed of all the individual $\phi_{1..K}$; must sum to 1

$z_{i=1..N}$

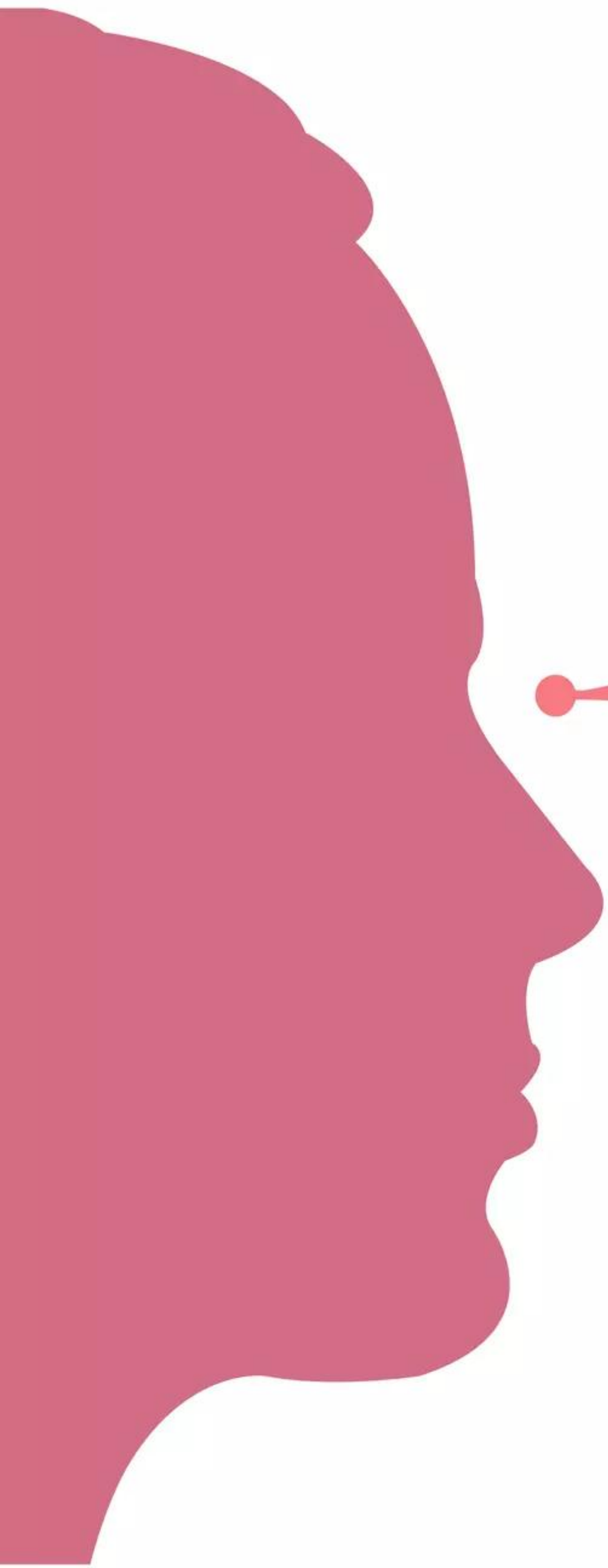
= component of observation i

$x_{i=1..N}$

= observation i

Fraud Detection





Payment Fraud (phishing, account take-over, carding)

System abuse (promo, content, account, logistic and payment methods especially **COD**)

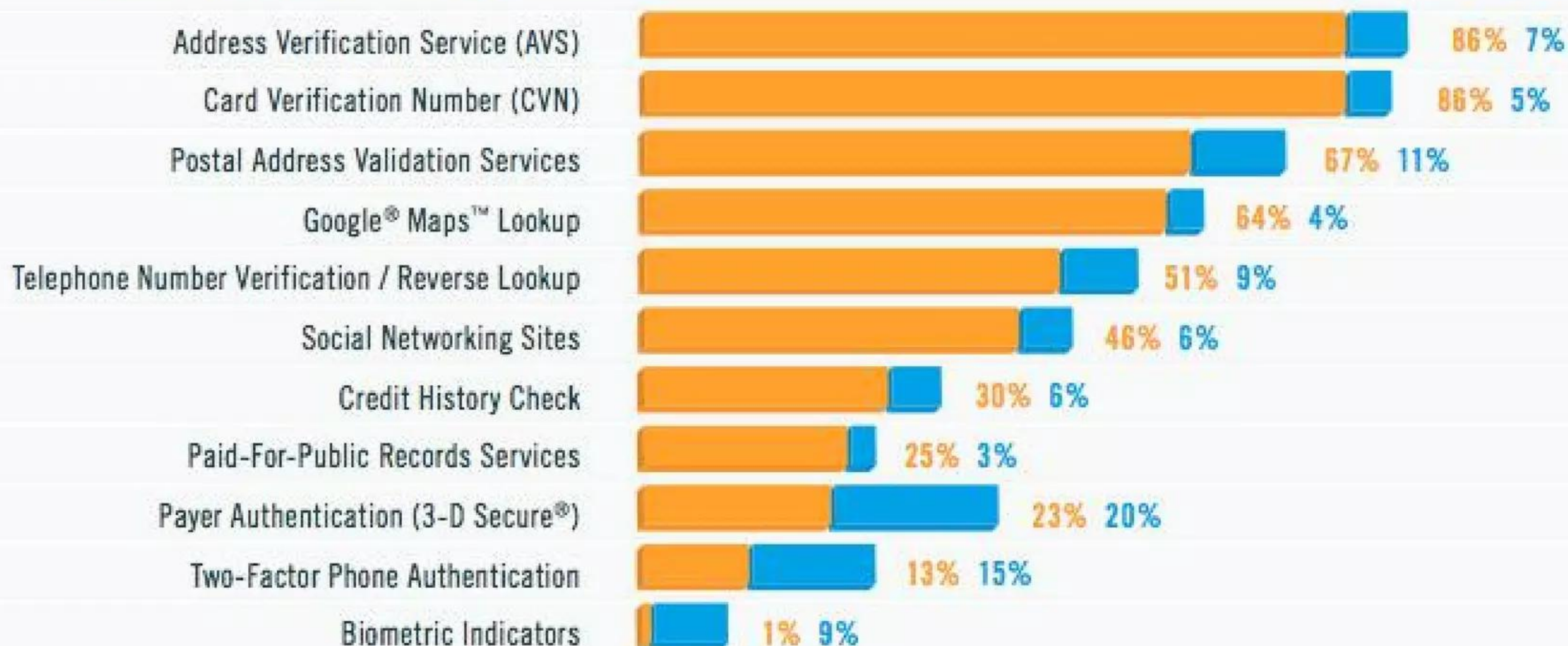
Fraud not only result in financial losses but also produce some reputational risk.

Some security measures has been taken by bank or another multinational finance service.

[E. Duman et al, 2013]

MOST ADOPTED FRAUD DETECTION TOOLS

VALIDATION SERVICES



CURRENTLY USING



PLANNING NEW IMPLEMENTATION

MOST ADOPTED FRAUD DETECTION TOOLS

YOUR PROPRIETARY DATA / CUSTOMER HISTORY



MULTI-MERCHANT DATA / PURCHASE HISTORY



PURCHASE DEVICE TRACKING



CURRENTLY USING  PLANNING NEW IMPLEMENTATION 



THANK YOU

Any question?