Disclaimer

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.

blibli.com
BIG CHOICES BIG DEALS

Machine Learning:
an Introduction and cases

Hendri Karisma
hendri.karisma@gdn-commerce.com

Hendri Karisma

- 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

Materials

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

The Definition of Informatics

"*Automation of Information*" – Prof. Dr. Ing. Iping Supriana

Artificial Intelligence

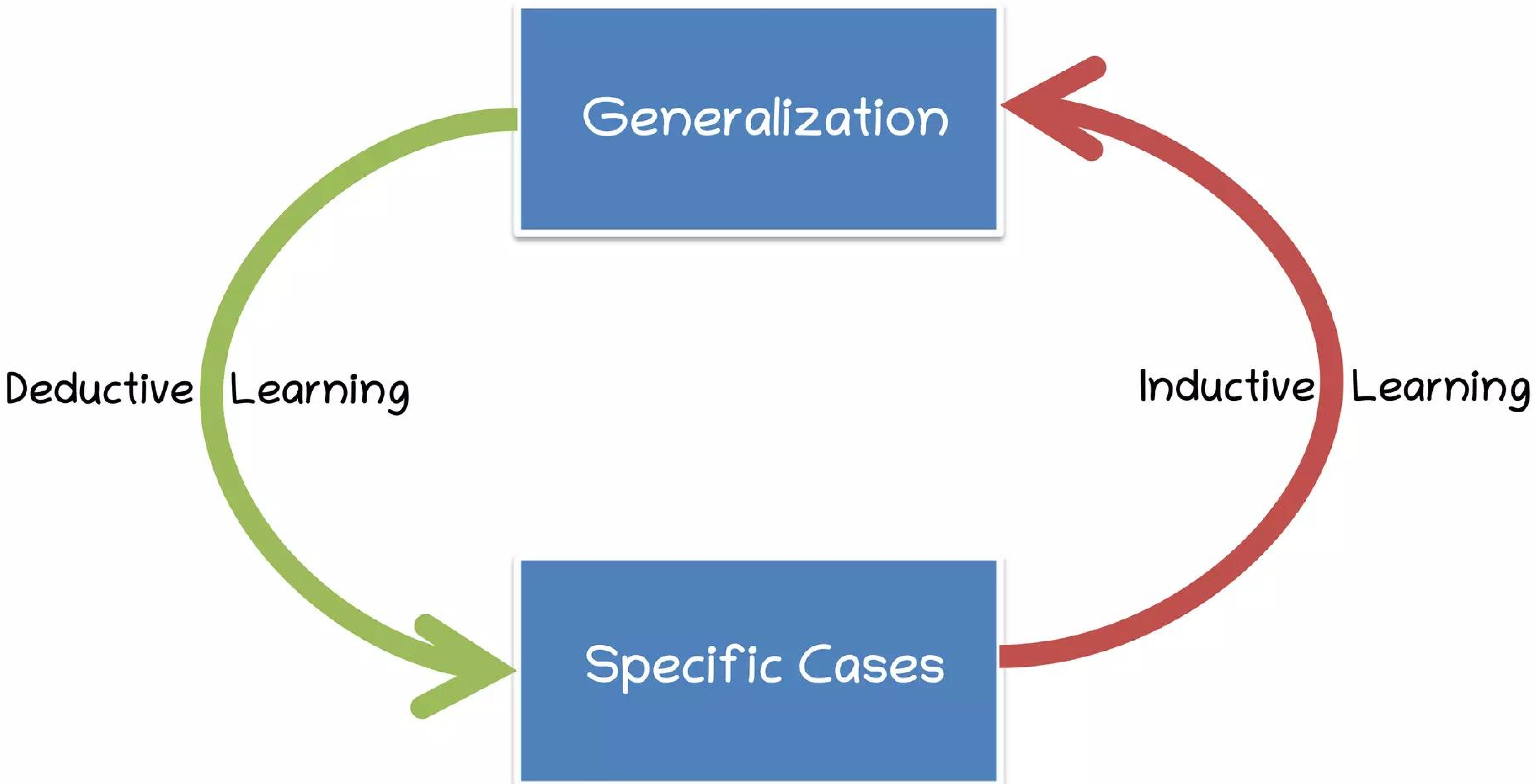
- S. Rusel and P. Norvig, Artificial Intelligence in Modern Approach

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

Problem Solving agent

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

Learning


Generalization
Specific Cases
Deductive Learning
Inductive Learning

Data ???

Machine Learning Definition

"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

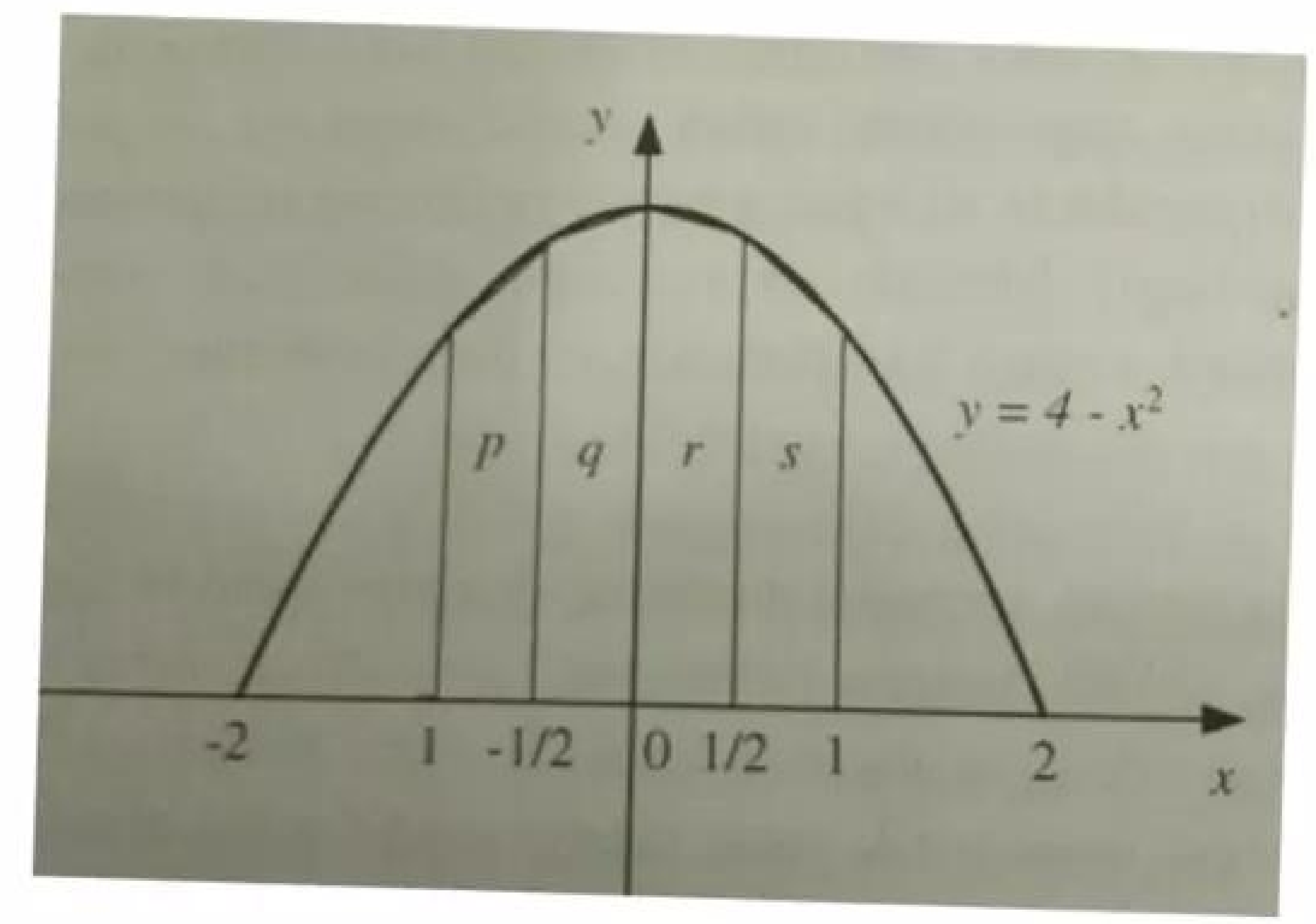
Problem and Solutions

- 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

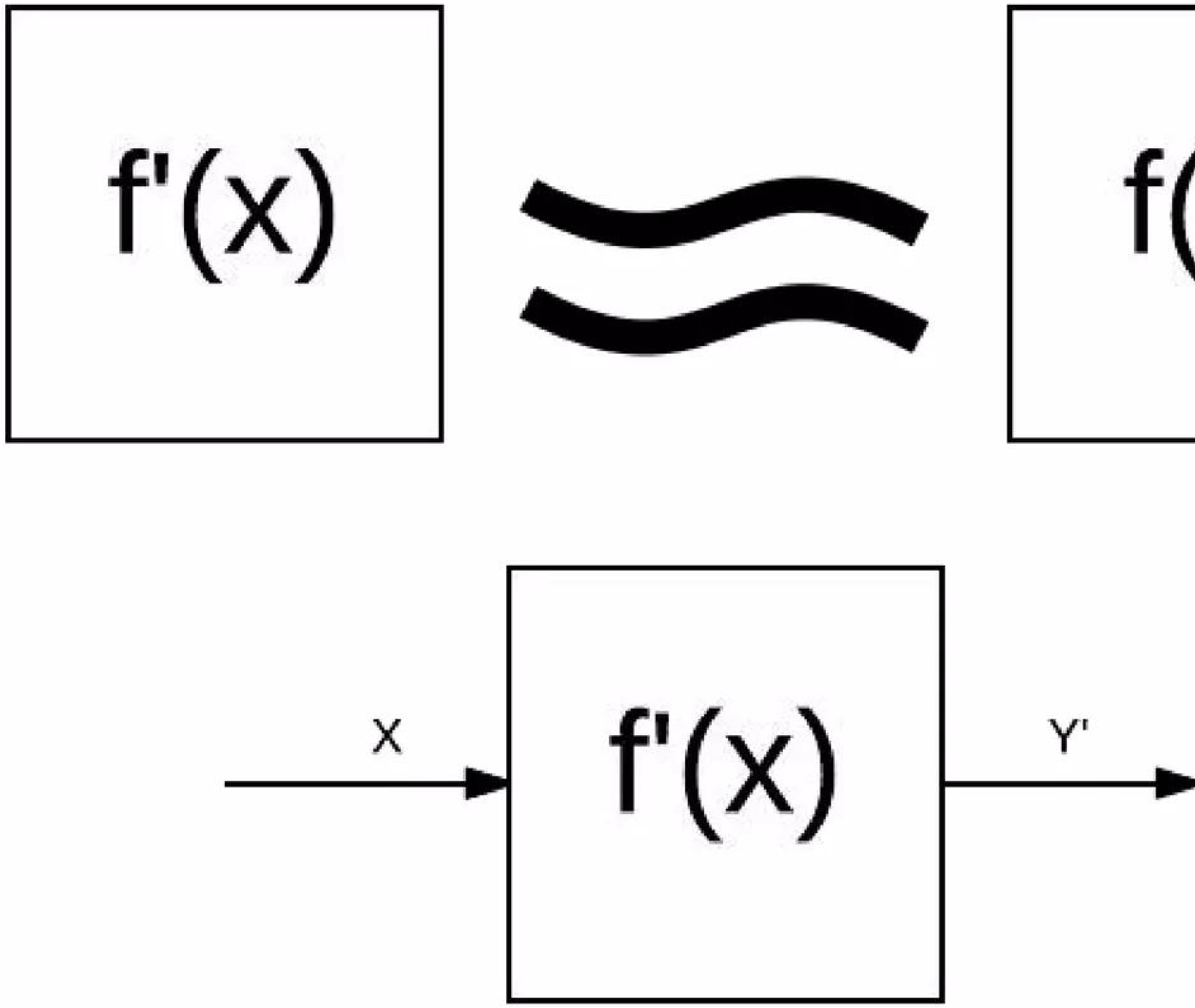
$$
\begin{aligned}
I &\approx p + q + r + s \\
&\approx \{[f(-1) + f(-1/2)] \times 0.5/2\} + \{[f(-1/2) + f(0)] \times 0.5/2\} + \\
&\quad \{[f(0) + f(1/2)] \times 0.5/2\} + \{[f(1/2) + f(1)] \times 0.5/2\} \\
&\approx 0.5/2\ \{f(-1) + 2f(-1/2) + 2f(0) + 2f(1/2) + f(1)\} \\
&\approx 0.5/2\ \{3 + 7.5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



  - 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.

How it works



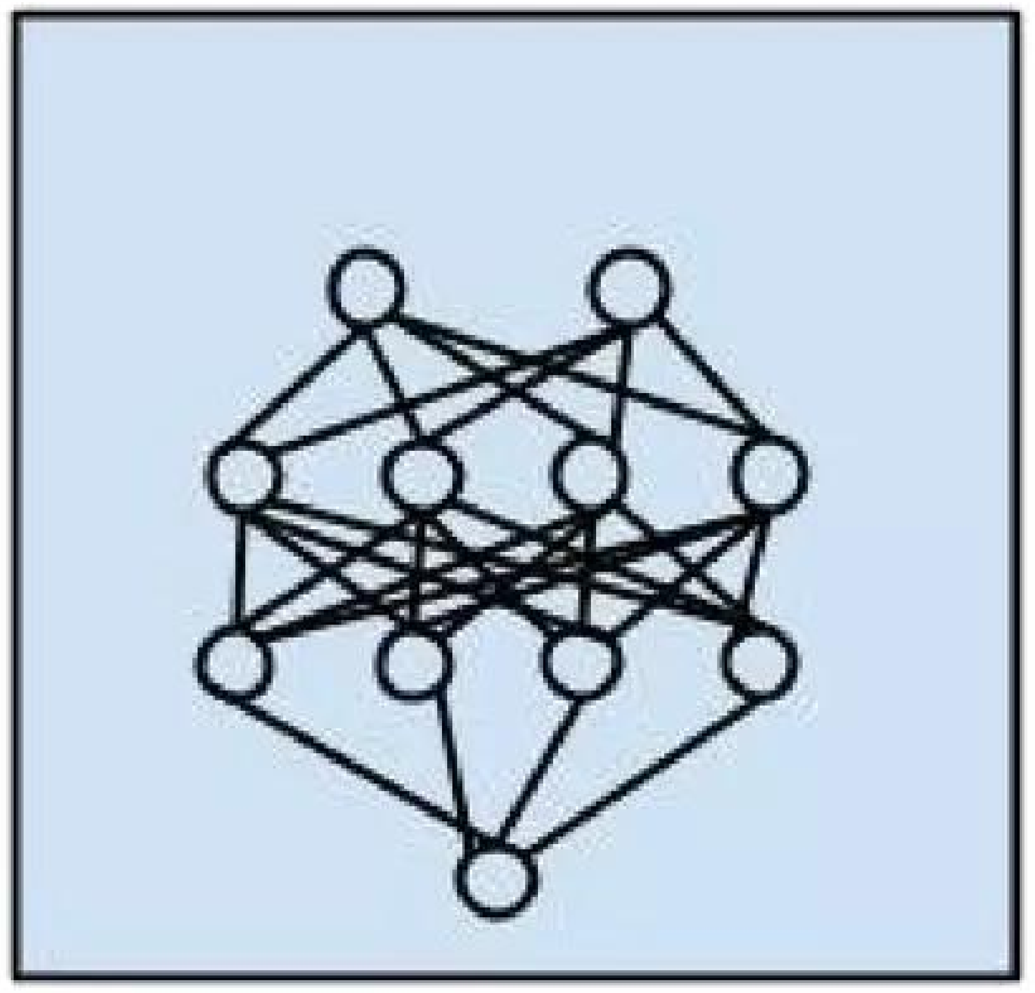
Count the error (y - y')
Then minimize the error
or
Maximize the likelihood

Machine Learning Function

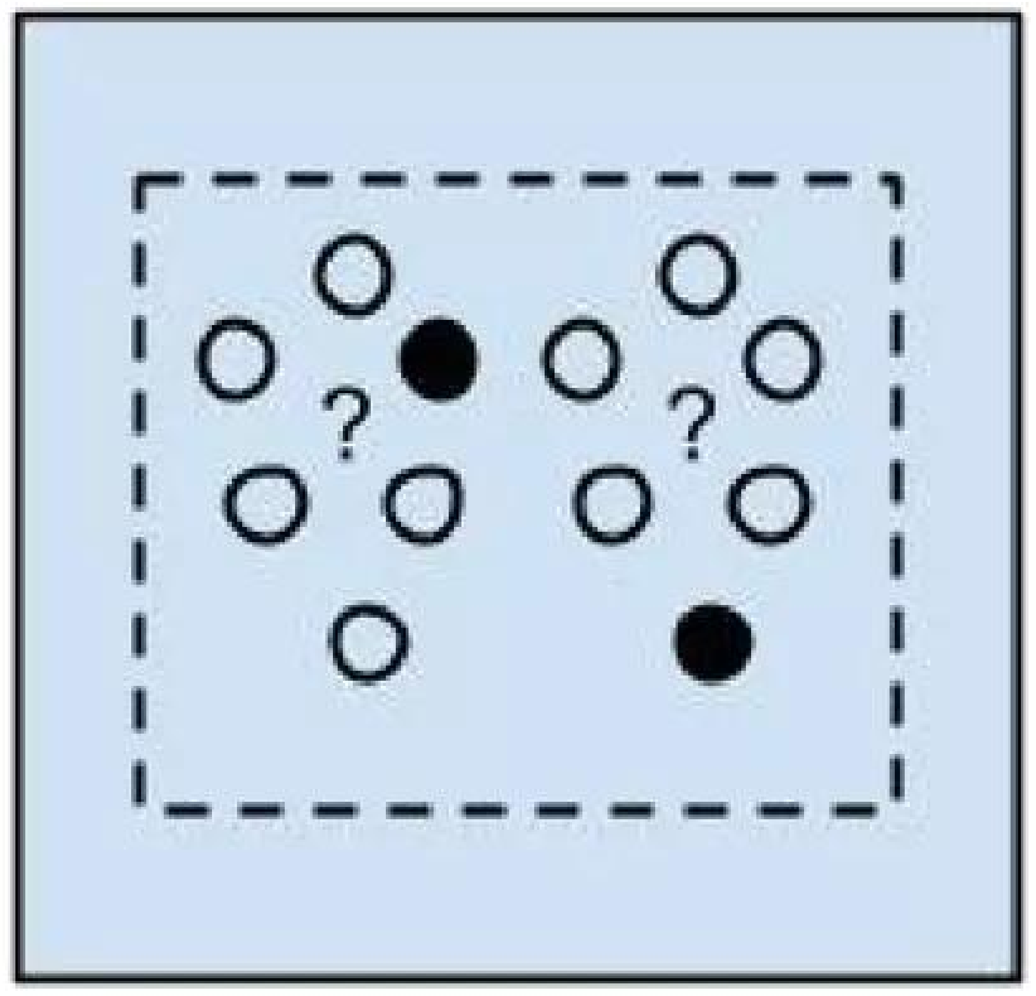
- Information Theory (Decission Tree : ID-Tree, C4.5, etc)
- Probability (Bayessian : Naive Bayes, Belief Network, etc)
- Graphical Model (Belief network, HMM, CRF, Neural Network, etc)
- Numerical Method / Regression (Stochastic Gradient Descent: Linear Regression, Multiple Linear Regression, Neural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

Machiner Learning Taxonomy

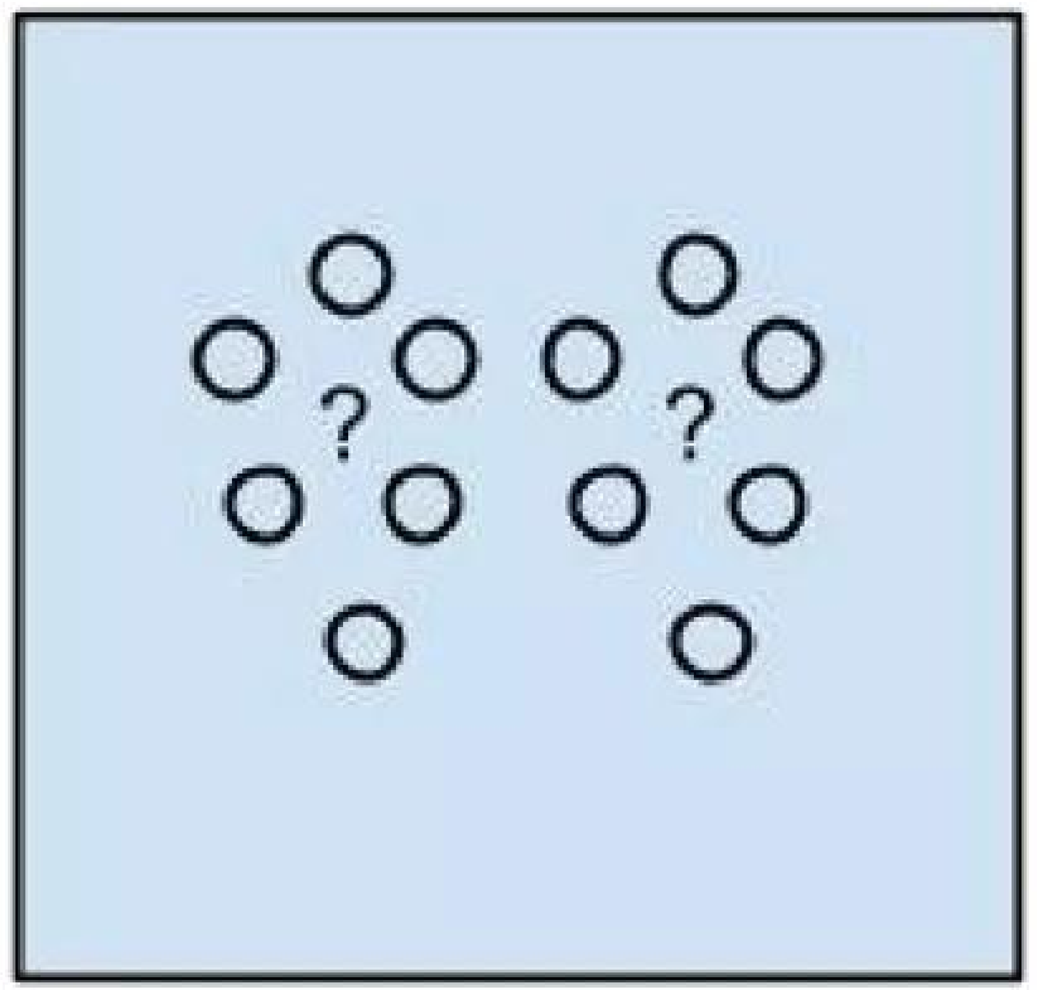
- Supervised
- Unsupervised
- Reinforcement Learning
- Semi-Supervised
- Deep Learning



