

Factory Analytics Automation

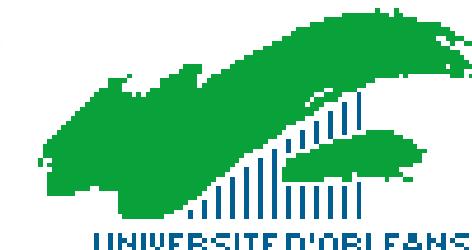
Grégoire de Lassence





- Grégoire de Lassence
Responsable Pédagogie et Recherche (09/2002)
- Département Académique
Tel : +33 6 15 64 04 25
gregoire.delassence@sas.com ou fragrd@sas.com

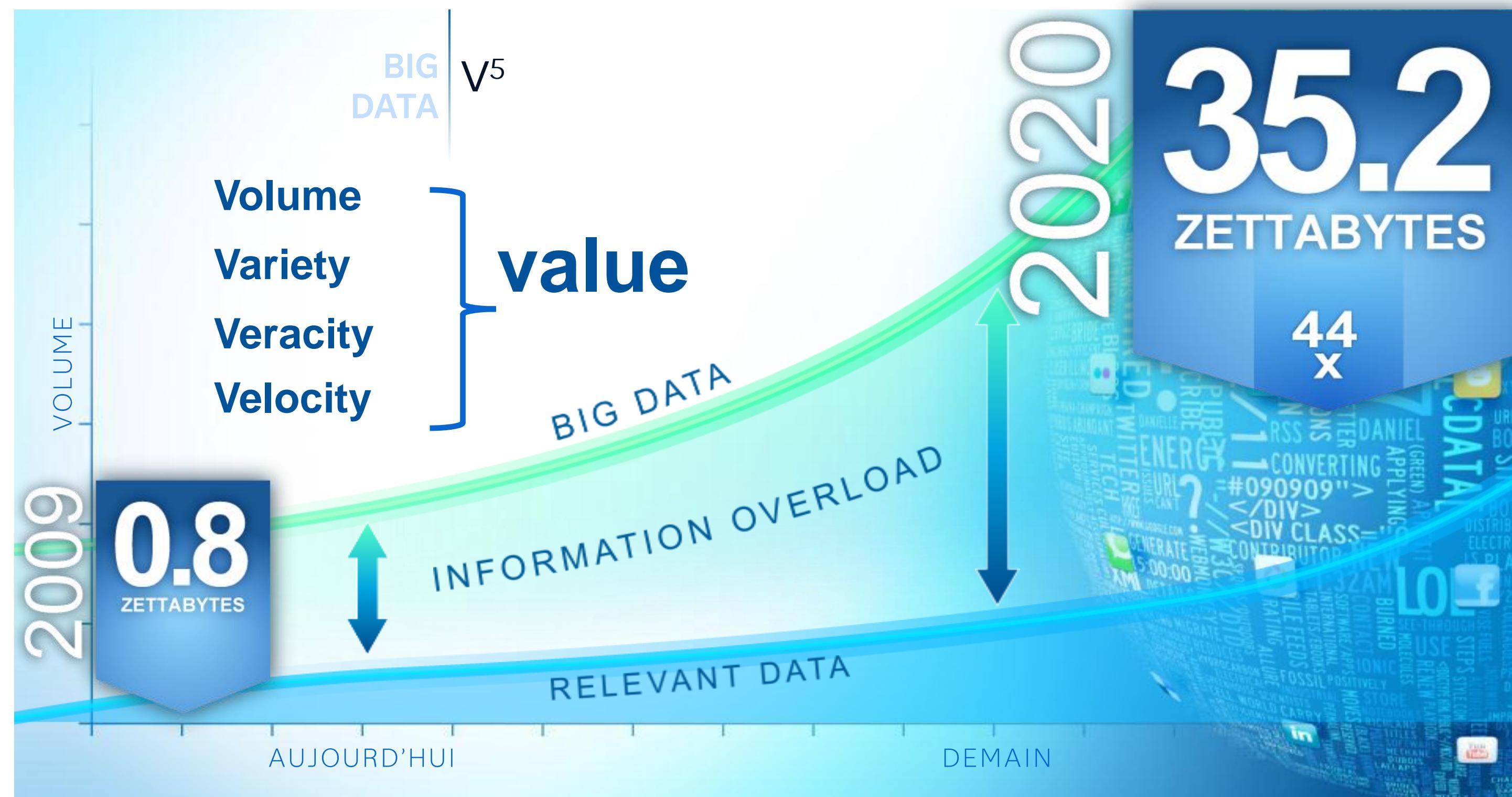
INSEAD



SAS

- Created in 1976