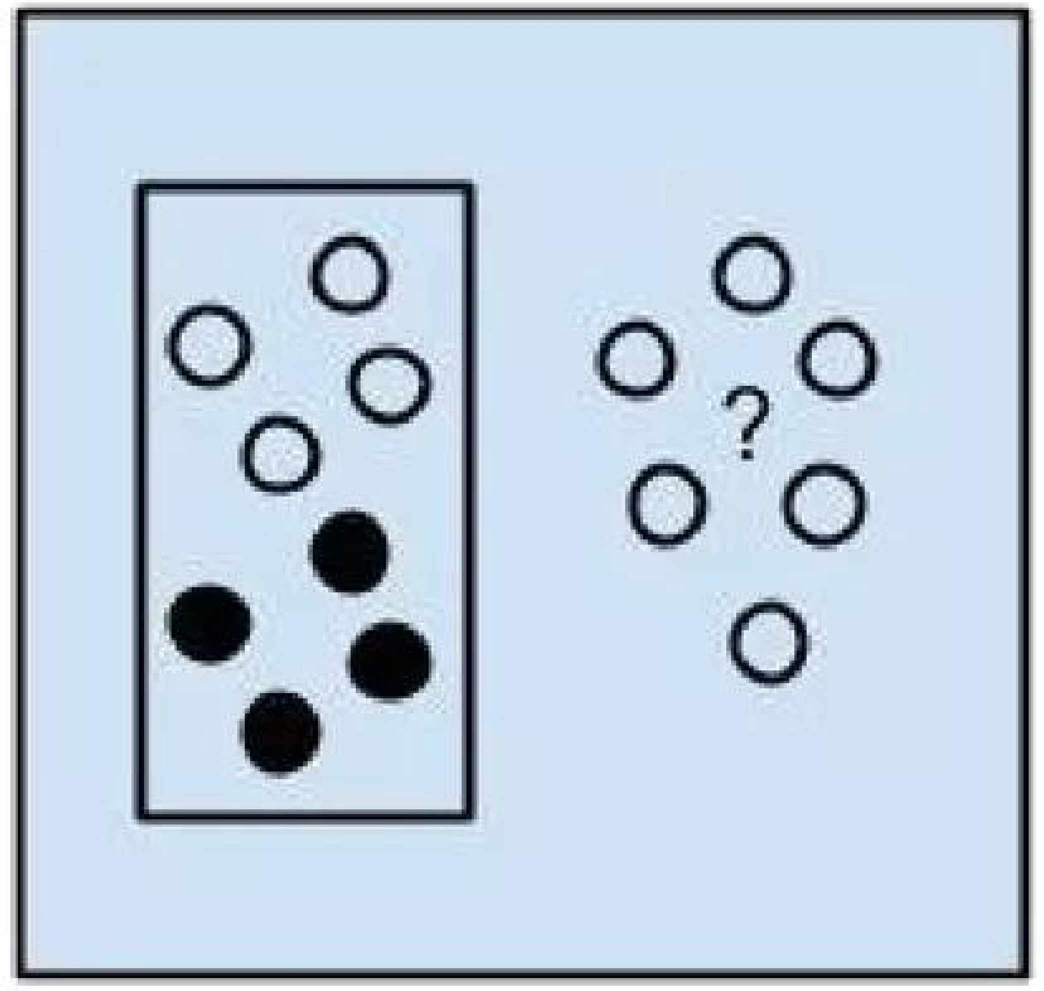
Deep Learning Algorithms



Semi-supervised Learning Algorithms

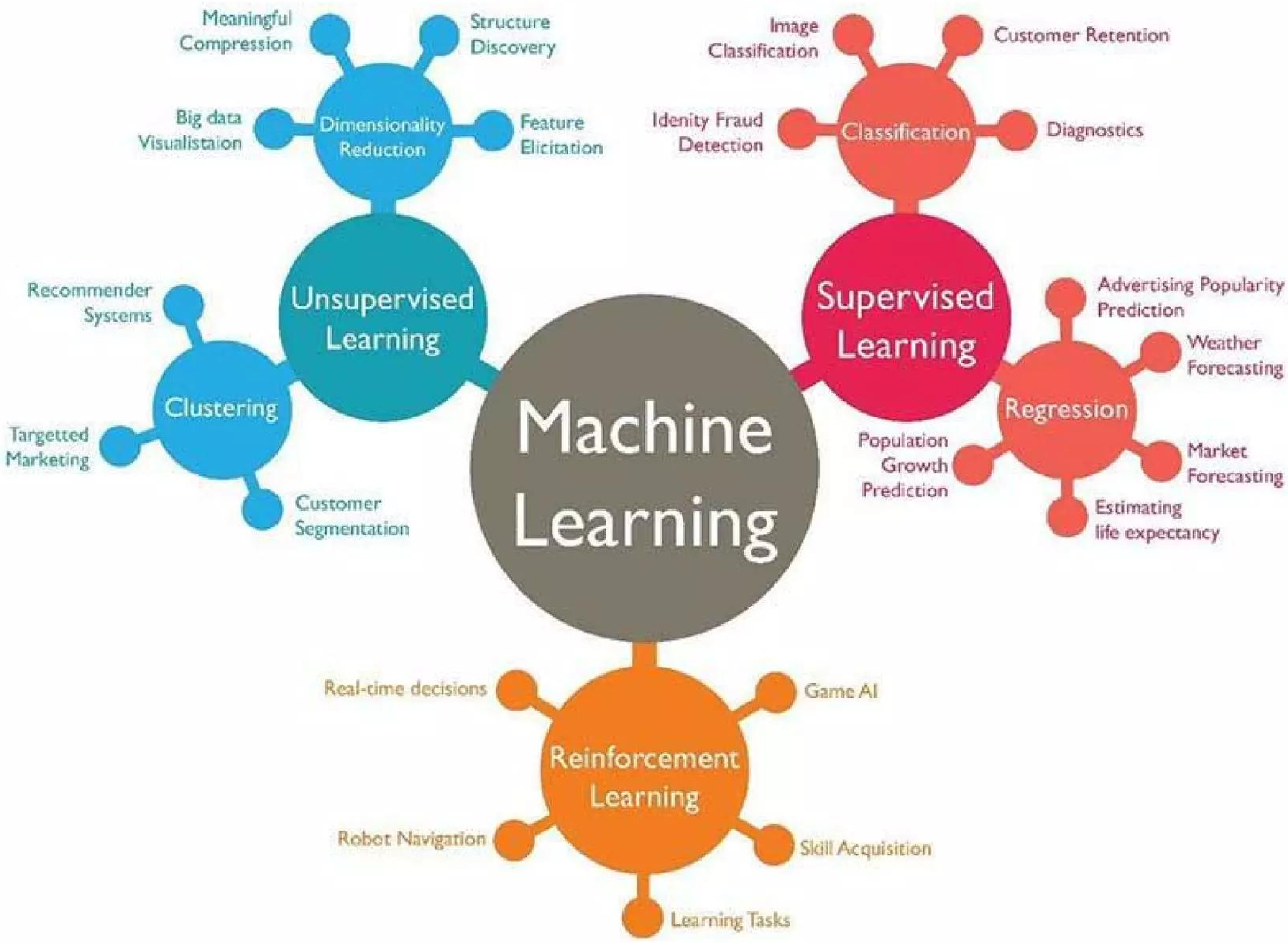


Unsupervised Learning Algorithms

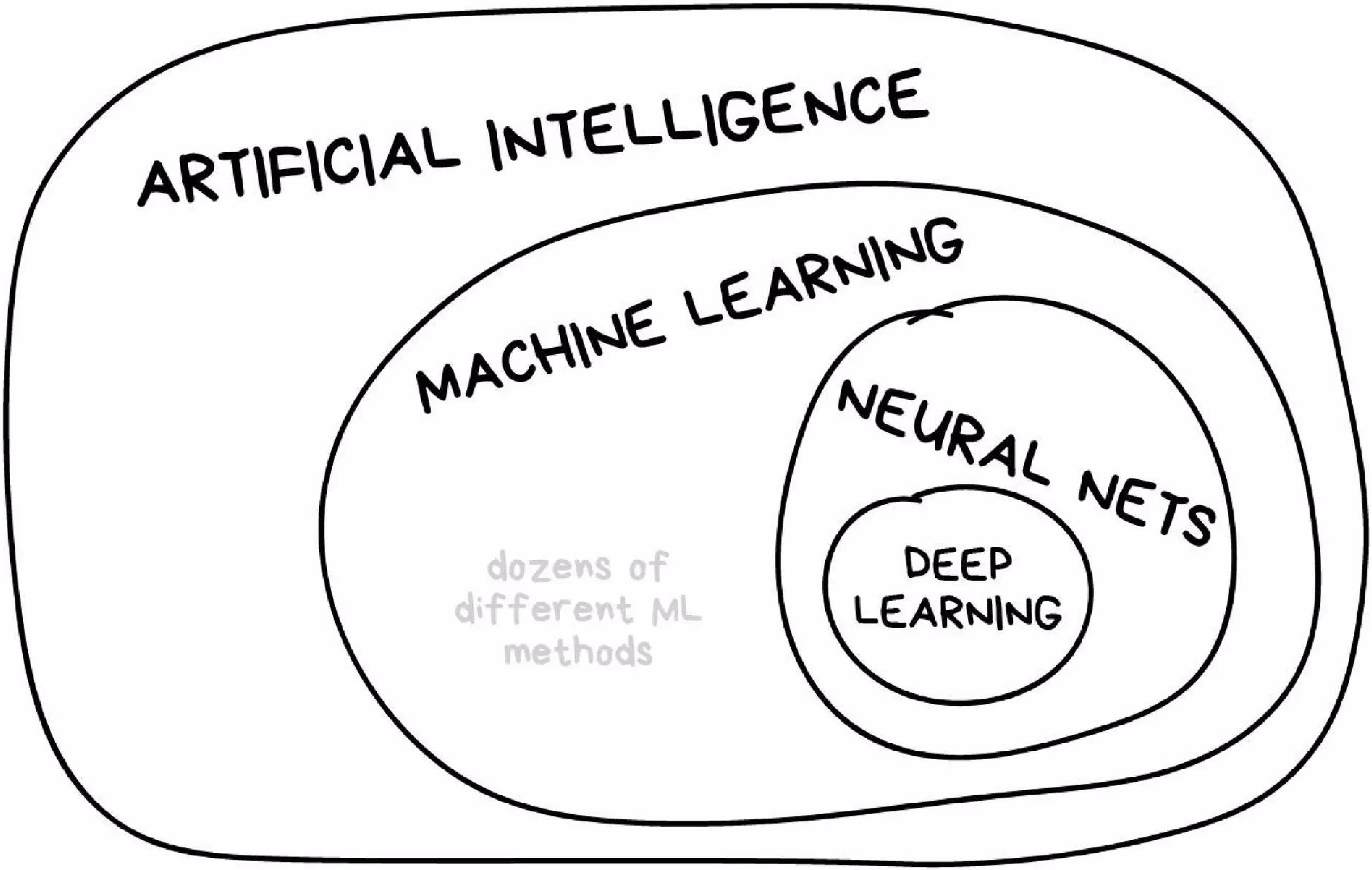


Supervised Learning Algorithms

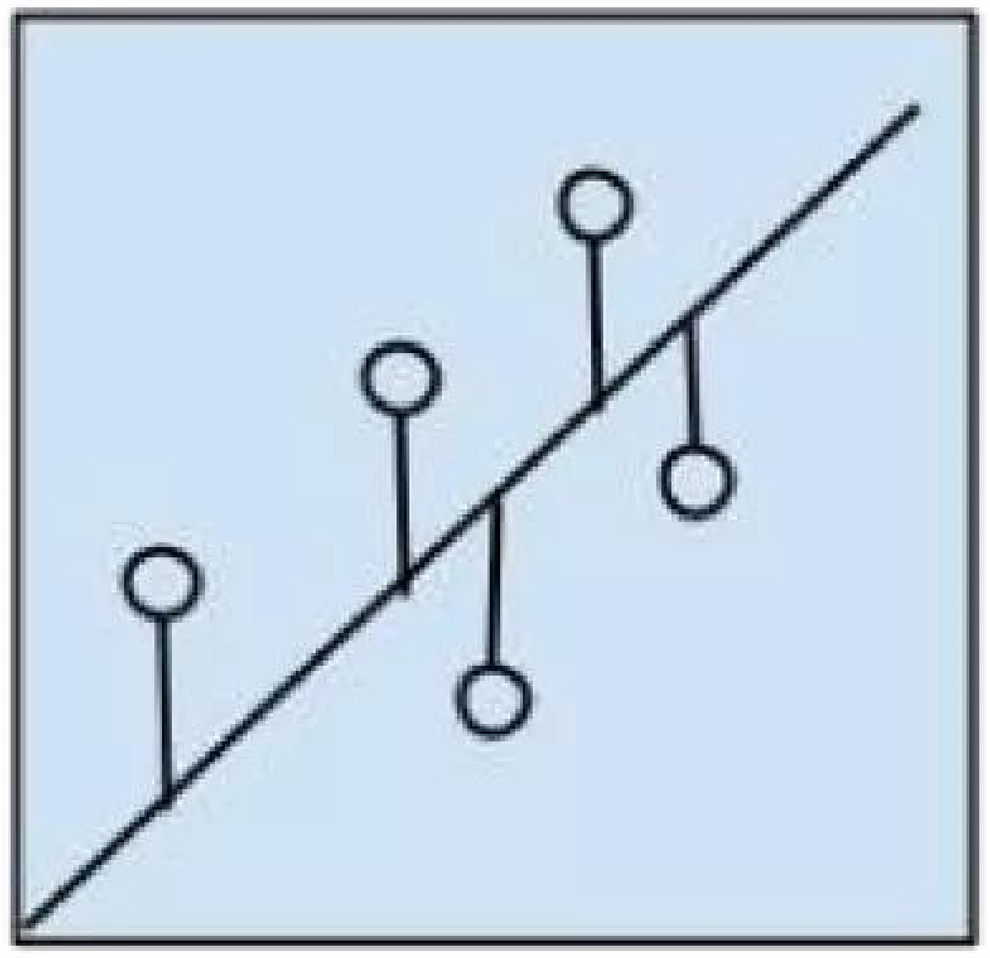
Machine Learning Taxonomy #2



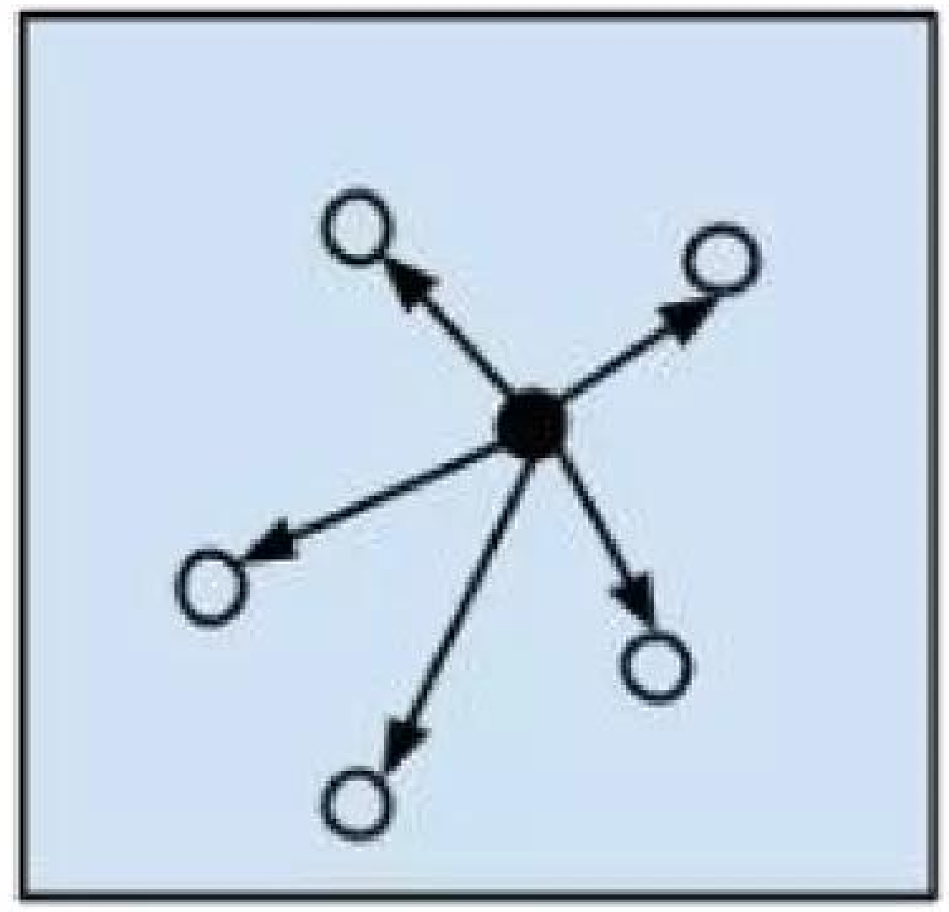
Methods


ARTIFICIAL INTELLIGENCE
MACHINE LEARNING
NEURAL NETS
DEEP LEARNING
dozens of different ML methods