- Revenue 2022 : 3.3 milliards \$ (+8%)
- 12 000 employees
 - 220 employees in France
- 25 % of revenue reinvest in R&D





BIG DATA



“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.”

- Big Data
 - 1%
 - Disruptifs

Masaru Kitsuregawa

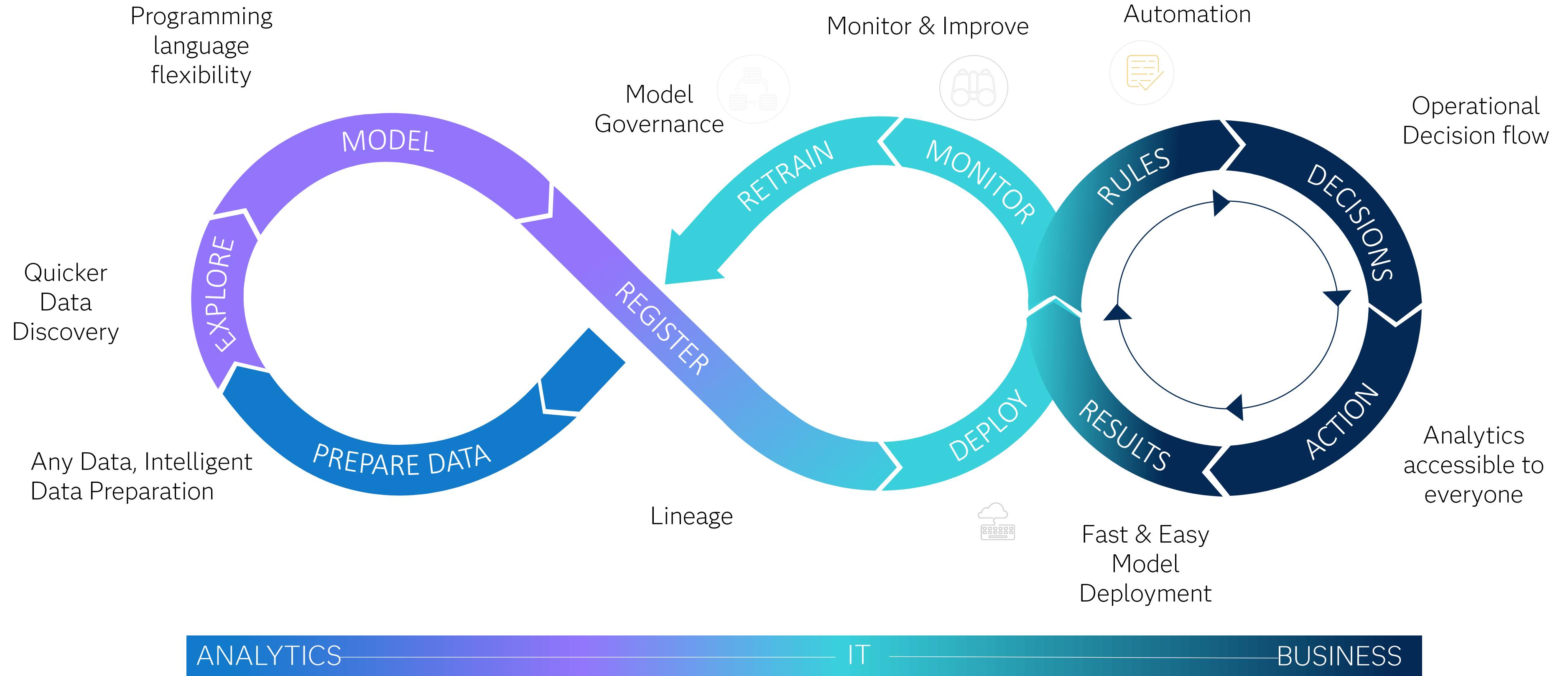
CEO National Institute of Informatics –
Japon

- « In a world of fake news Big Data is probably the most confusing expression »

- « Machine Learning Models are more live animal than Object »

Grégoire de Lassence
SAS Academic France

Analytics, IT and Business

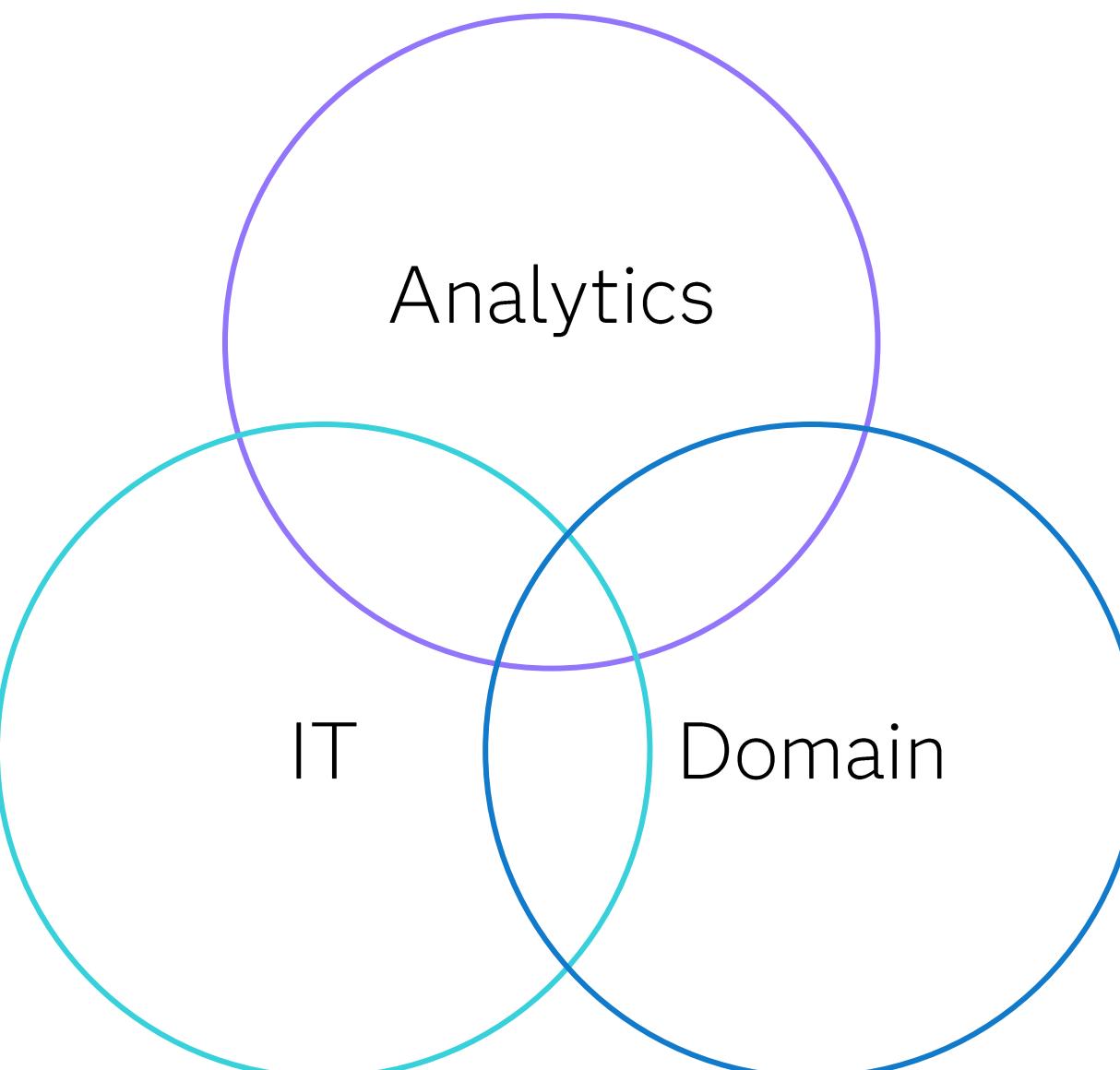


Required Expertise Working Together

A Whole Company Approach

Analytics

- Statistician
- Data Scientist
- Forecasting Modeler
- Text Miner
- Mathematician
- ...