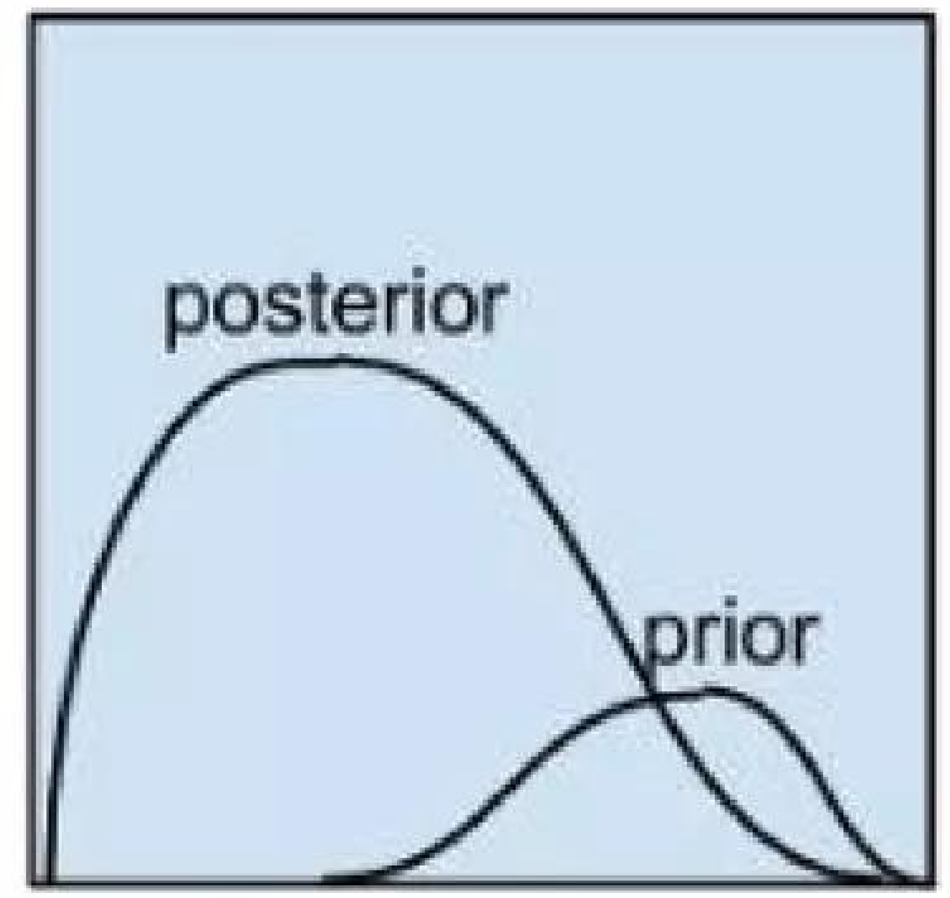
Machiner Learning Taxonomy



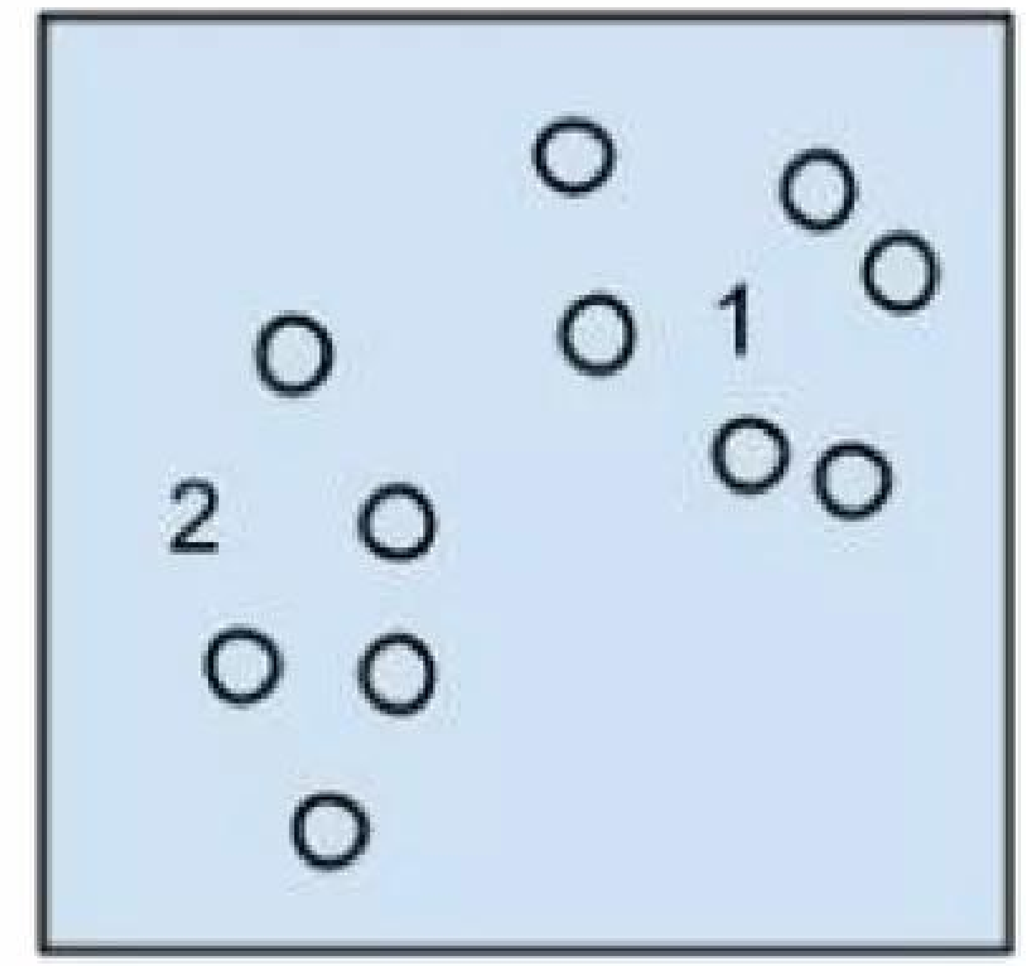
Regression Algorithms



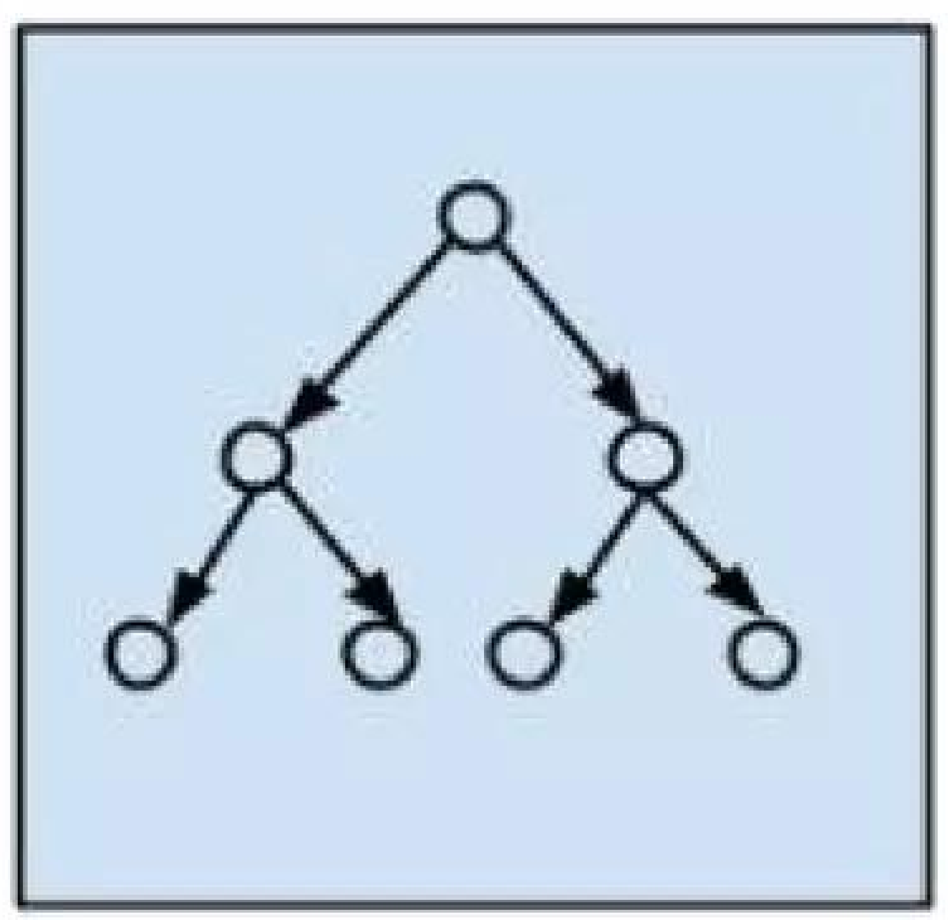
Instance-based Algorithms



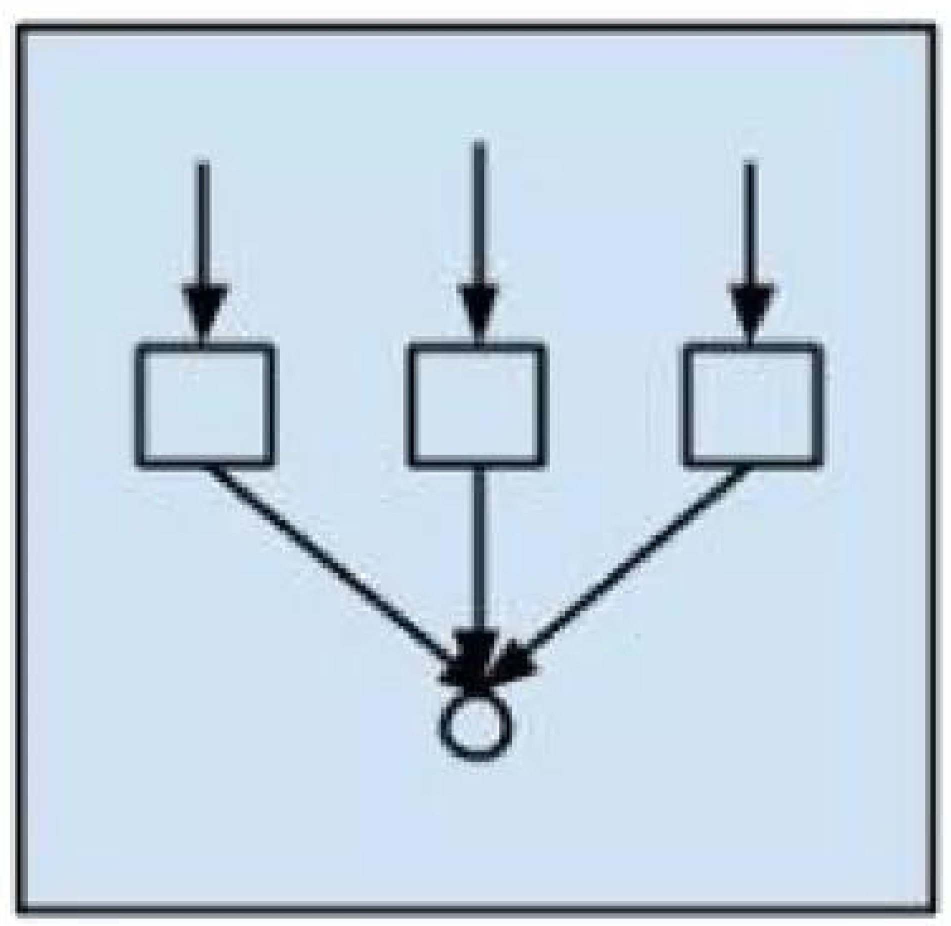
Bayesian Algorithms



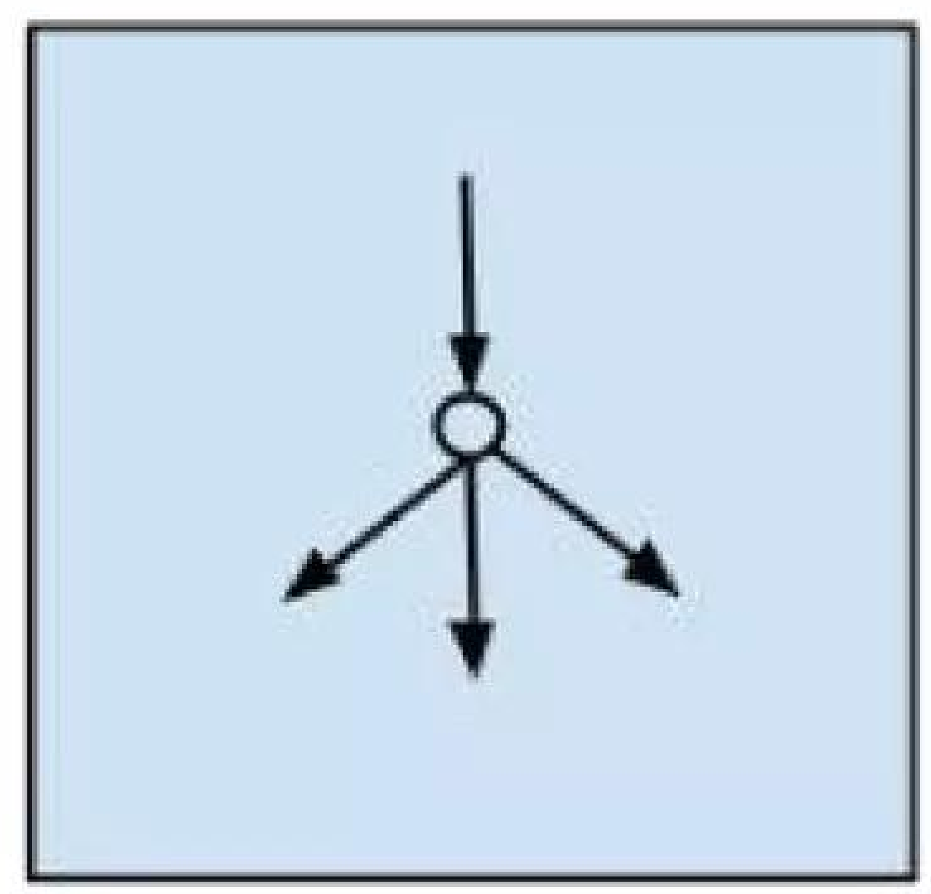
Clustering Algorithms



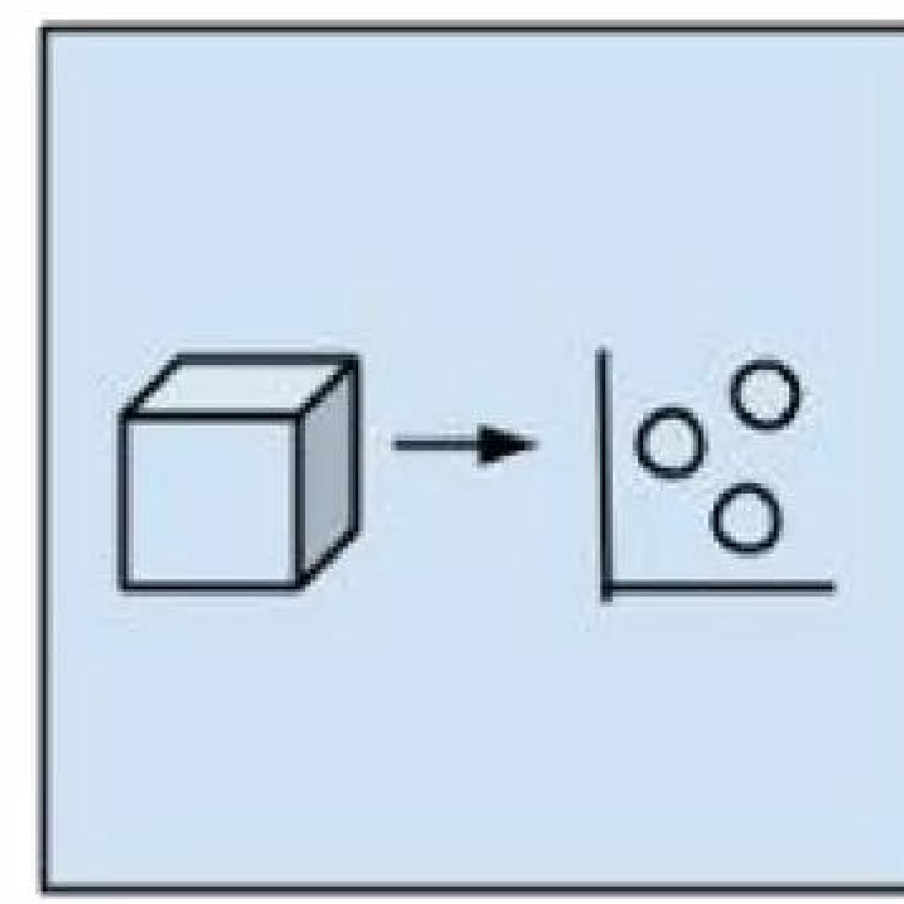
Decision Tree Algorithms



Ensemble Algorithms



Artificial Neural Network Algorithms



Dimensional Reduction Algorithms

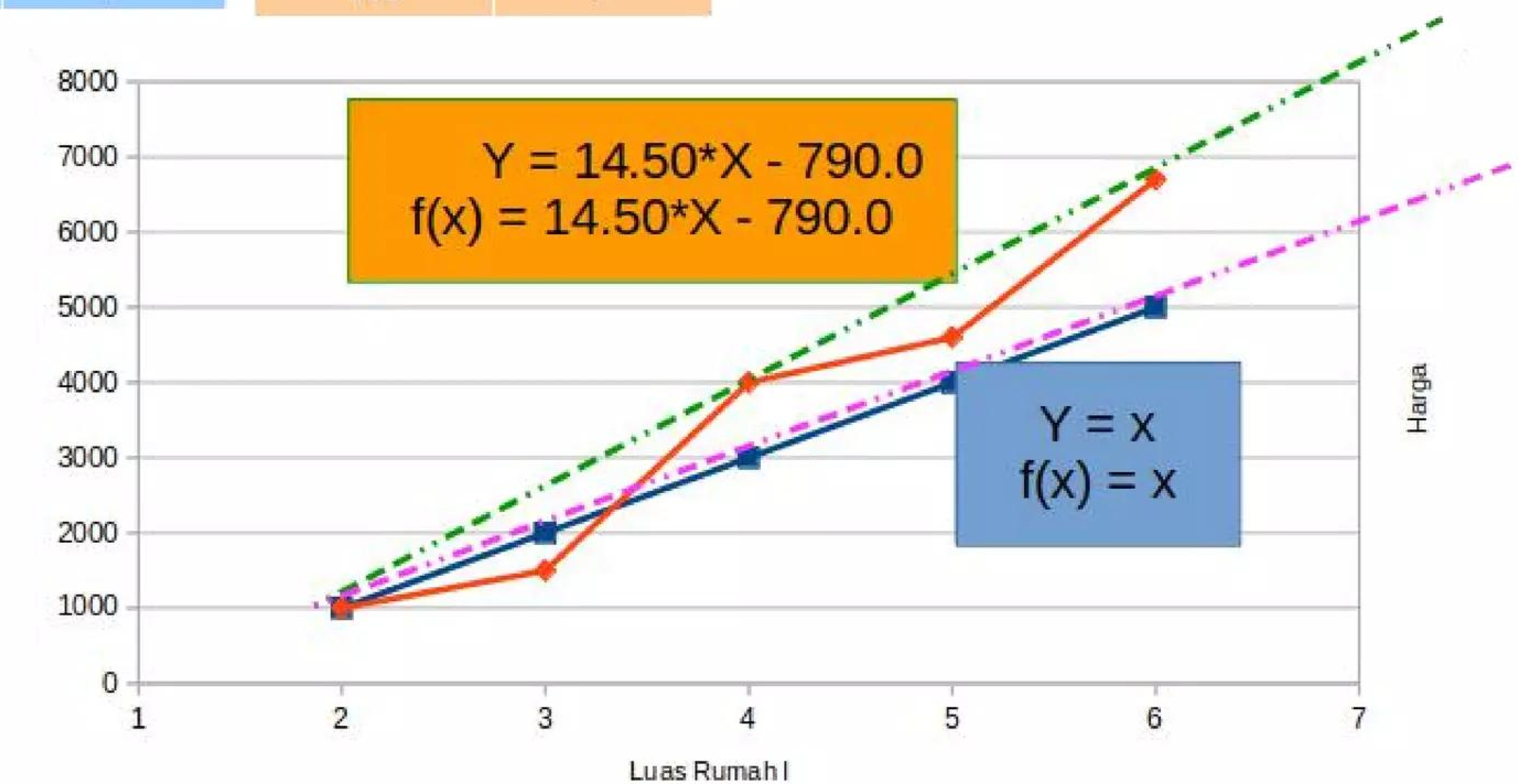
Regression

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

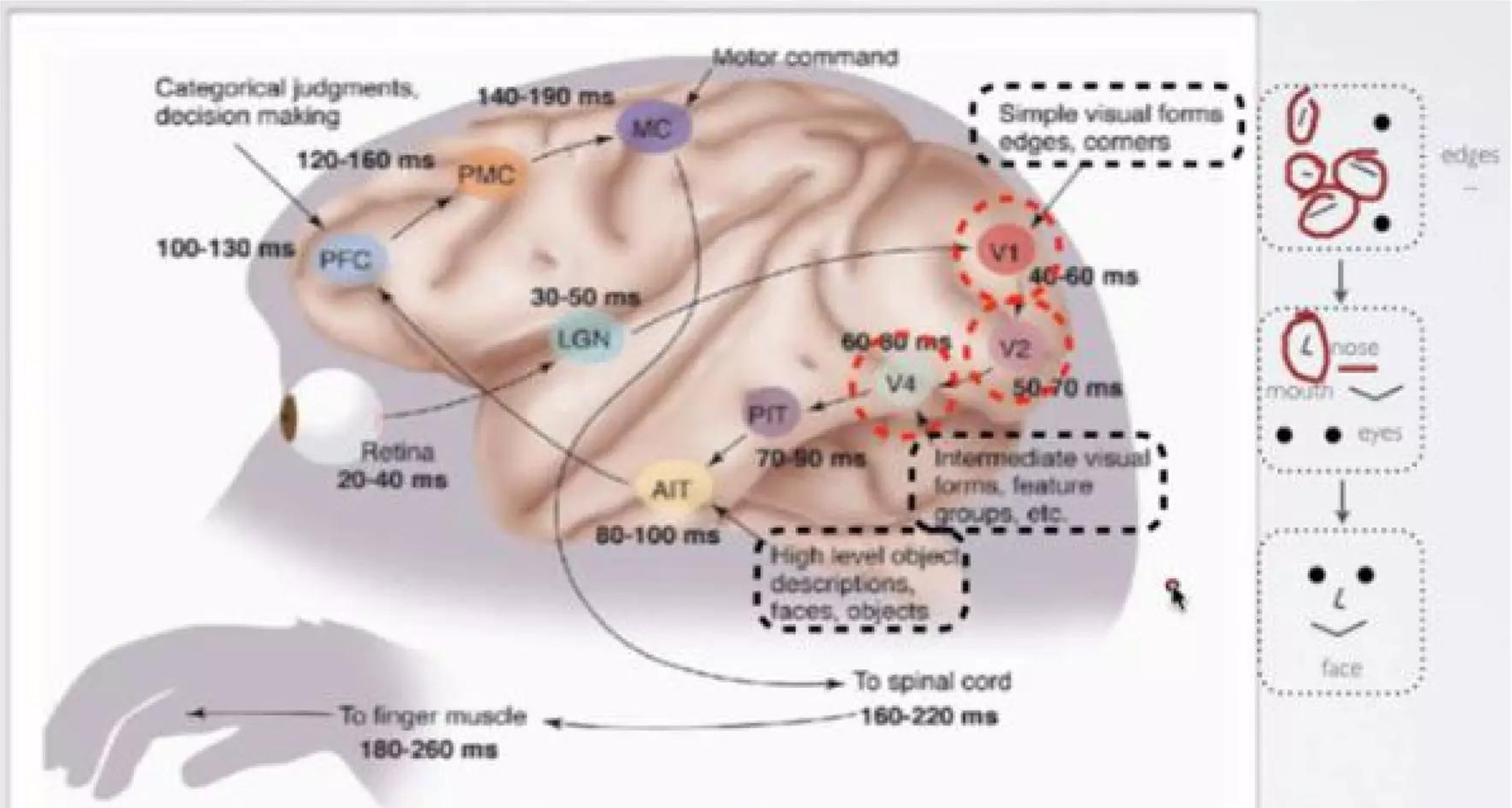
Luas Rumah (m^2)	Harga (Juta)
100	1000
200	1500
300	4000
400	4600
500	6700
600	?

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

Y = 1.500*X - 0.5000

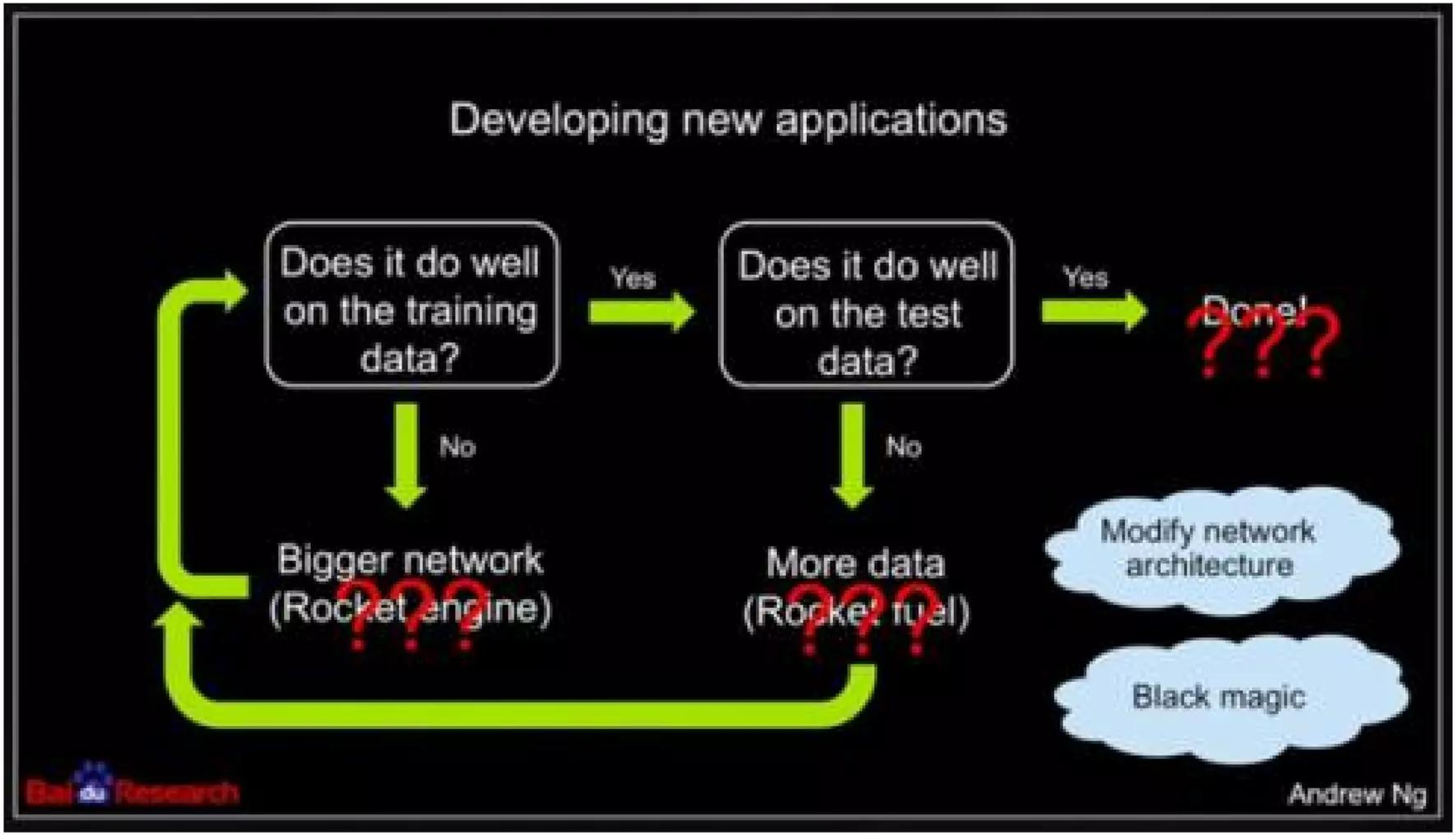


Deep Learning

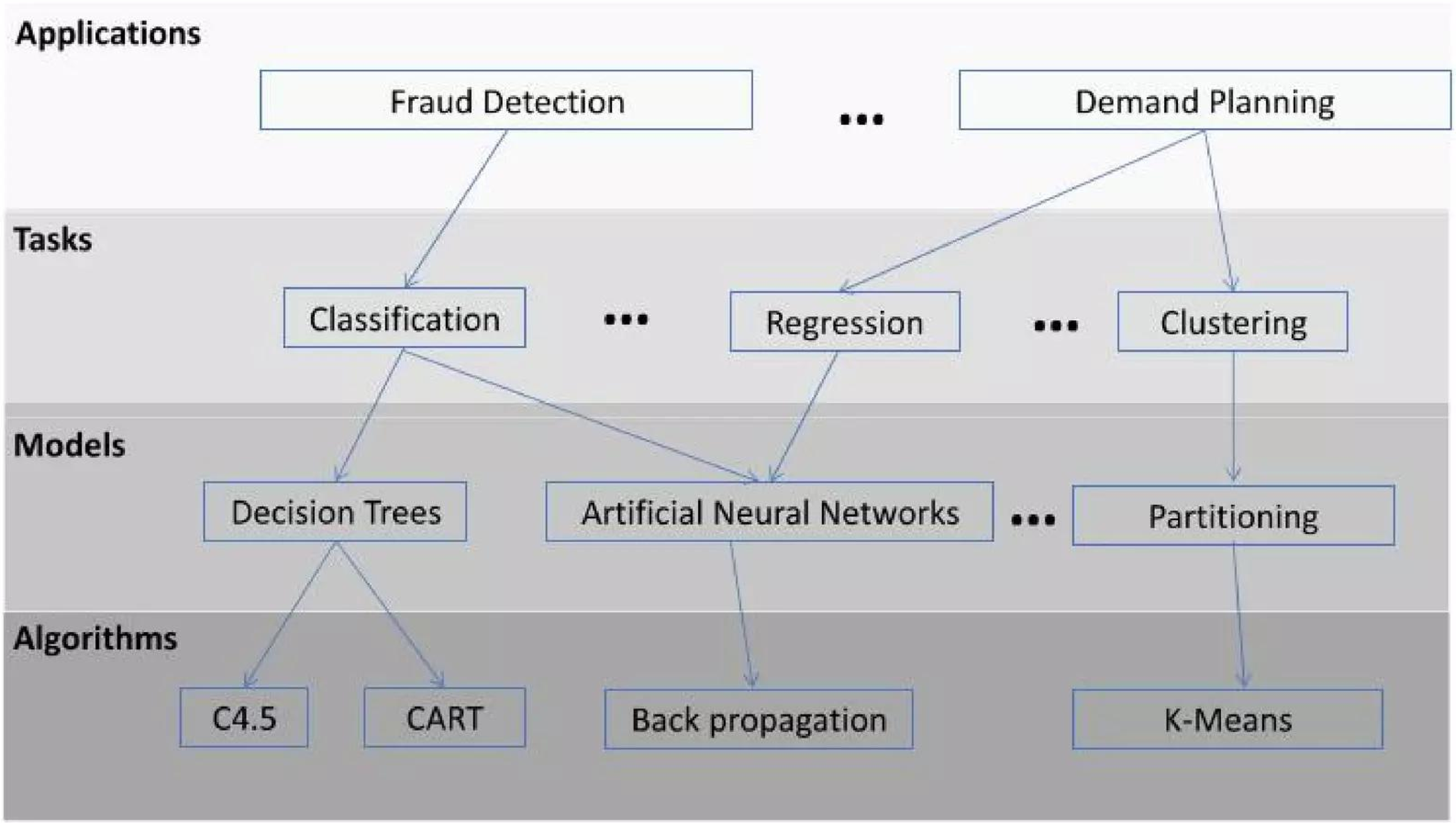


[picture from Simon Thorpe]