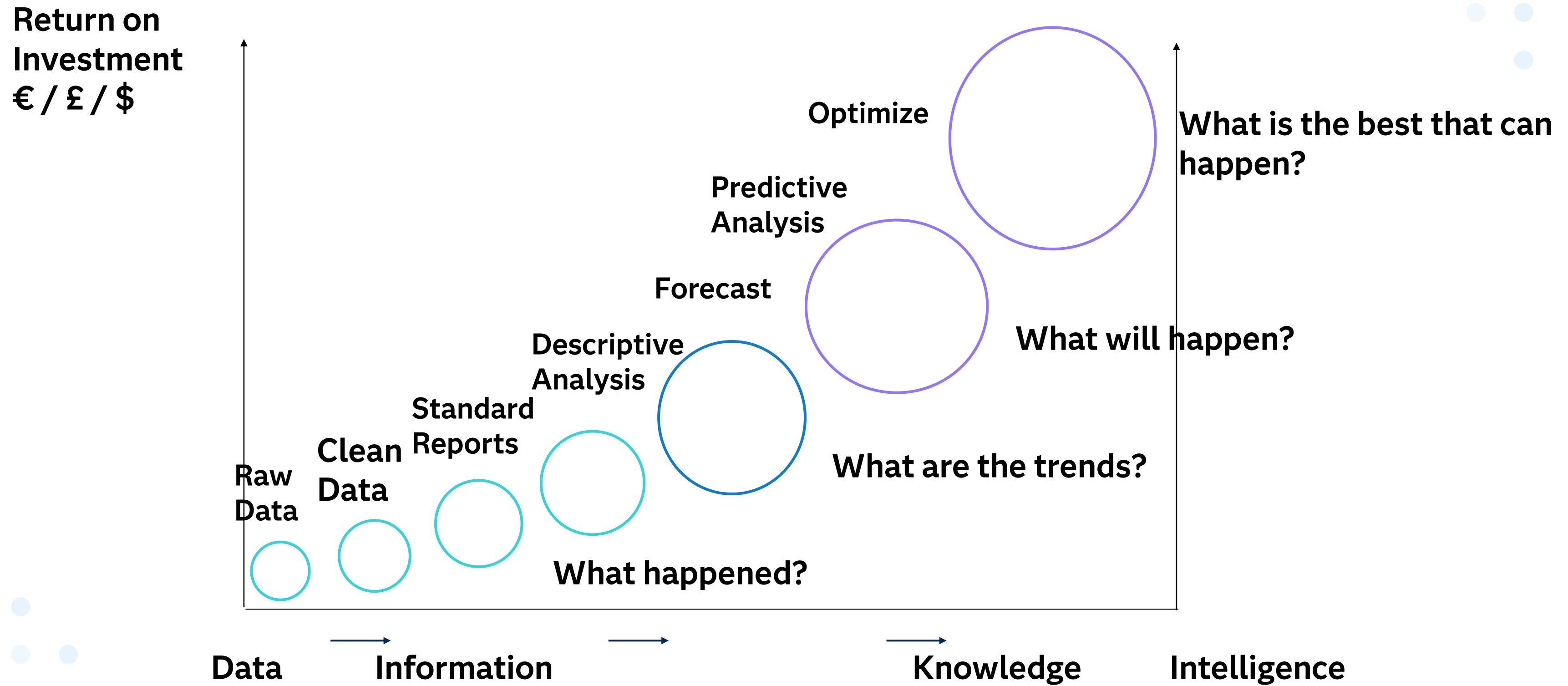
IT

- Architecture
- Security
- Data Integration
- ...

Domain

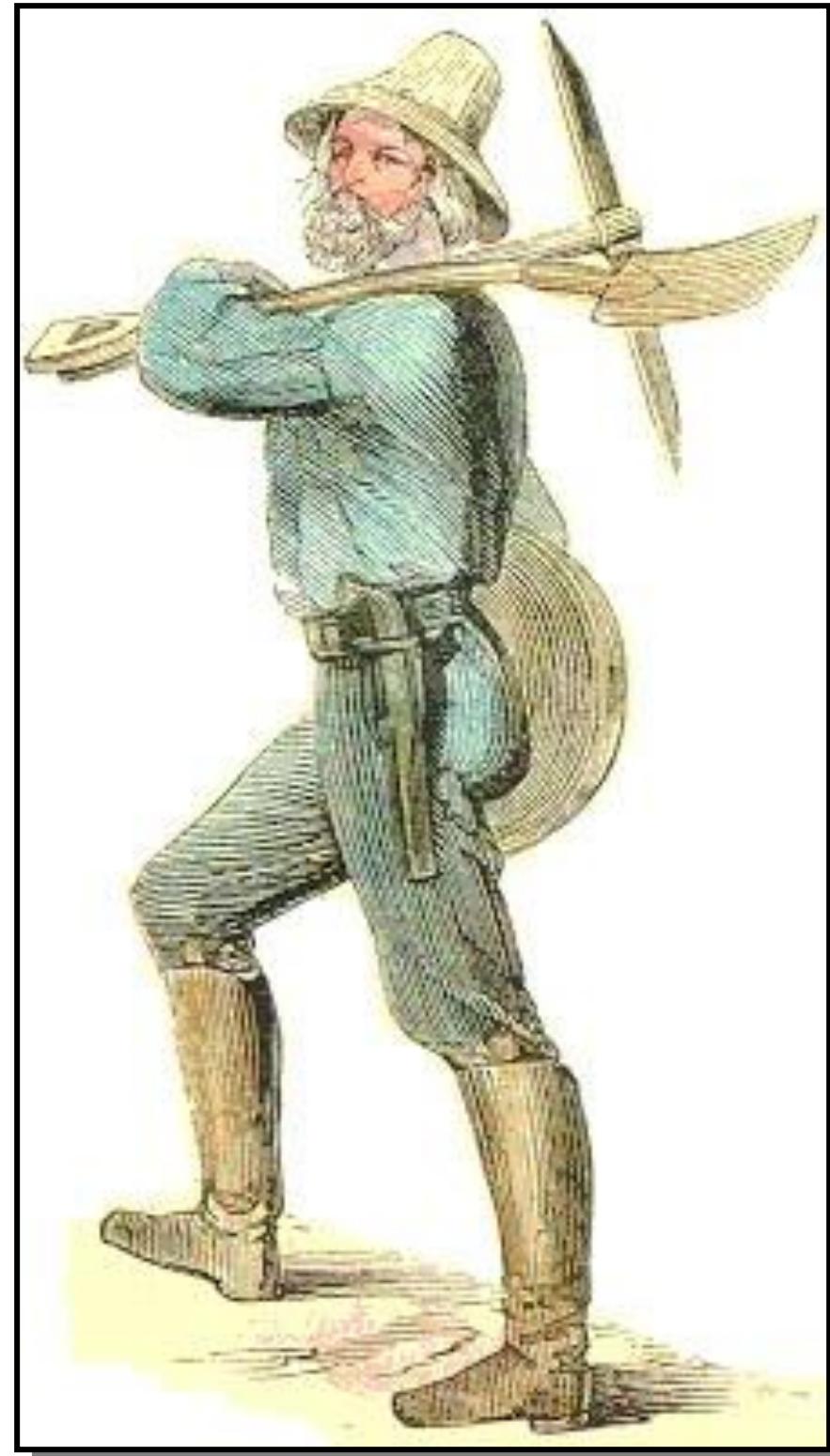
- Data Analyst
- Domain Specialist
- ...