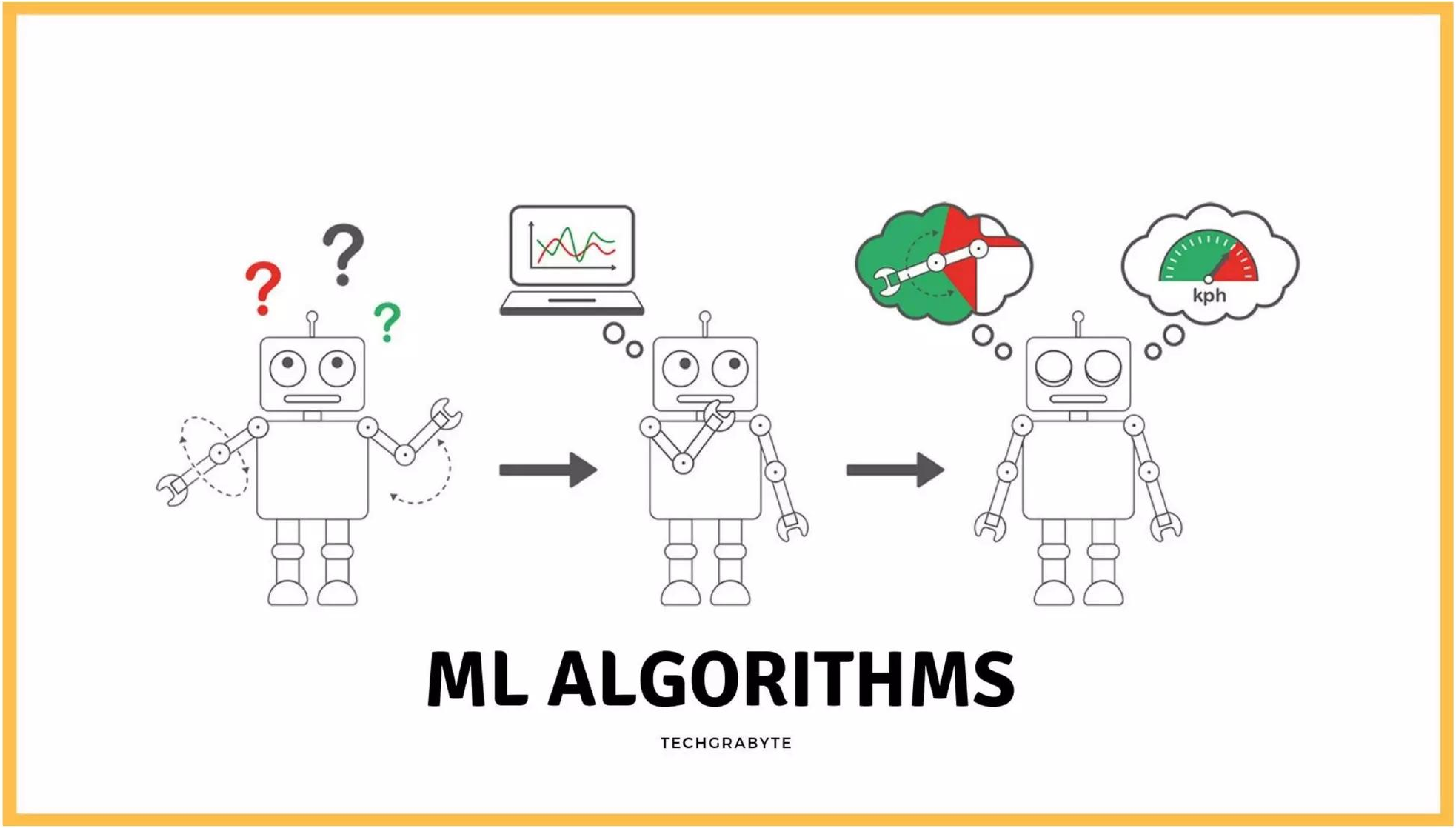
Deep Learning



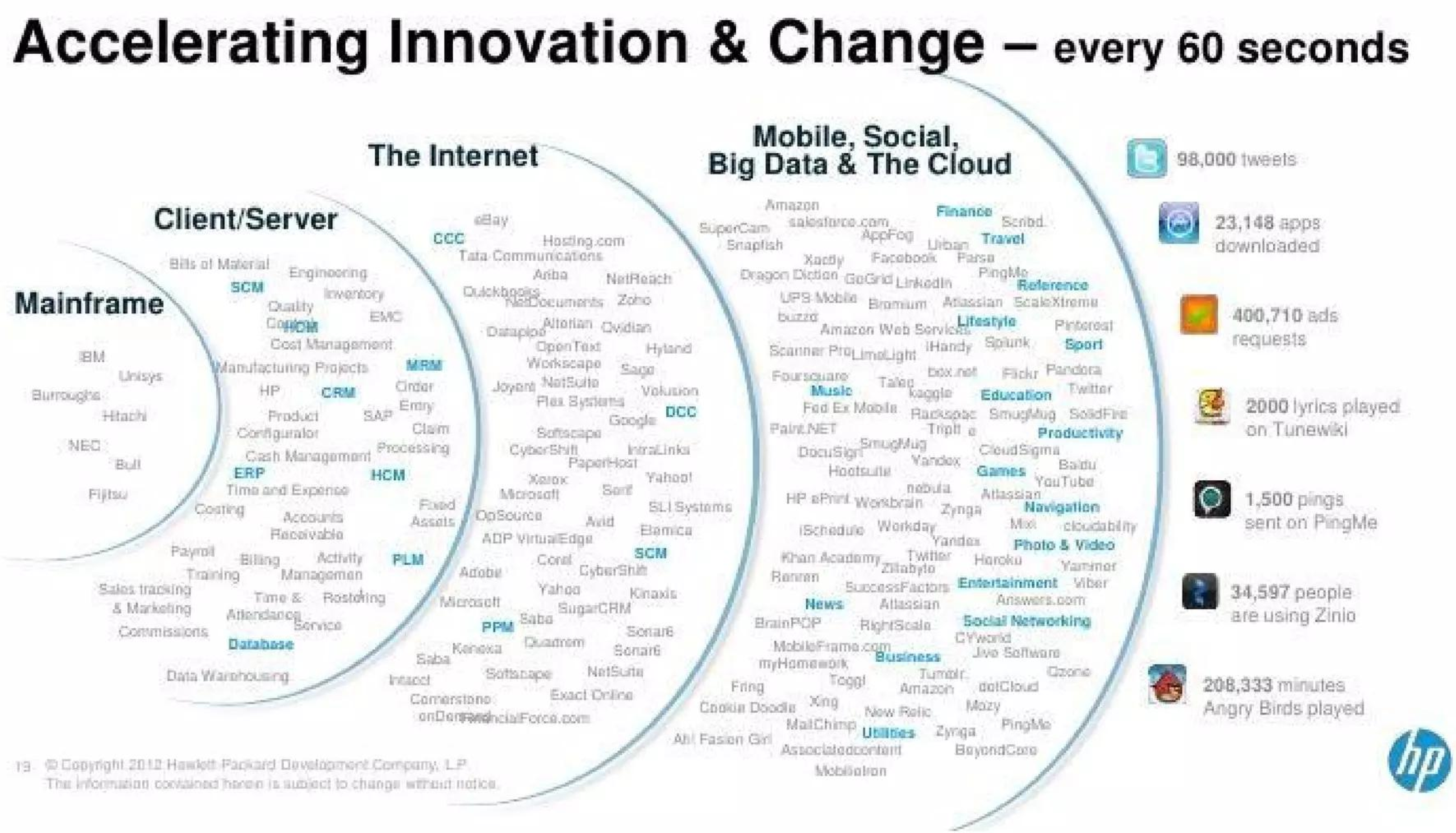
The four layer of Datamining



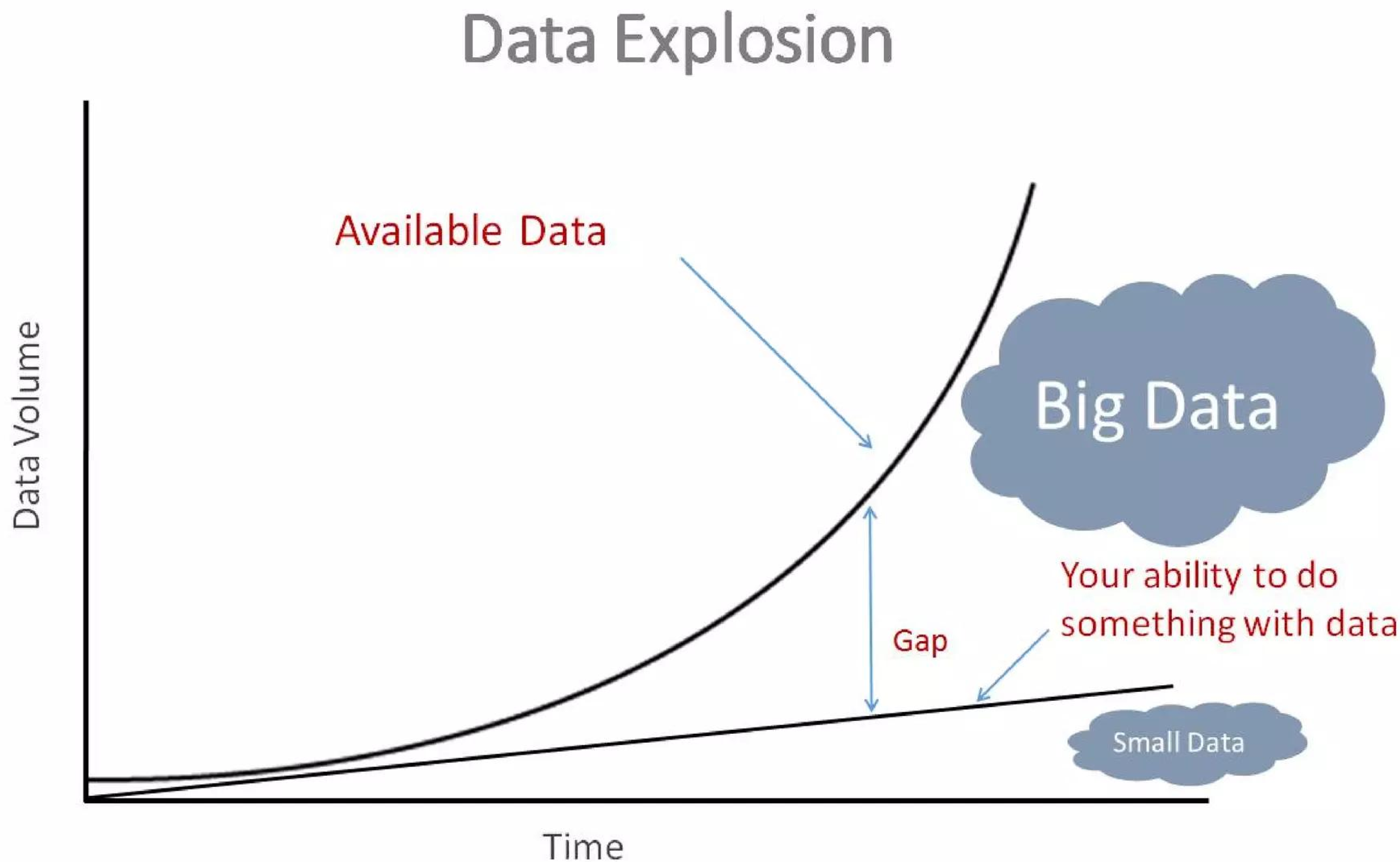
Problems



Complexity #1

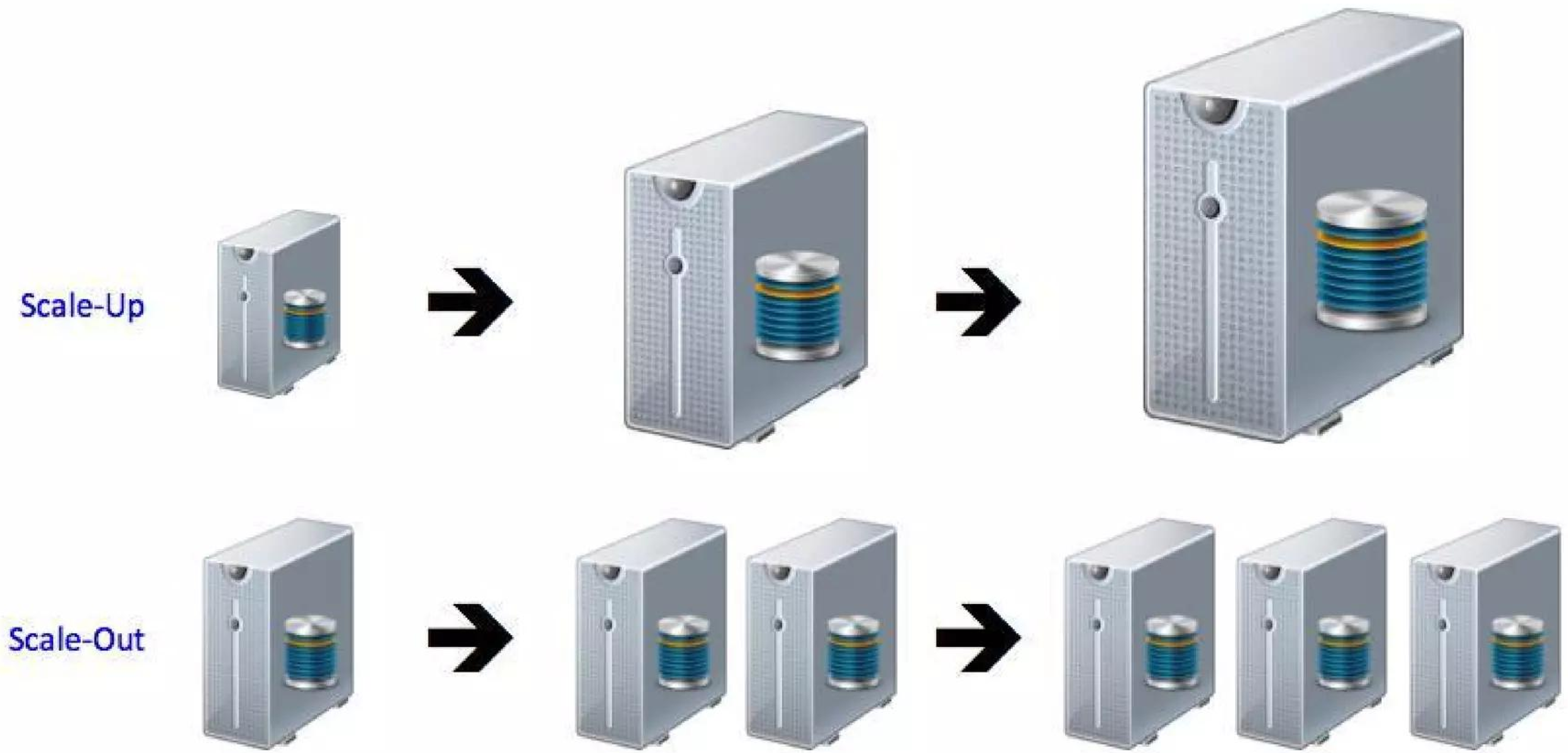


Complexity #2



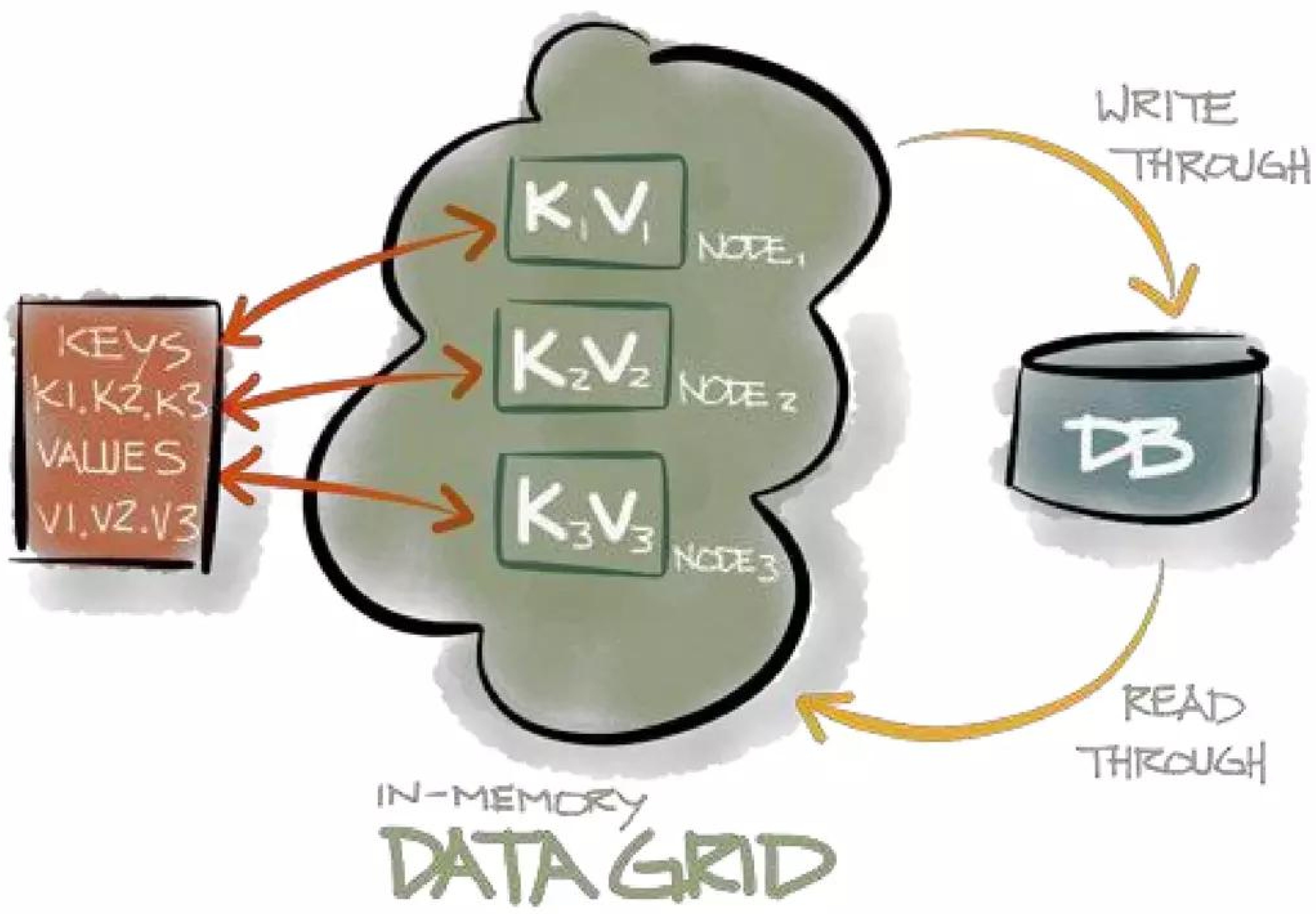
Solutions

Solutions #1



Solutions #2

- **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



Solutions #3

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

Technologies

Tools on python

- 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

Tools on java virtual machine (jvm)

- Weka
- Deeplearin4j (working with spark and gpu)
- H2O (working with spark and GPU, support tensor, mxnet, and cafe)
- JcuDNN (JNI wrapping nvidia cuDNN)
- Mahout (hadoop)
- Mllib Spark