Levels of Analytics



Required Expertise

- Domain
 - » Marketing, Fraud, Risk, etc.
- IT
 - » Data Preparation, Data Quality
 - » Hardware & Software architecture
- Analytical Methods
 - » Statistics
 - » Machine Learning
 - » Operational research, Forecasting
 - » Text mining, Deep Learning
 - » Etc.

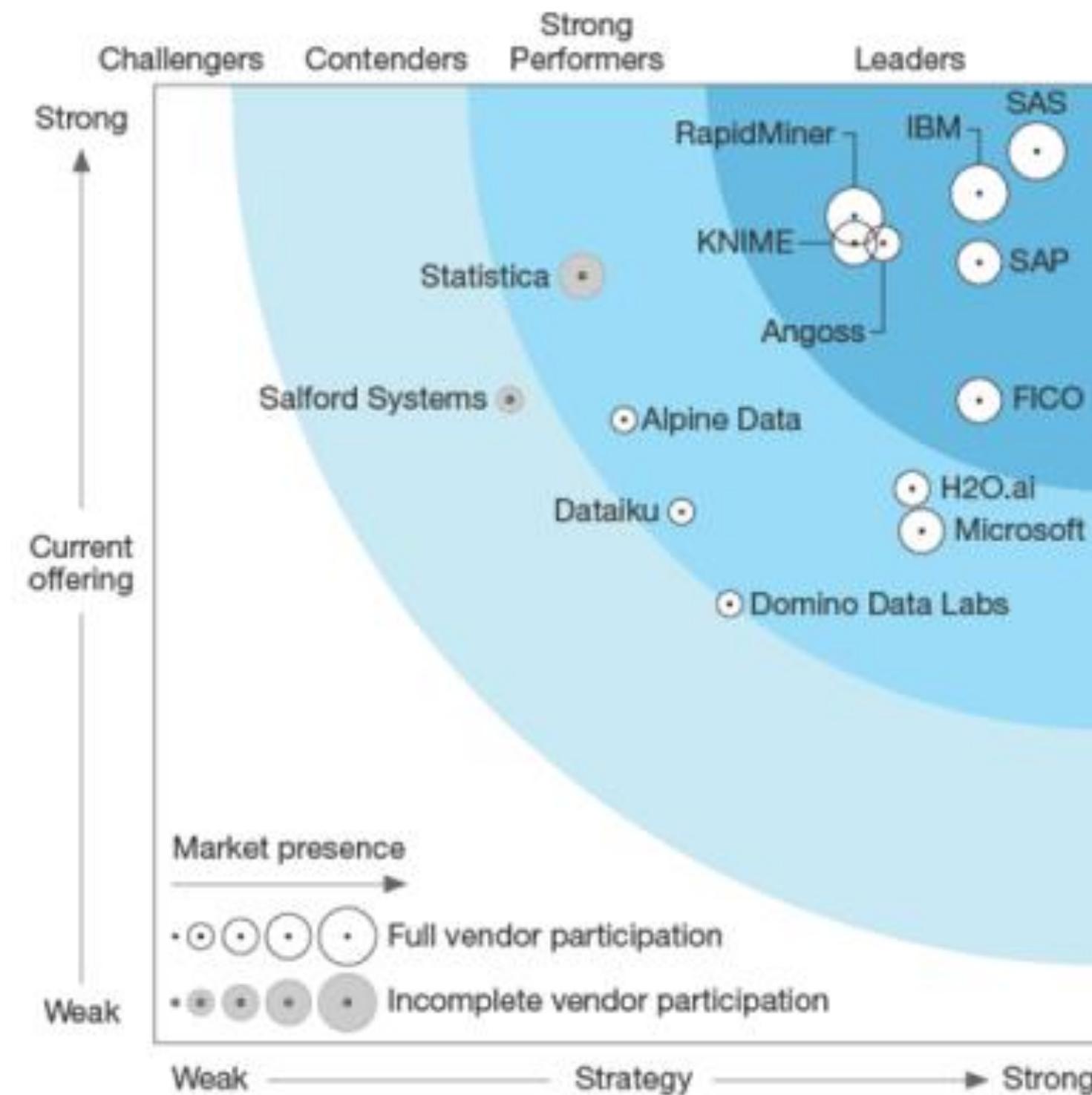


- Data + Science + Statistics + Communication + Production = Money
 - Data + Science + Statistics + Communication = Sexy Dream
 - Data + Science + Statistics = Interesting for brain, open source is enough
 - Data + Science - Statistics = Big failure
 - Data - Science - Statistics = IT
 - Big Data - Data - Science - Statistics = LOL
-
- Conclusion → Stop listening nonproduction stuff on Big Data, Begin to learn useful things