Tools

blibli.com
BIG CHOICES BIG DEALS

NumPy

Seaborn

scikit learn

TensorFlow™

Pandas

elasticsearch

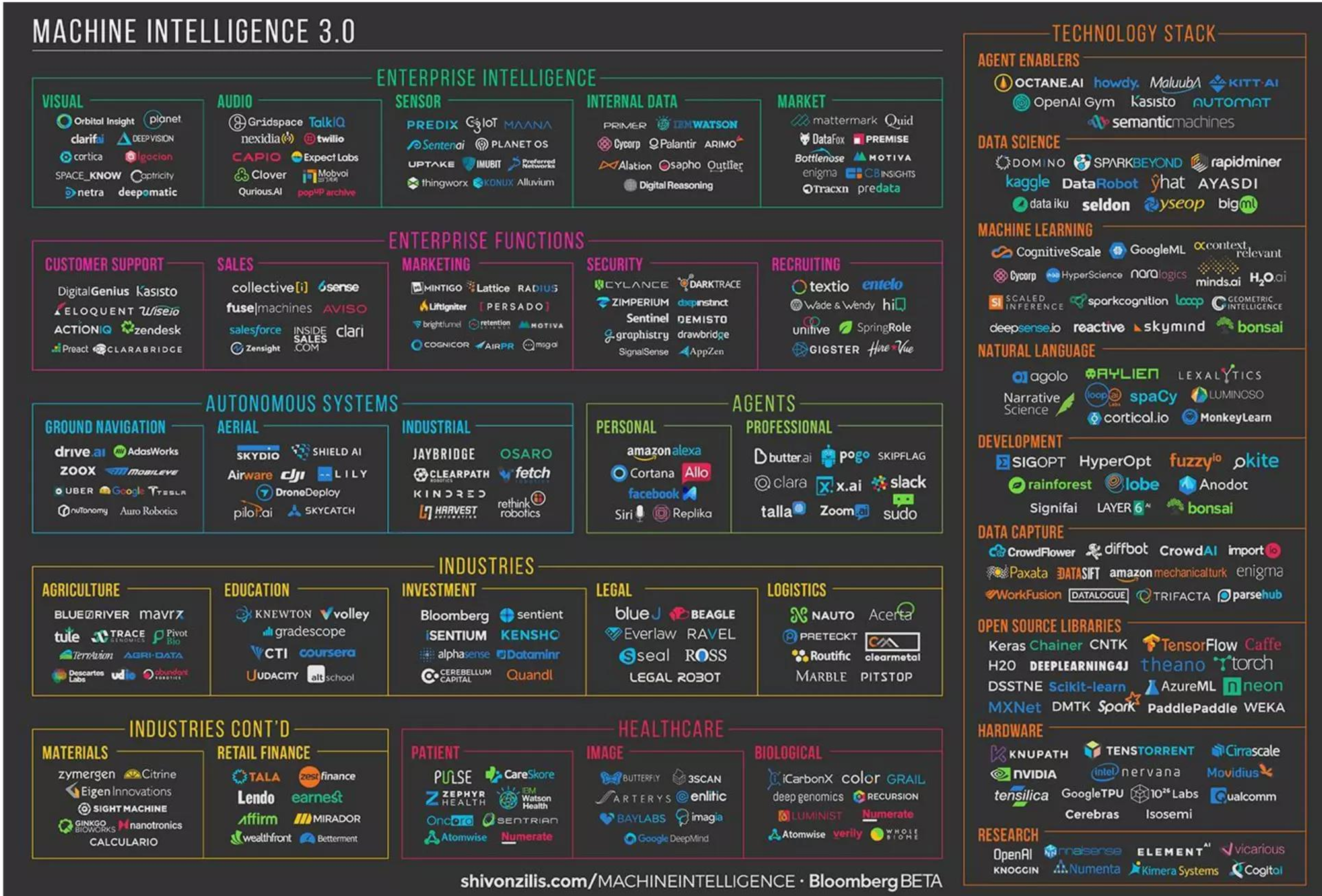
APACHE Spark™

NVIDIA CUDA

OpenCL™

APACHE kafka®

Stack and Services



Stack and Services

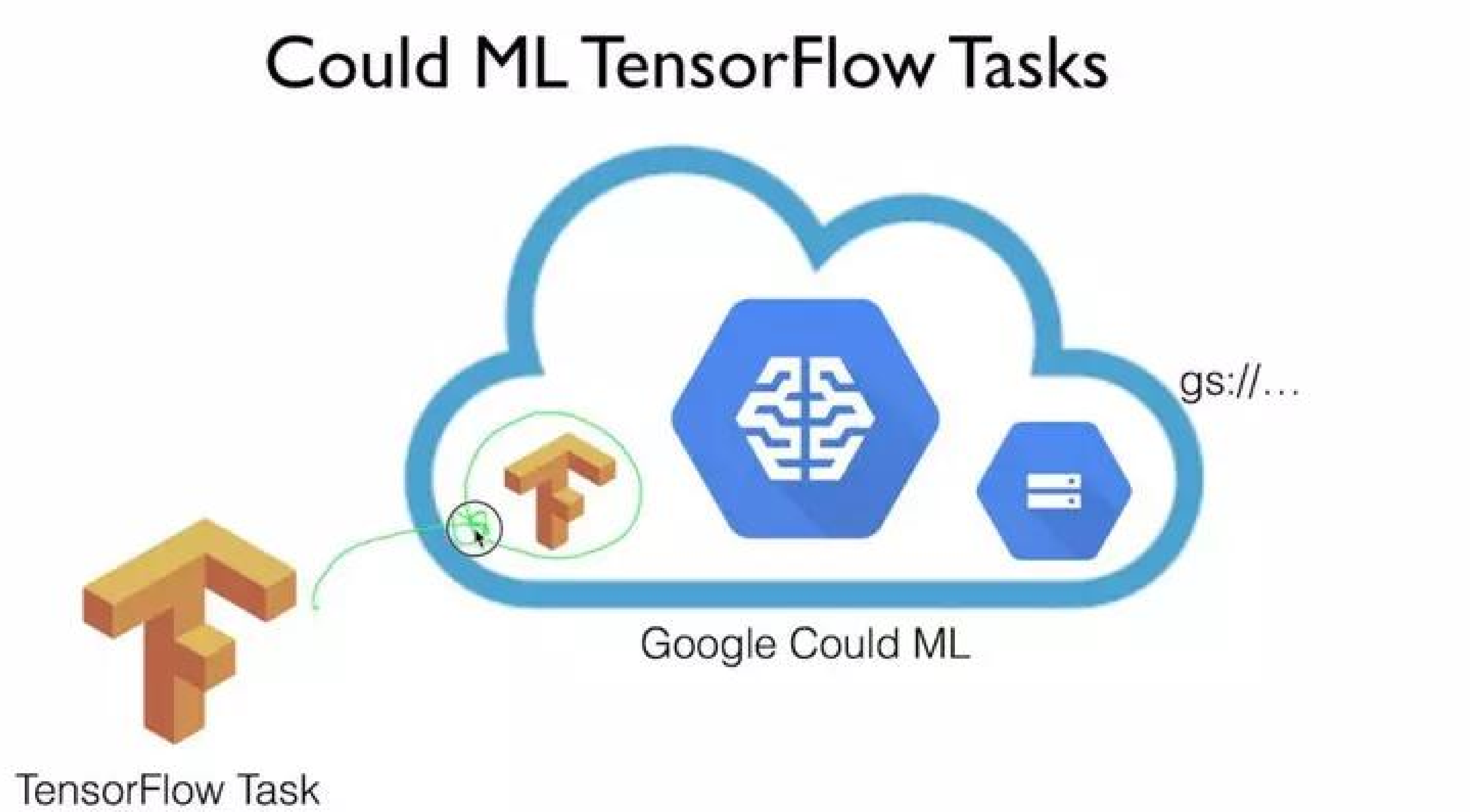
Google Cloud Platform

Microsoft Azure

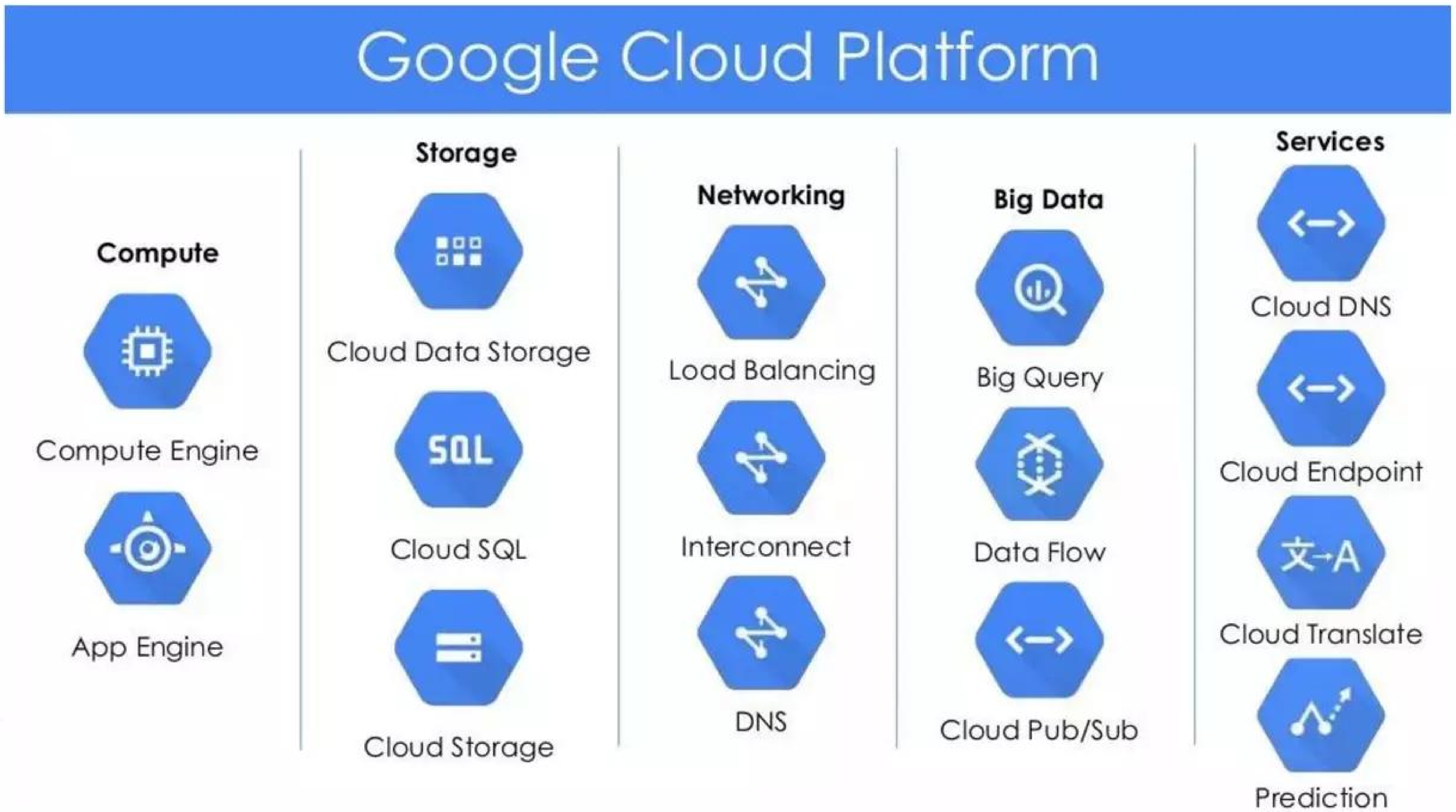
amazon web services™

GCP

Google Cloud Platform
Compute
App Engine
Compute Engine
Container Engine
Storage
Bigtable
Cloud Storage
Cloud SQL
Cloud Datastore
Big Data
Big Query
Pub/Sub
Dataflow
Dataproc
Datalab
Machine Learning
Vision API
Machine Learning
Speech API
Translate API

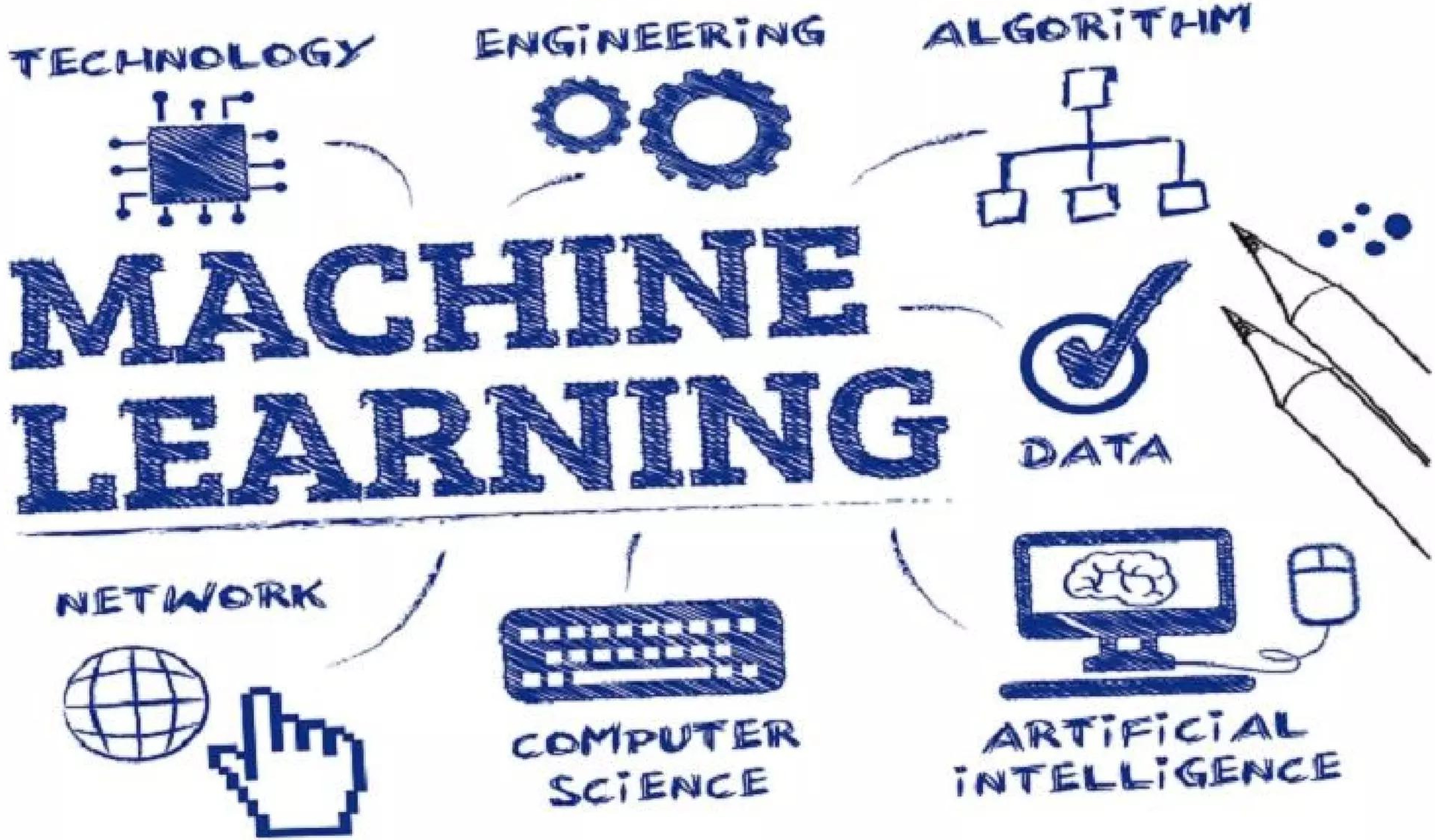


Could ML TensorFlow Tasks
gs://...
Google Could ML
TensorFlow Task



Google Cloud Platform
Compute
Compute Engine
App Engine
Storage
Cloud Data Storage
Cloud SQL
Cloud Storage
Networking
Load Balancing
Interconnect
DNS
Big Data
Big Query
Data Flow
Cloud Pub/Sub
Services
Cloud DNS
Cloud Endpoint
Cloud Translate
Prediction