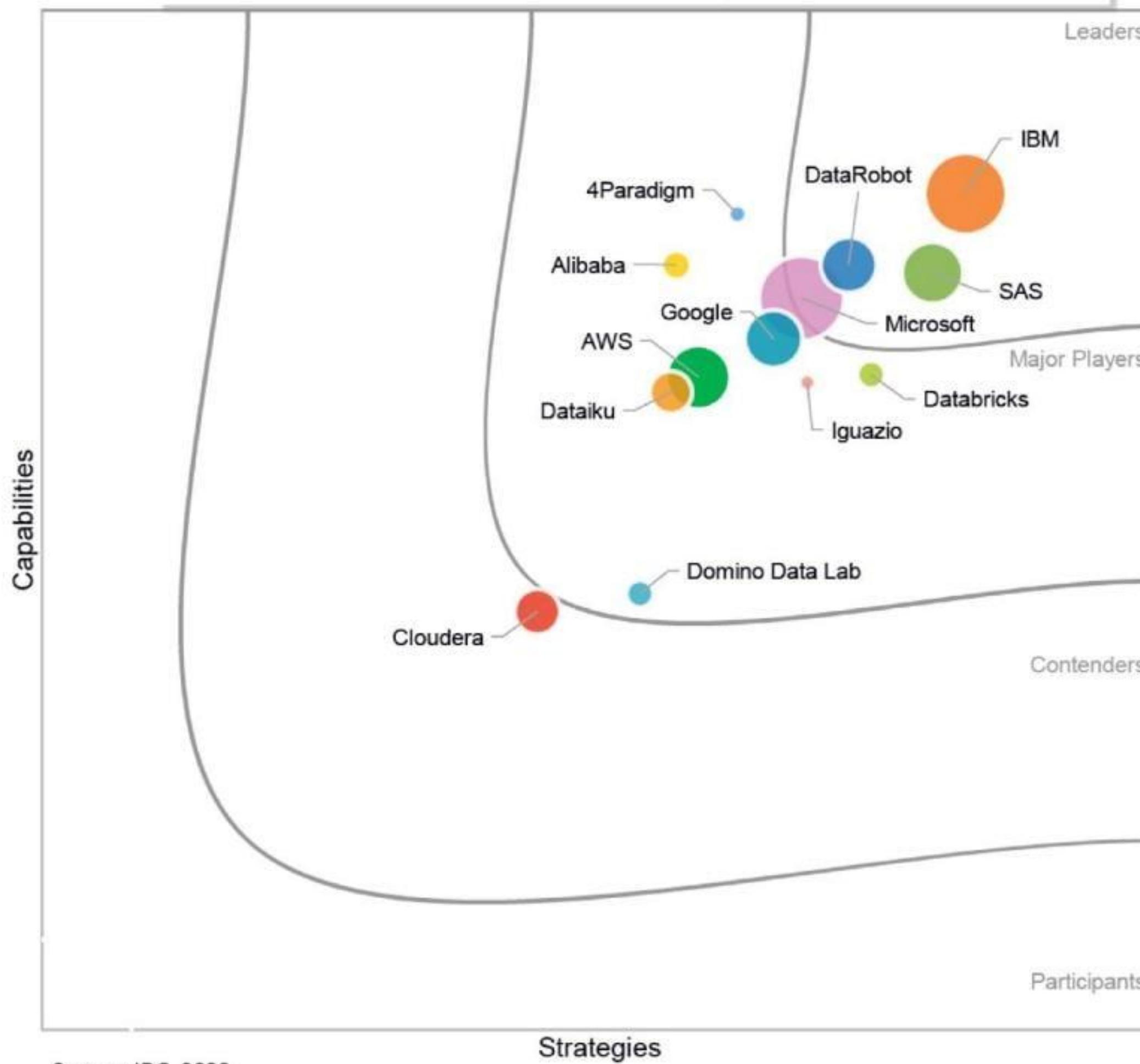
Market



Predictive Analytics and Machine Learning Solutions, Q1 2017



IDC MarketScape Worldwide Machine Learning Operations Platforms
Vendor Assessment, 2022