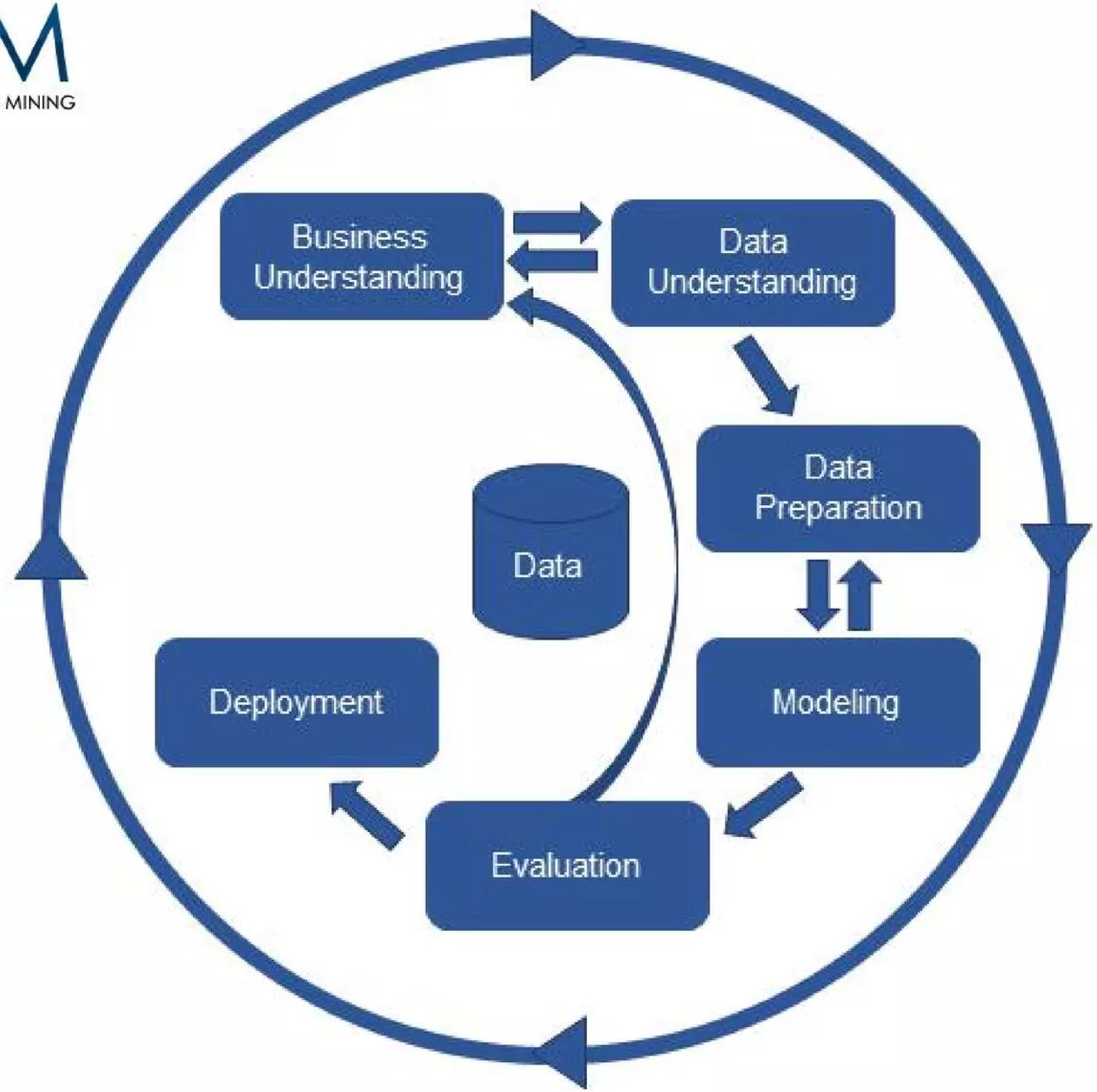
Compile all components



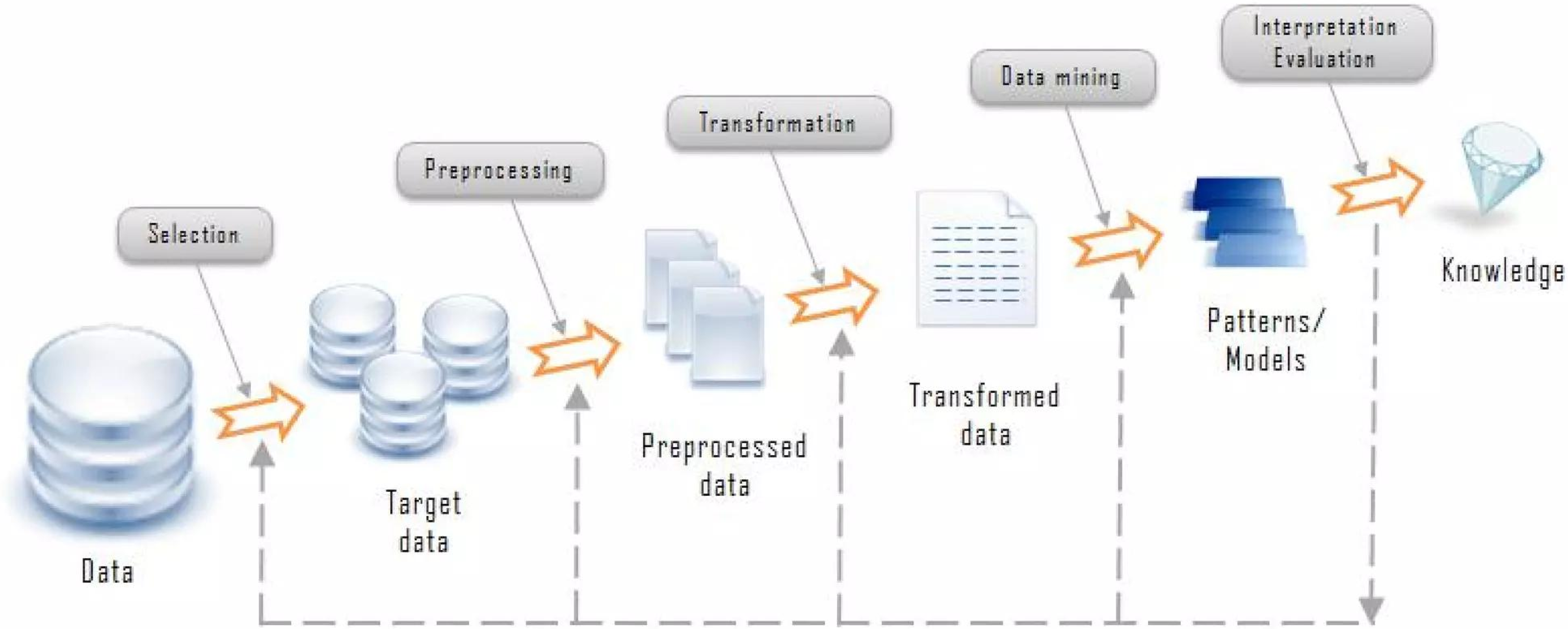
How we applied Machine Learning

CRISP-DM

CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING

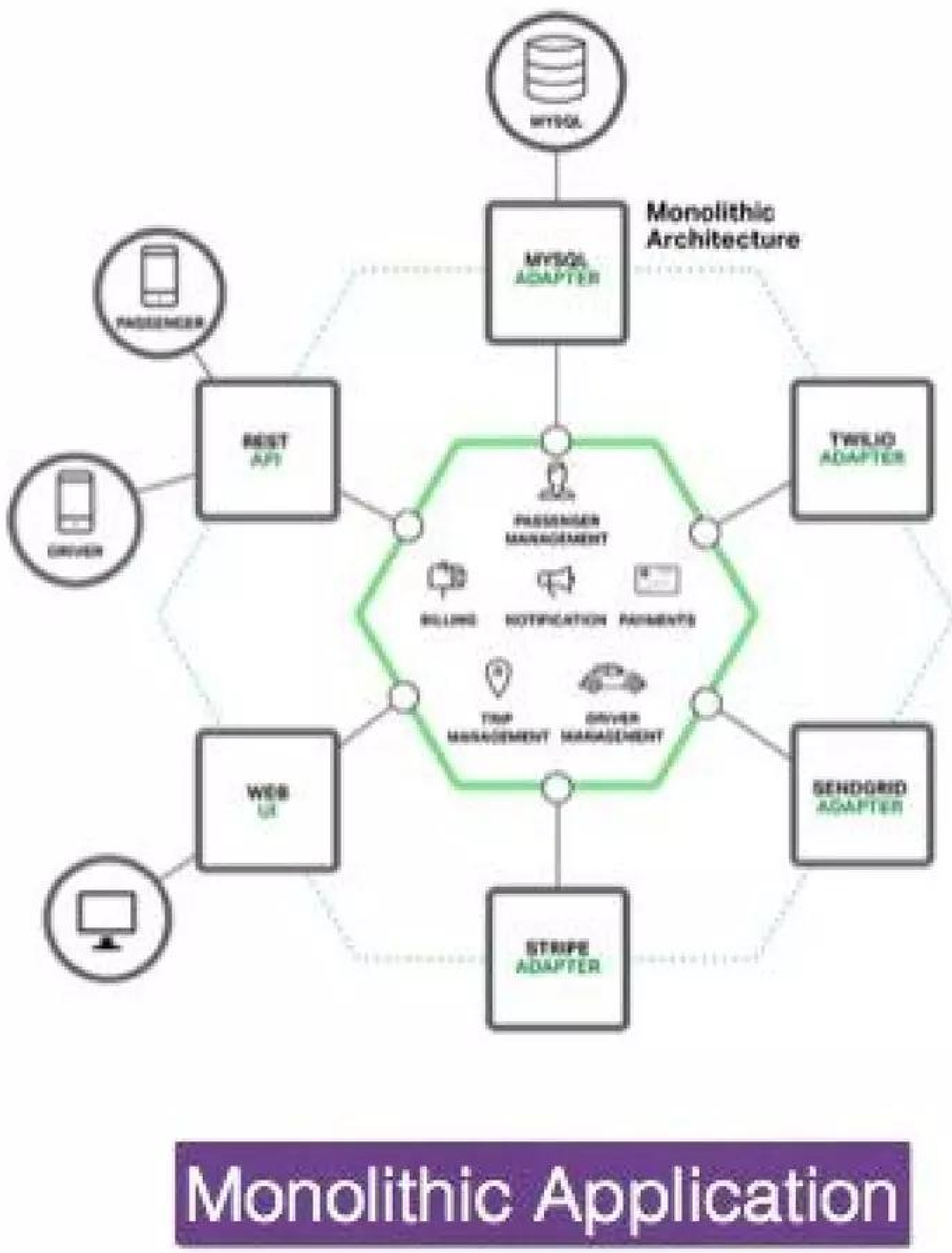


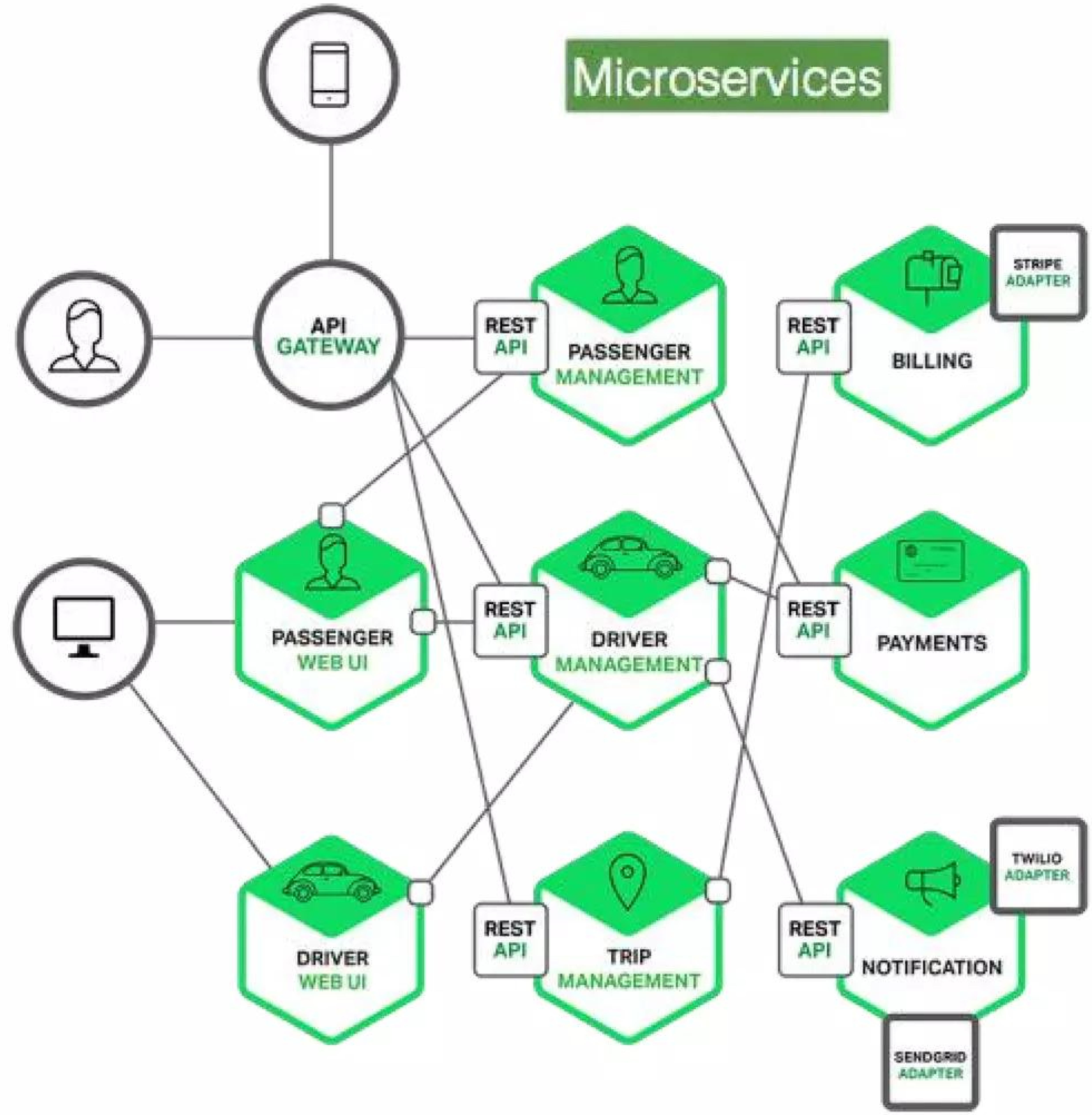
Process mining



Another problem

- We are using micro services

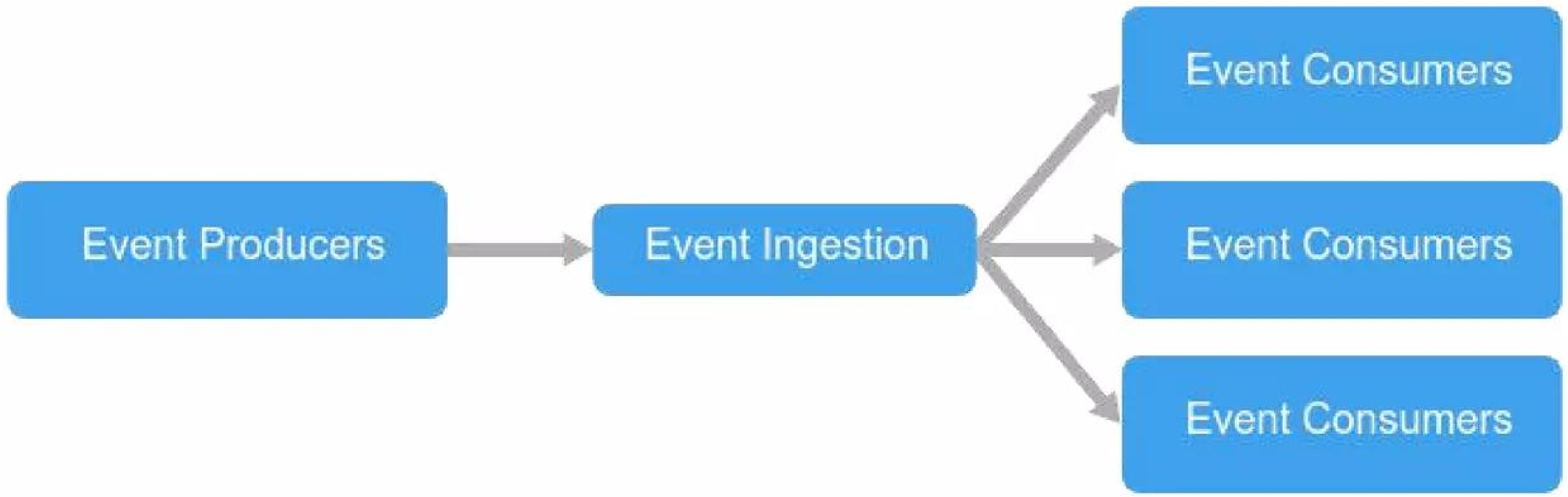


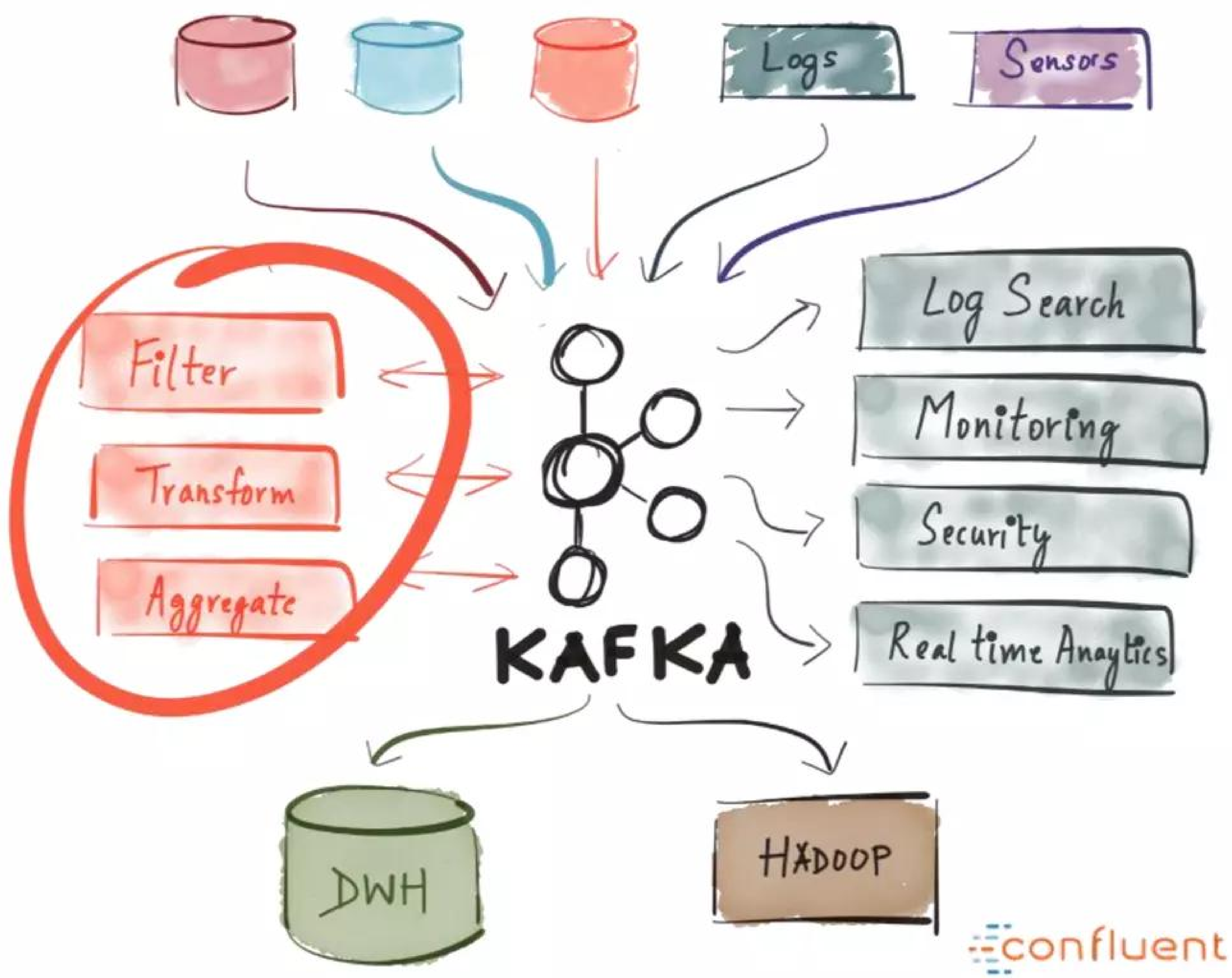


What data that we need???

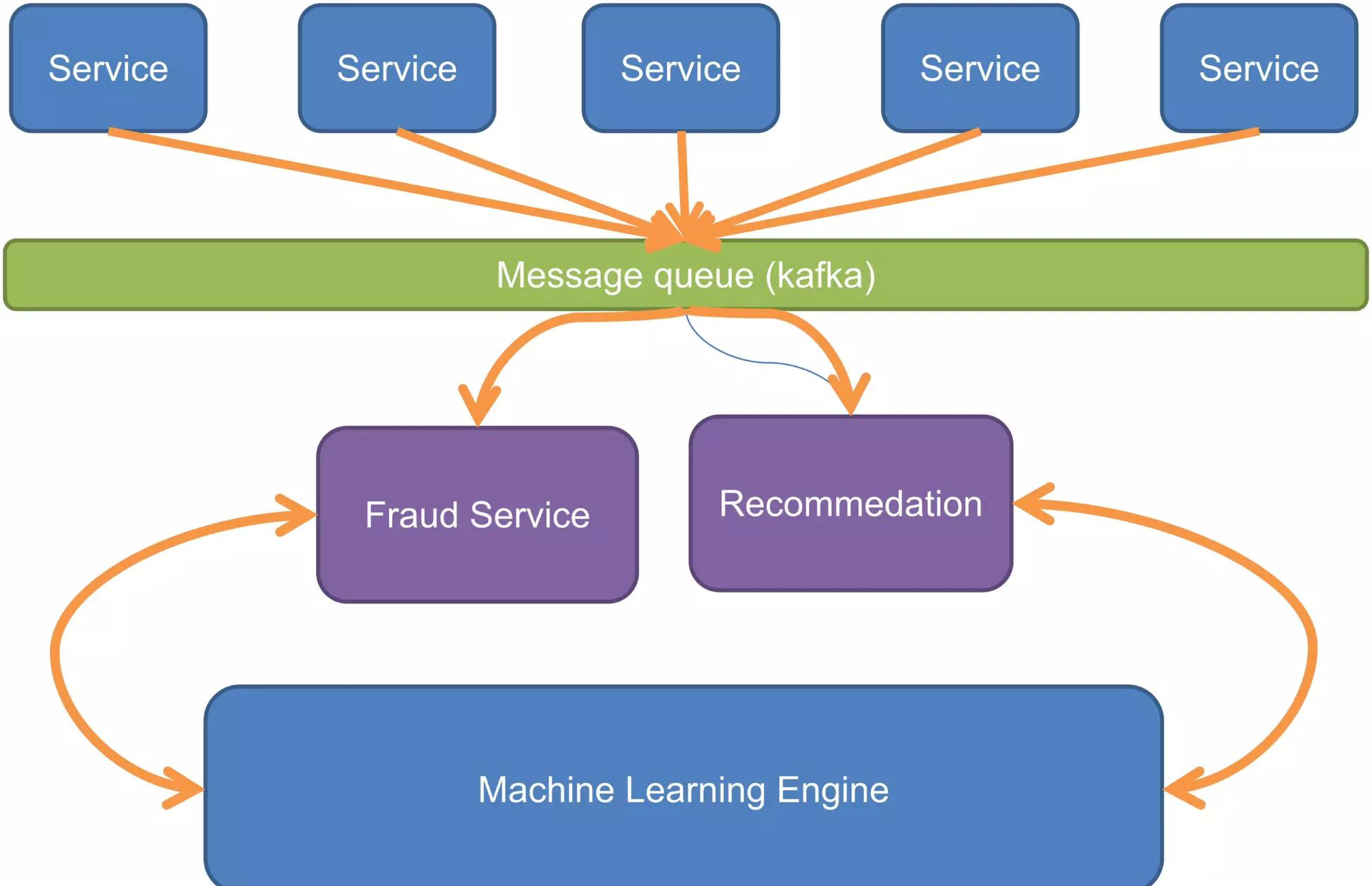
How to Collect

Event Drive Architecture

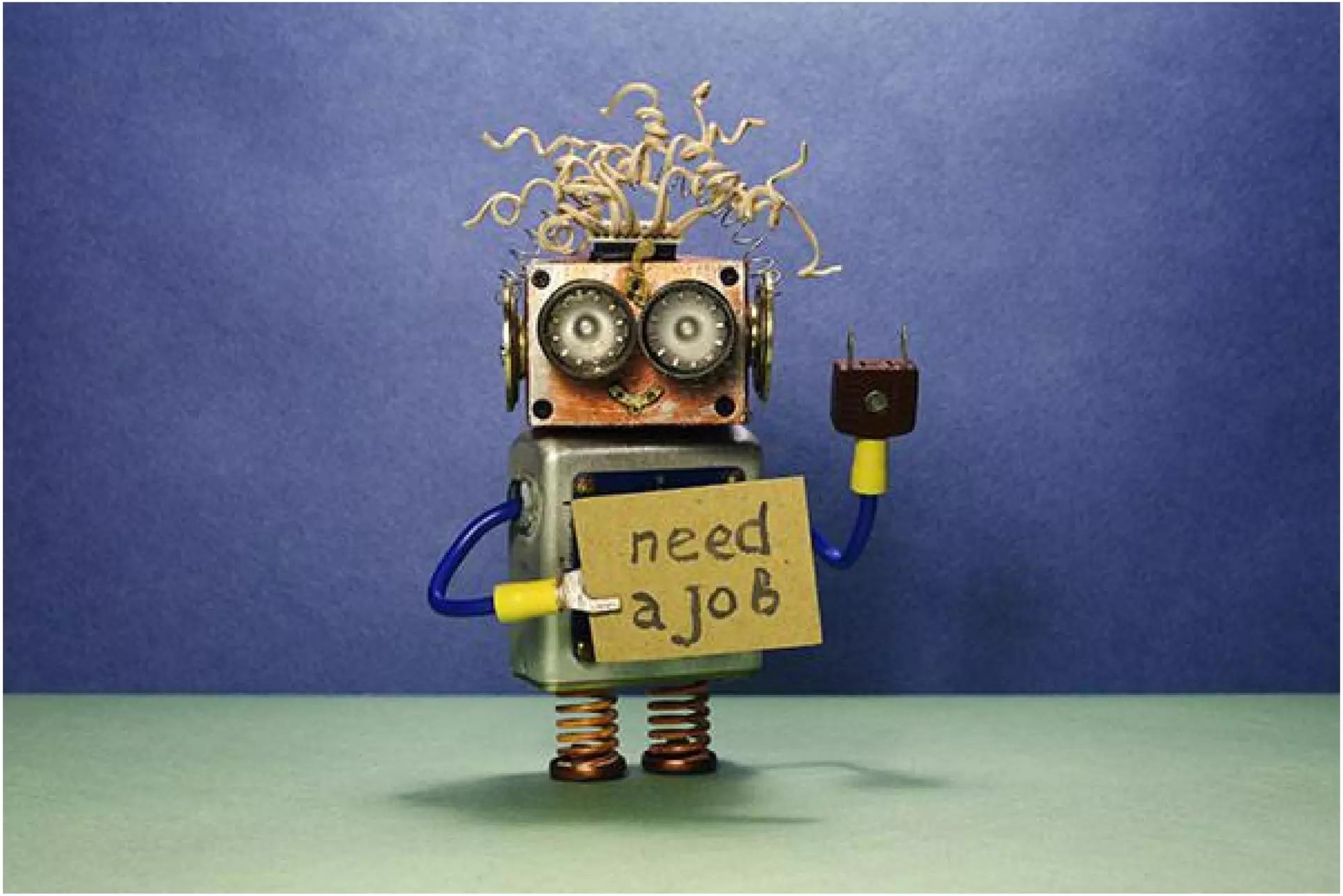




Event Drive Architecture



Cases



Anomaly Detection

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

Challanges

- 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

Anomaly Detection

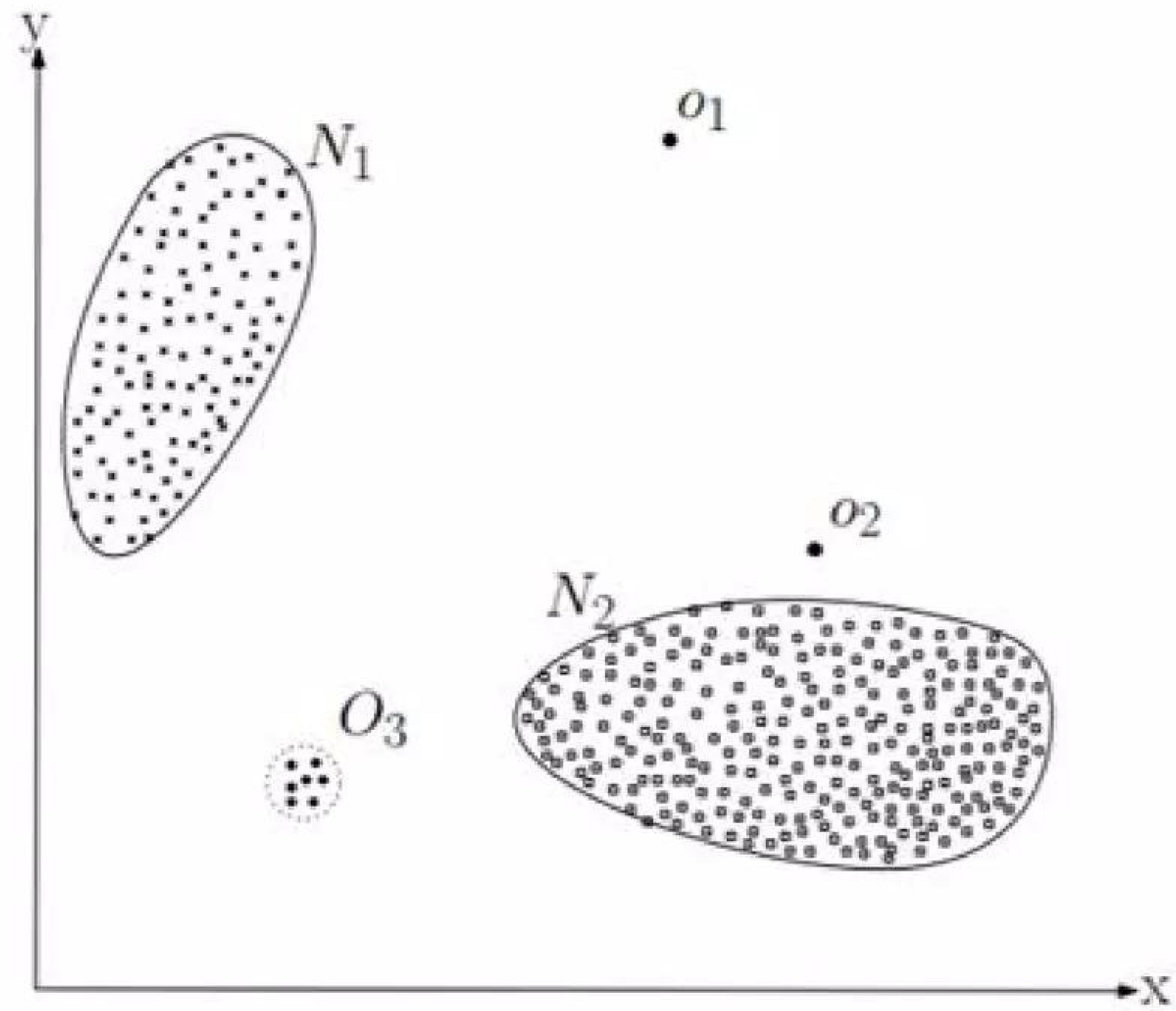


Fig. 1. A simple example of anomalies in a 2-dimensional data set.

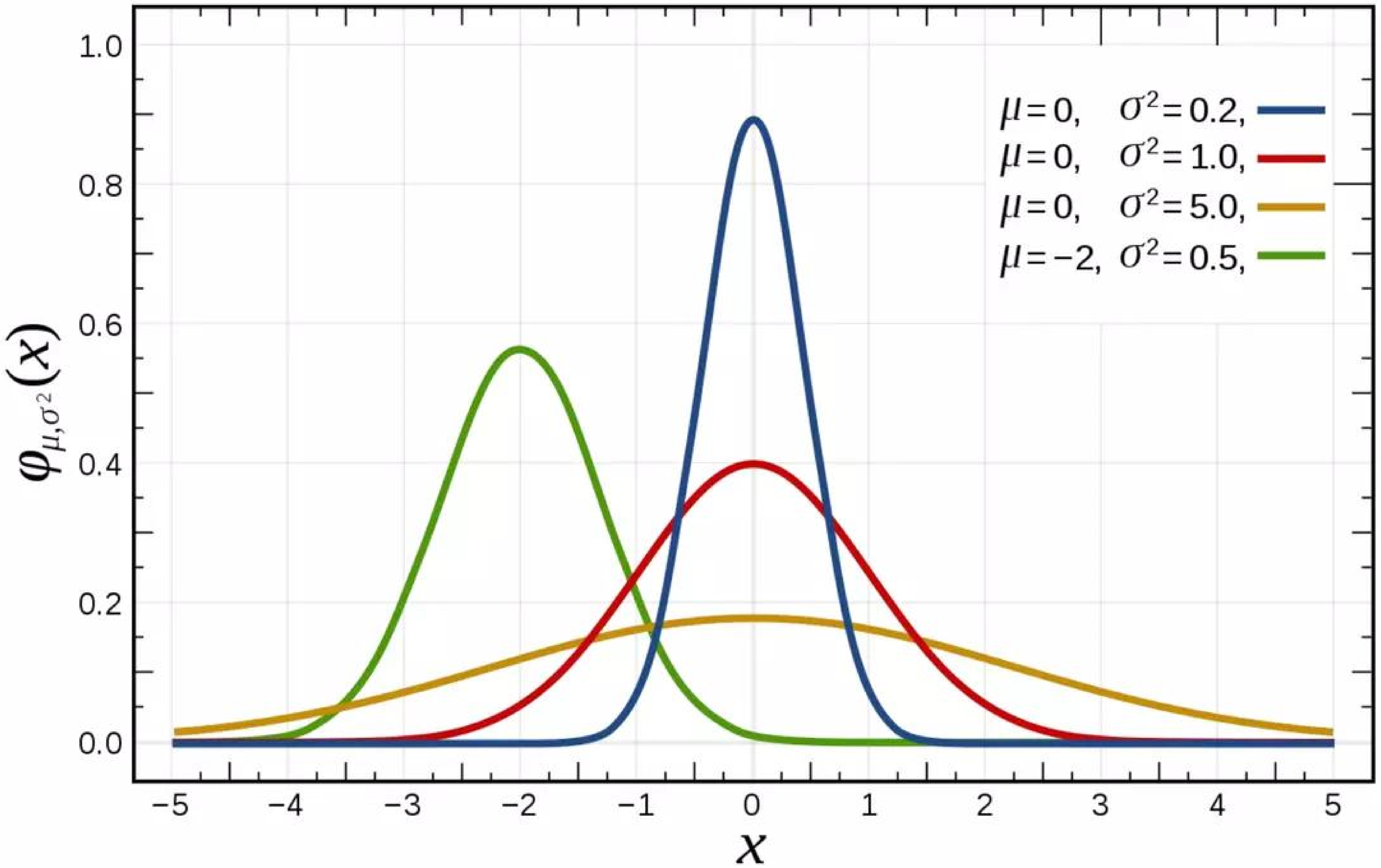
Solution Method :

- Gaussian MIxture Model
- Fitted by EM - Algorithm

Gaussian Distribution

blibli.com BIG CHOICES BIG DEALS

Before GMM, try to remember the gaussian distribution



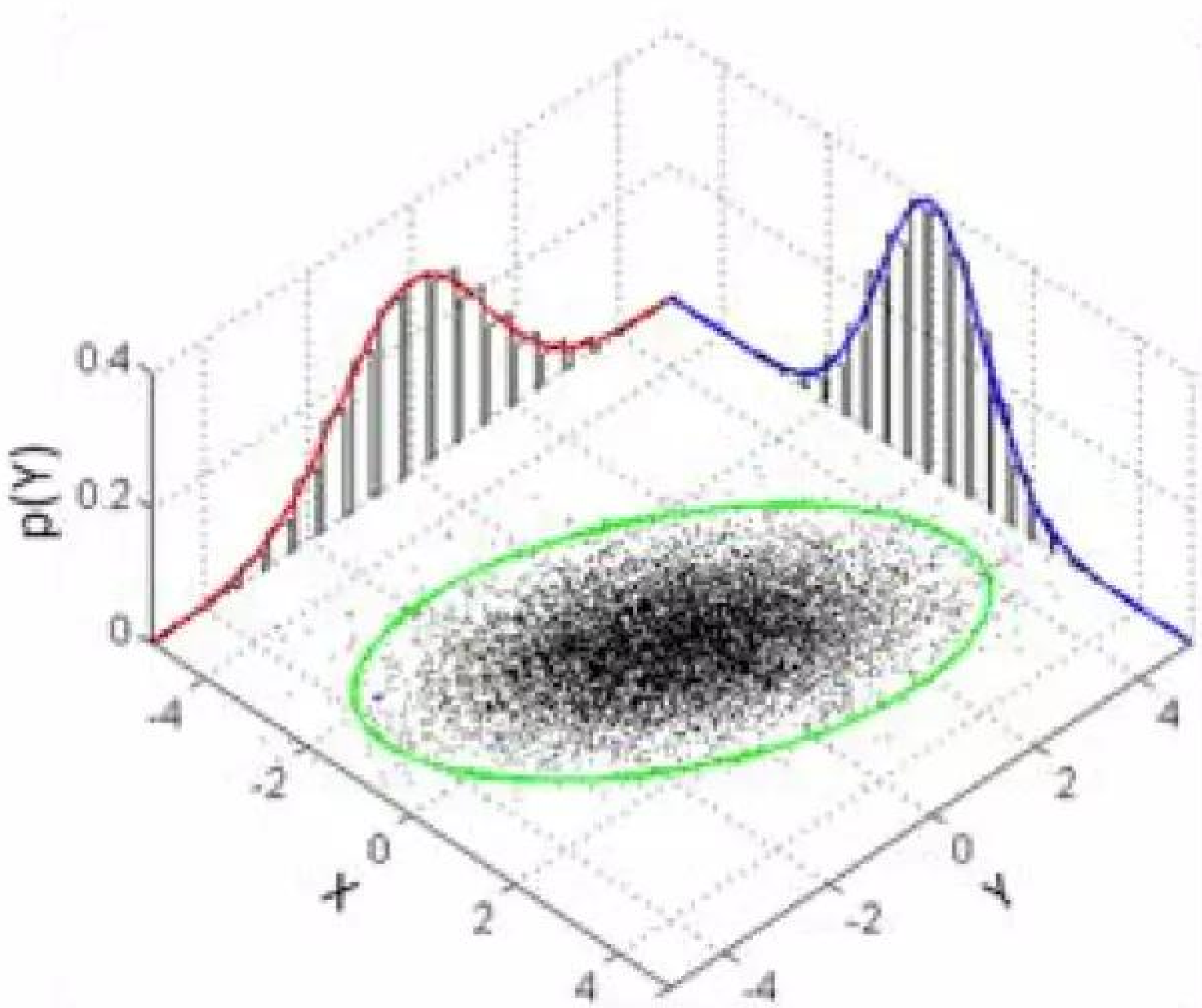
$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

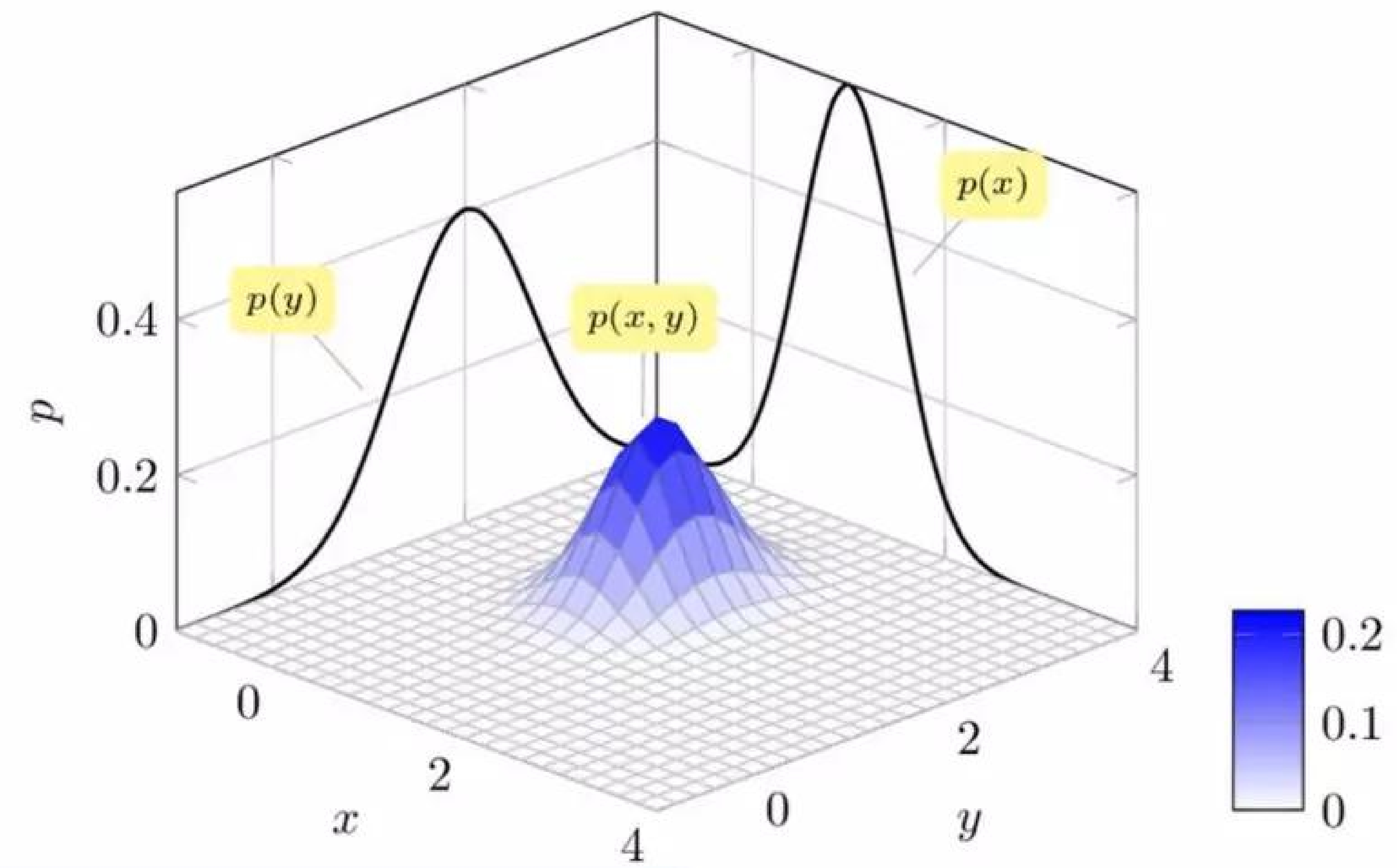
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

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

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

Gaussian Distribution Multivariate

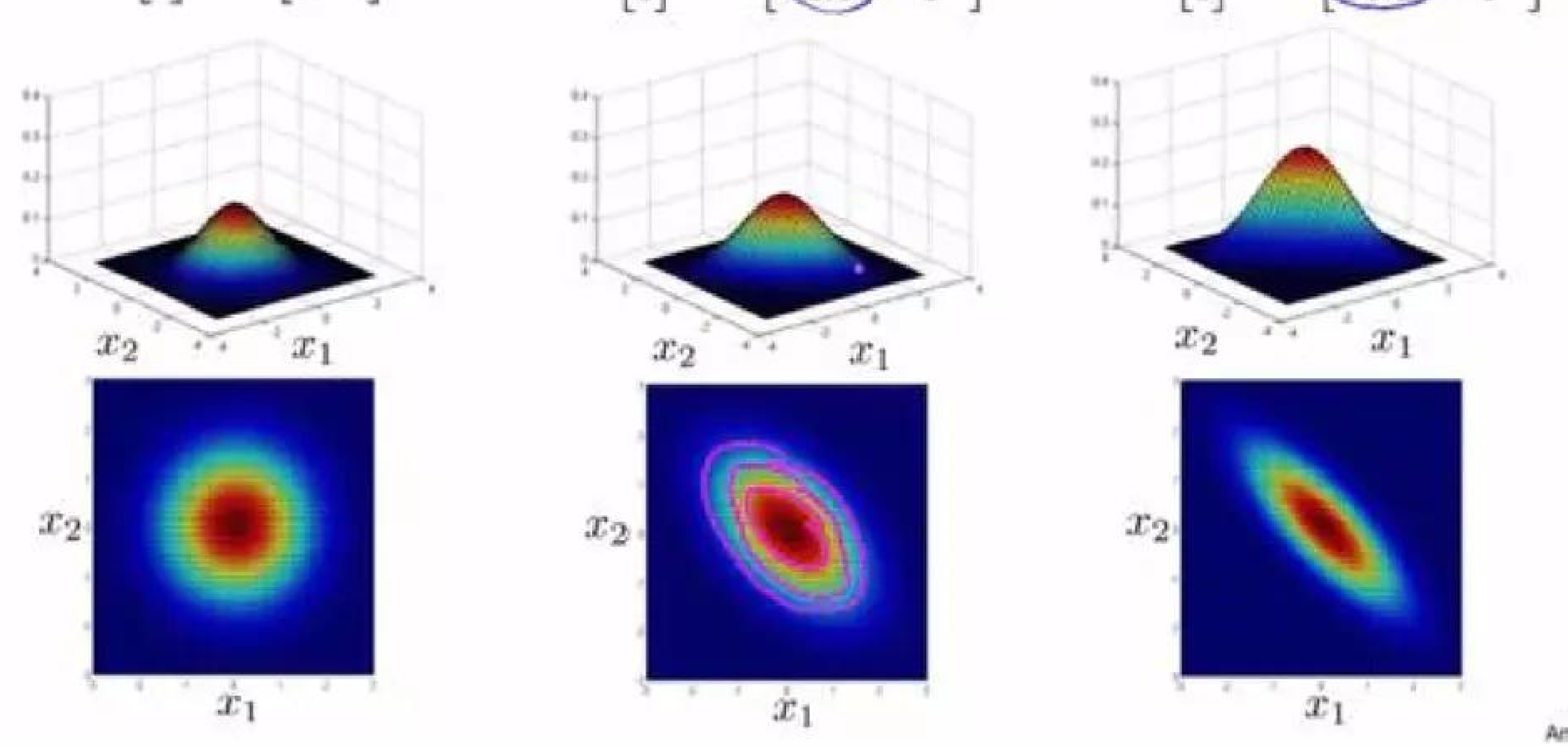




$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2} \parallel \boldsymbol{\Sigma} \parallel^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$$

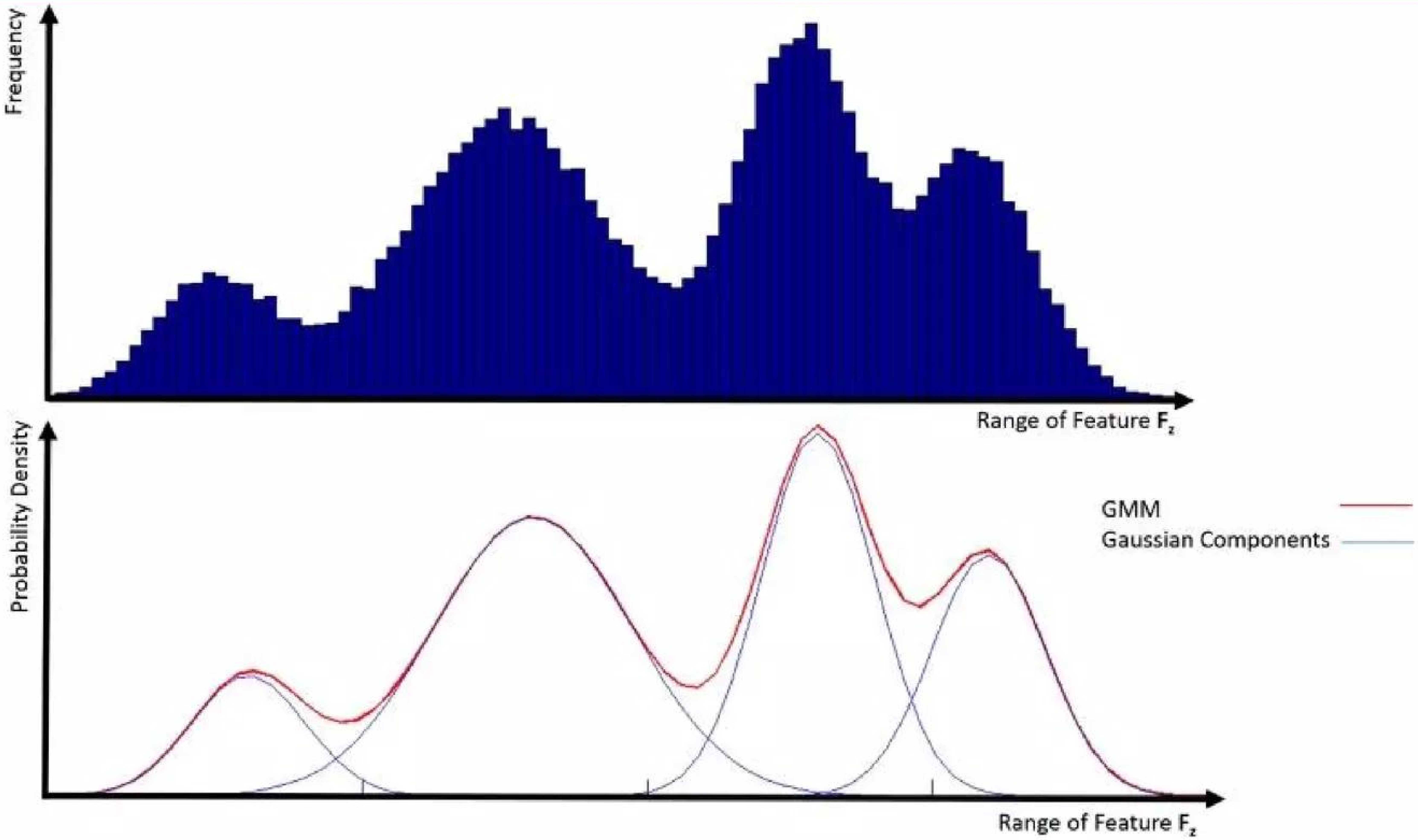
Multivariate Gaussian (Normal) examples

$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \; \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\quad$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \; \Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ $\quad$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \; \Sigma = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$



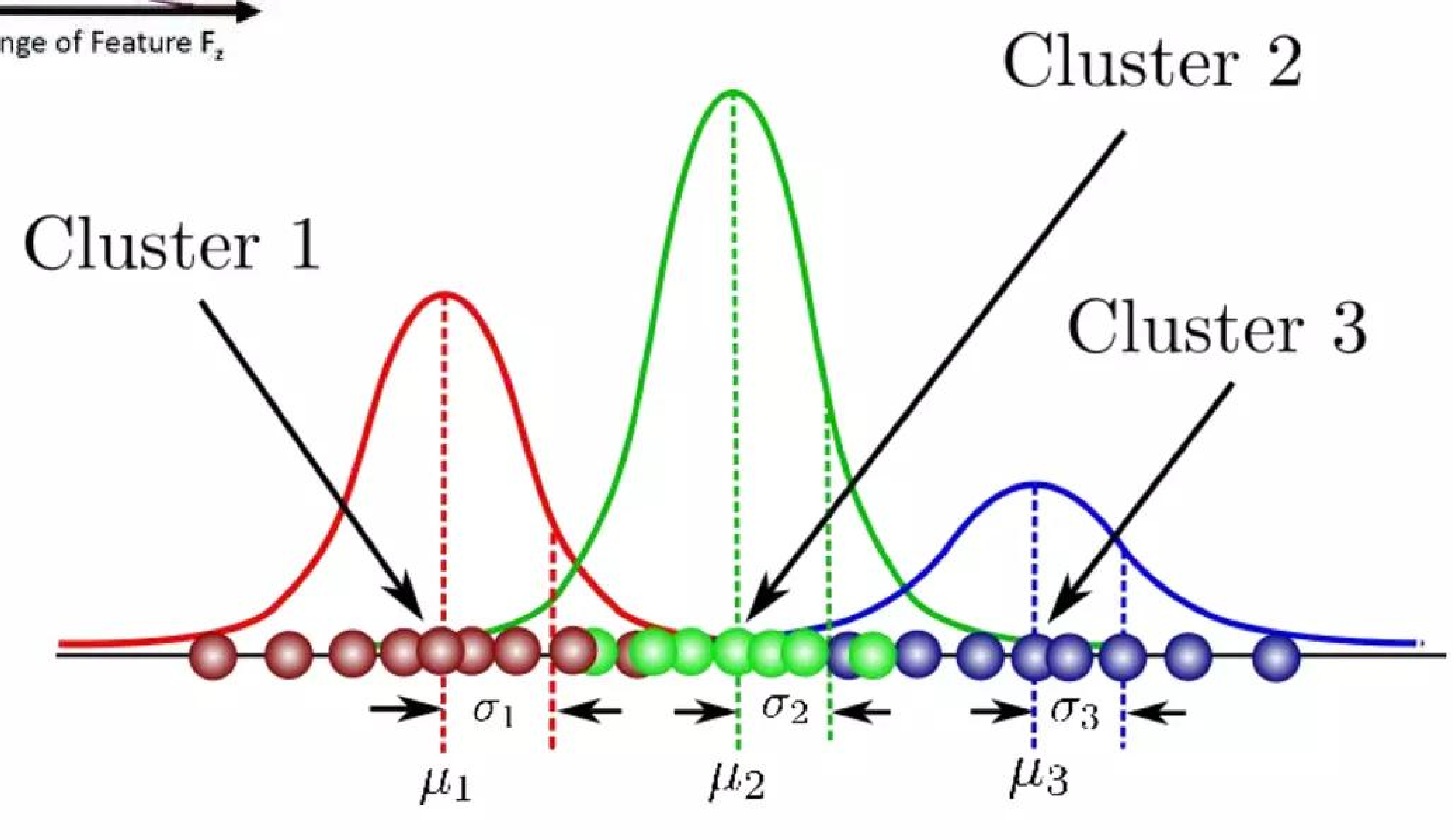
Andrew N

Gaussian Mixture

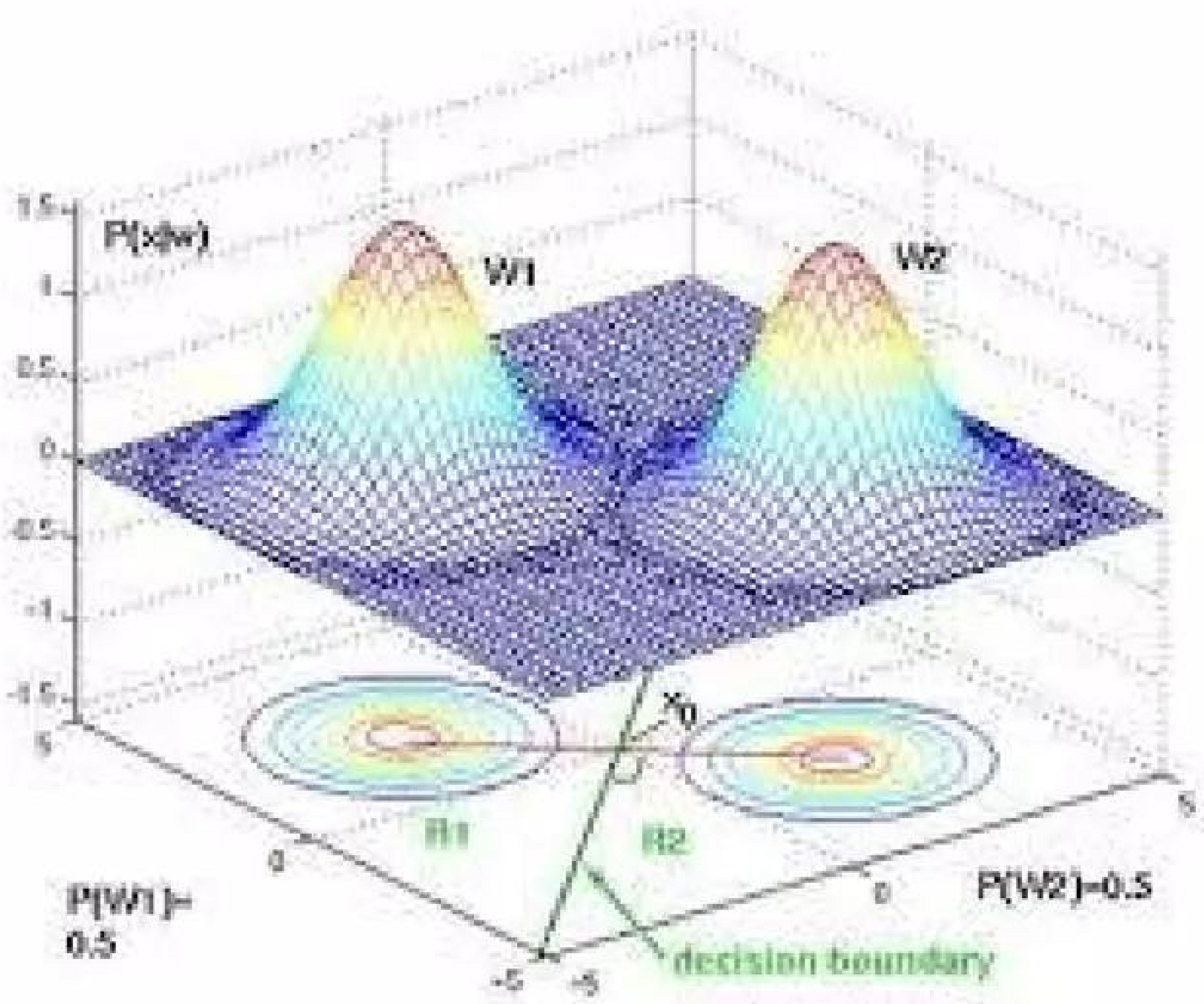


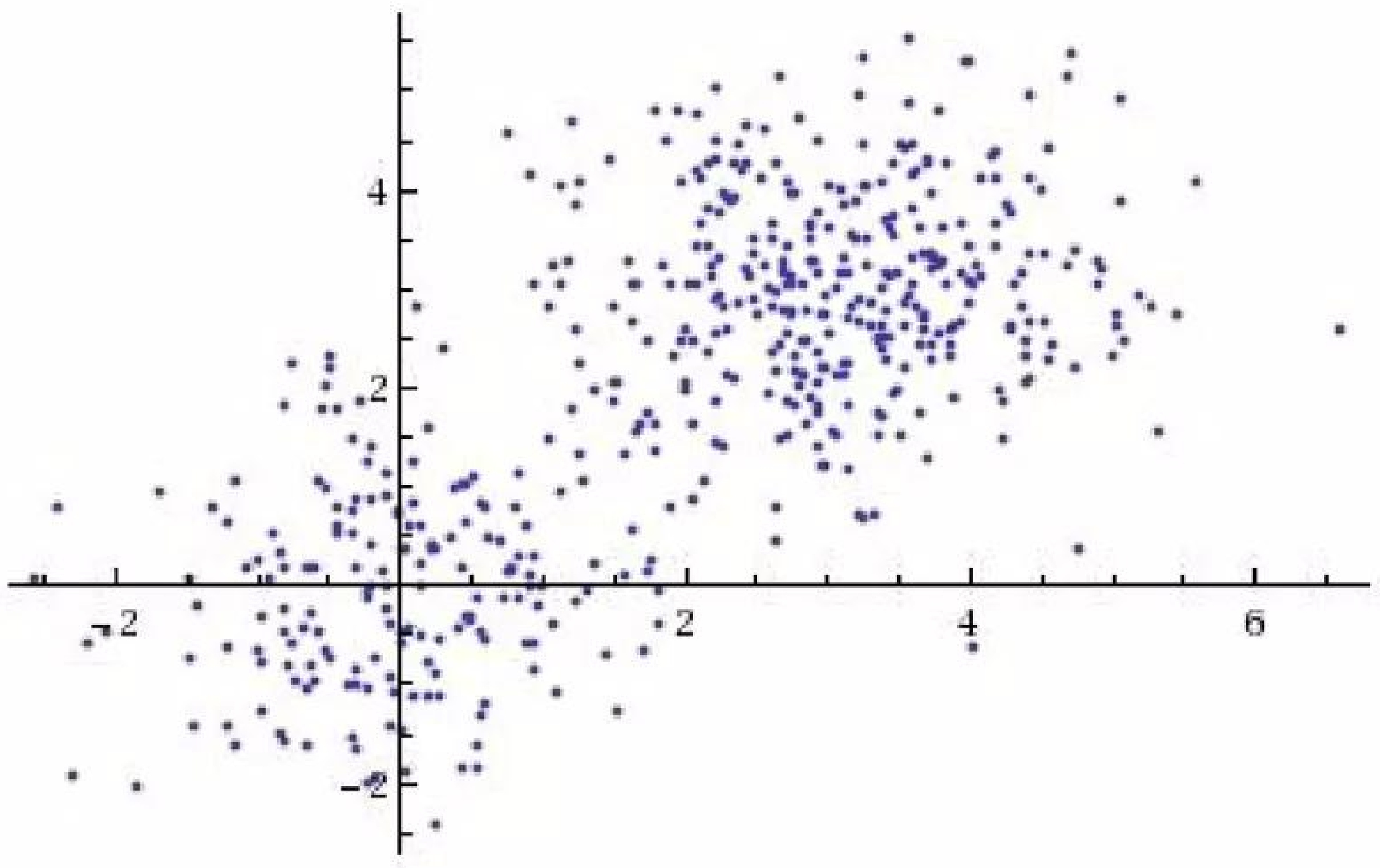
We have 4 gausians mean 4 clusters on the left picture.

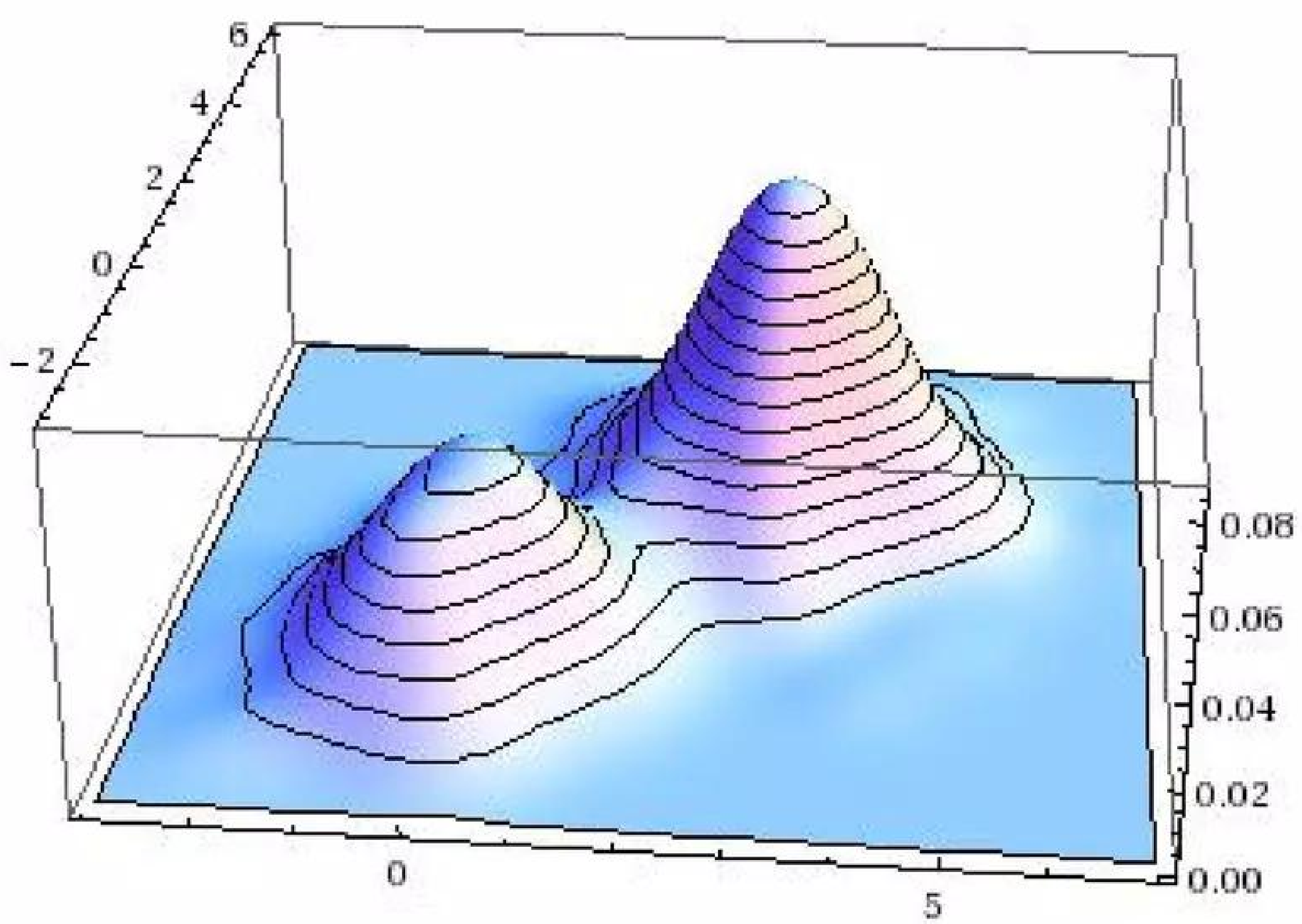
We have 3 gausians mean 3 clusters on the right picture.



Gaussian Mixture (Multivariate)







Gaussian Mixture Multivariate

blibli.com BIG CHOICES BIG DEALS

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.

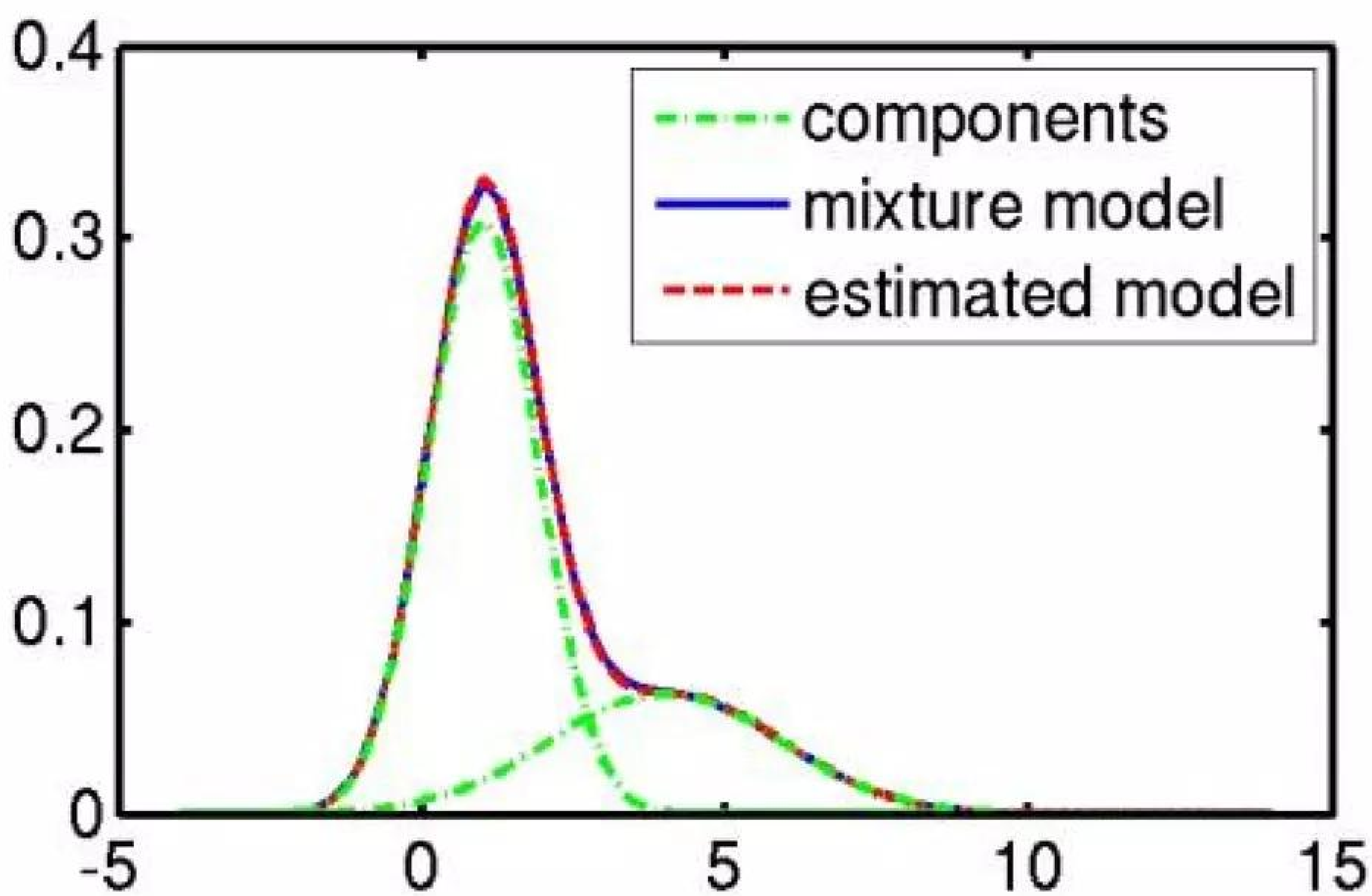
$$f(x) = \operatorname{argmax}\{p(x / \mu_1, \Sigma_1), p(x / \mu_2, \Sigma_2), p(x / \mu_3, \Sigma_3), \ldots p(x / \mu_{n+1}, \Sigma_{n+1})\}$$

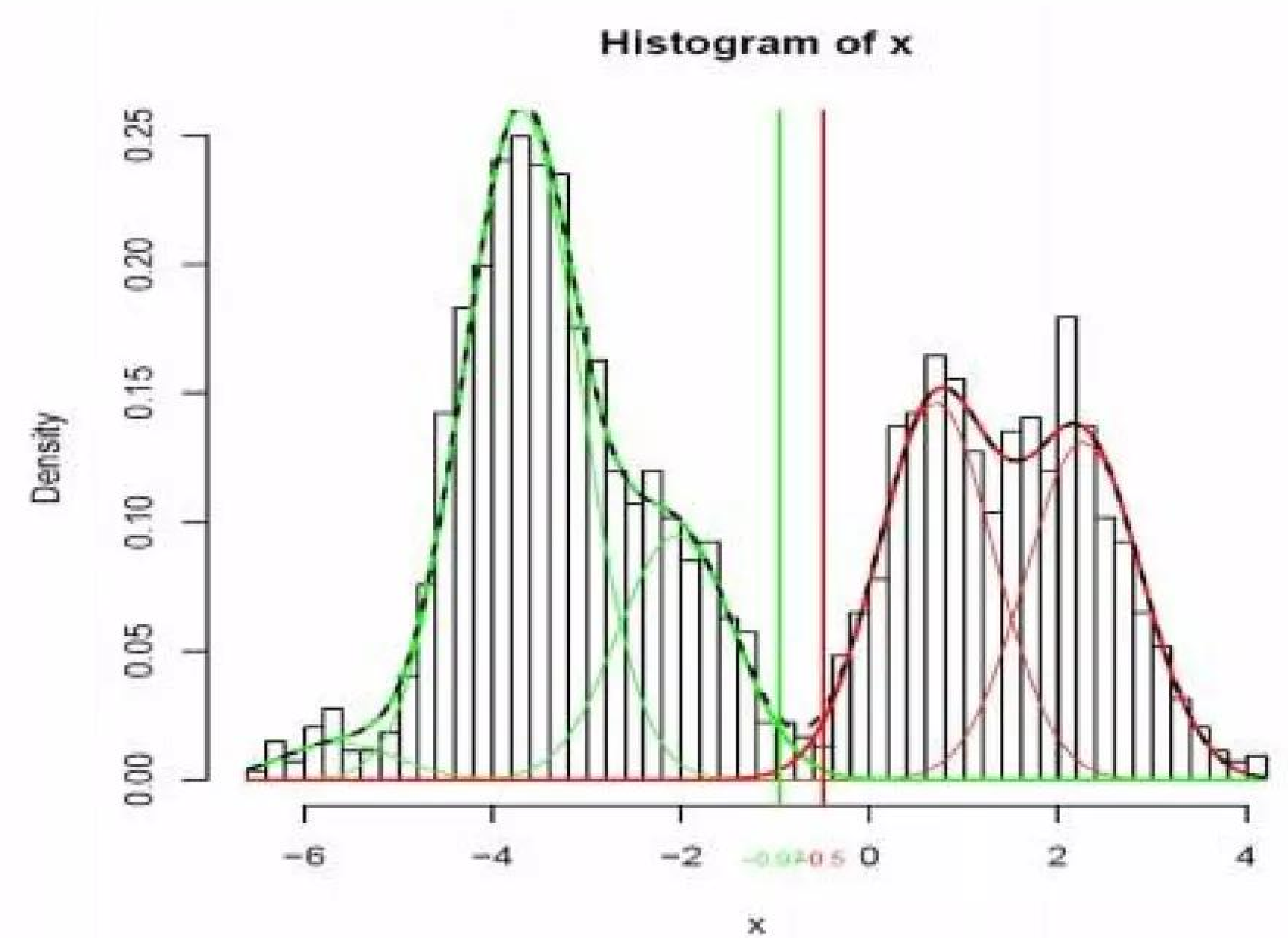
where p is :

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

EM (Expectation Maximization) Algorithm

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





EM (Expectation Maximization) Algorithm

blibli.com BIG CHOICES BIG DEALS

- 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 :

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

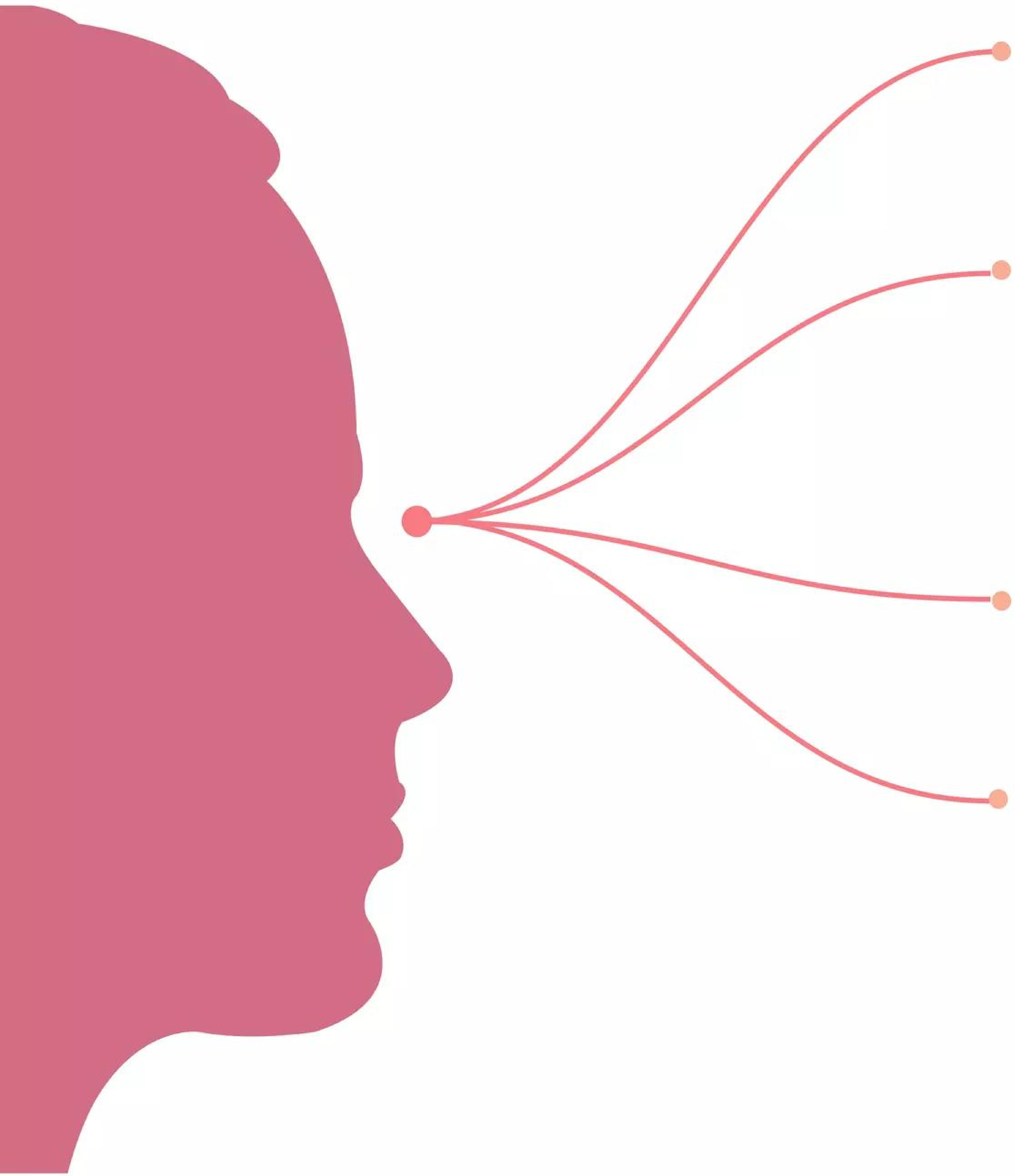
, Dimana $P(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma).$

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

$\theta_{i=1\ldots K}$	=	parameter of distribution of observation associated with component i
$\phi_{i=1\ldots K}$	=	mixture weight, i.e., prior probability of a particular component i
$\boldsymbol{\phi}$	=	K-dimensional vector composed of all the individual $\phi_{1\ldots K}$; must sum to 1
$z_{i=1\ldots N}$	=	component of observation i
$x_{i=1\ldots N}$	=	observation i

Fraud Detection

Fraud Detection



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

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

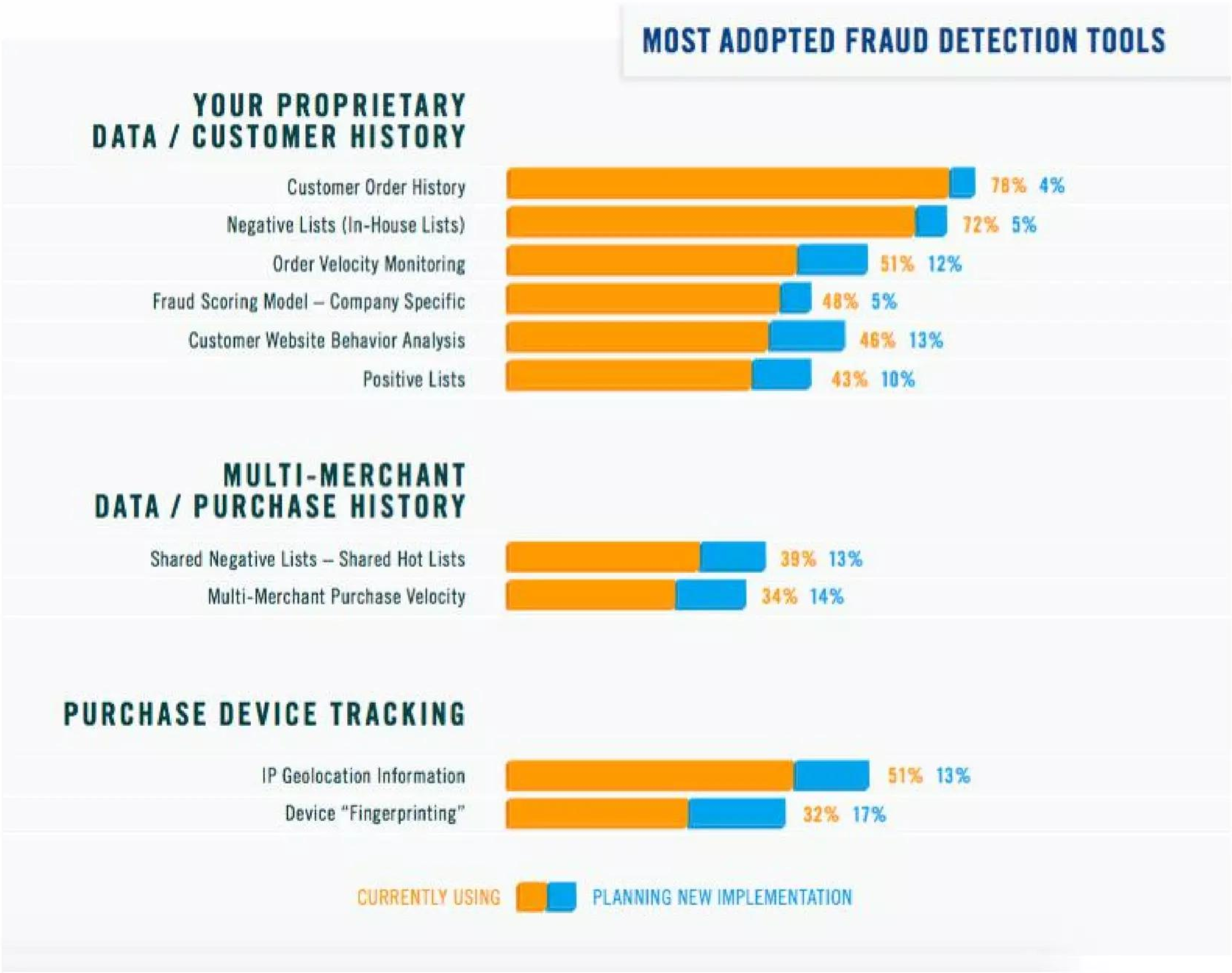
Fraud not only result in financial losses but also produce some reputational risk.

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

[E. Duman et al, 2013]

Annual Reports Cybersource

Annual Reports Cybersource



blibli.com
BIG CHOICES BIG DEALS

THANK YOU

Any question?