THE FORRESTER WAVE™

Multimodal Predictive Analytics And Machine Learning Solutions

Q3 2018



Figure 1. Magic Quadrant for Data Science Platforms



Figure 1. Magic Quadrant for Data Science and Machine-Learning Platforms



Figure 1. Magic Quadrant for Data Integration Tools



Source: Gartner (August 2016)

Figure 1. Magic Quadrant for Data Quality Tools



Figure 1. Magic Quadrant for Multichannel Marketing Hubs



Figure 1. Magic Quadrant for Digital Marketing Analytics



Machine Learning



“In the new world, it is not the big fish which eats the small fish, it’s the fast fish which eats the slow” fish.

Klaus Schwab
Founder and Executive Chairman
World Economic Forum

- Weak and Strong AI

Definition

- 2 family techniques
 - Descriptive – Unsupervised – 5%
 - **segmentation (« clustering »)**
 - Predictive – supervised – 95%
 - **Decision Trees – Forest – Gradient Boosting**
 - **Regression – Stepwise, LASSO, LARS ...**
 - **Neural Networks**
 - **SVM**
 - **Factorization Machine**
- Big data volume analysis and integration in the business process

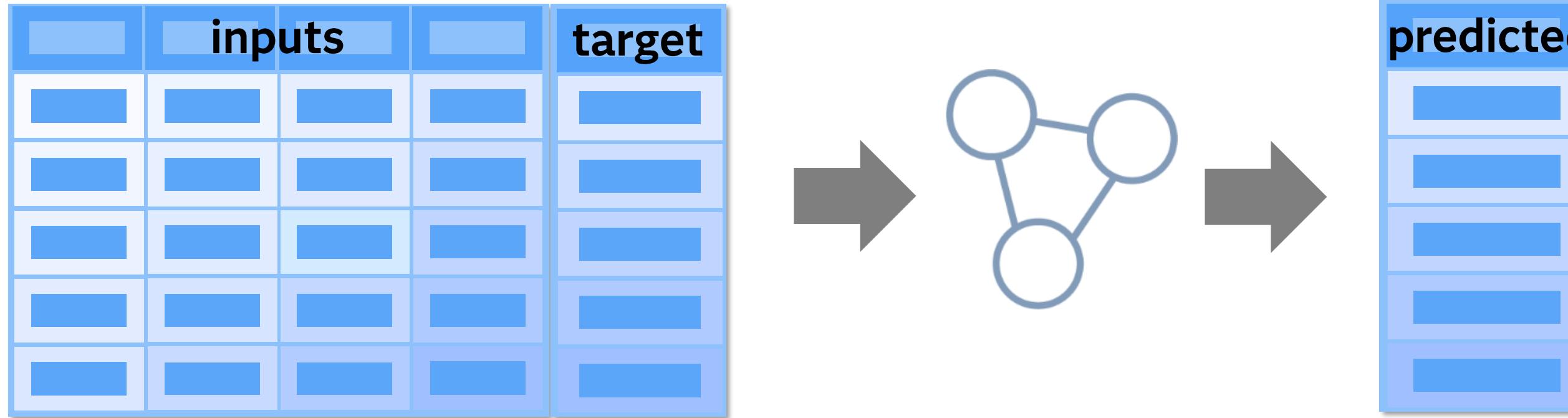
Required Expertise

- Domain
- Data
- Analytical Methods



Prediction Types

Training
Data



- decisions
- rankings
- estimates

Today



Virtual today in the past *

The target in
a virtual well
know future.
0 or 1 if the
customer buy

Variables

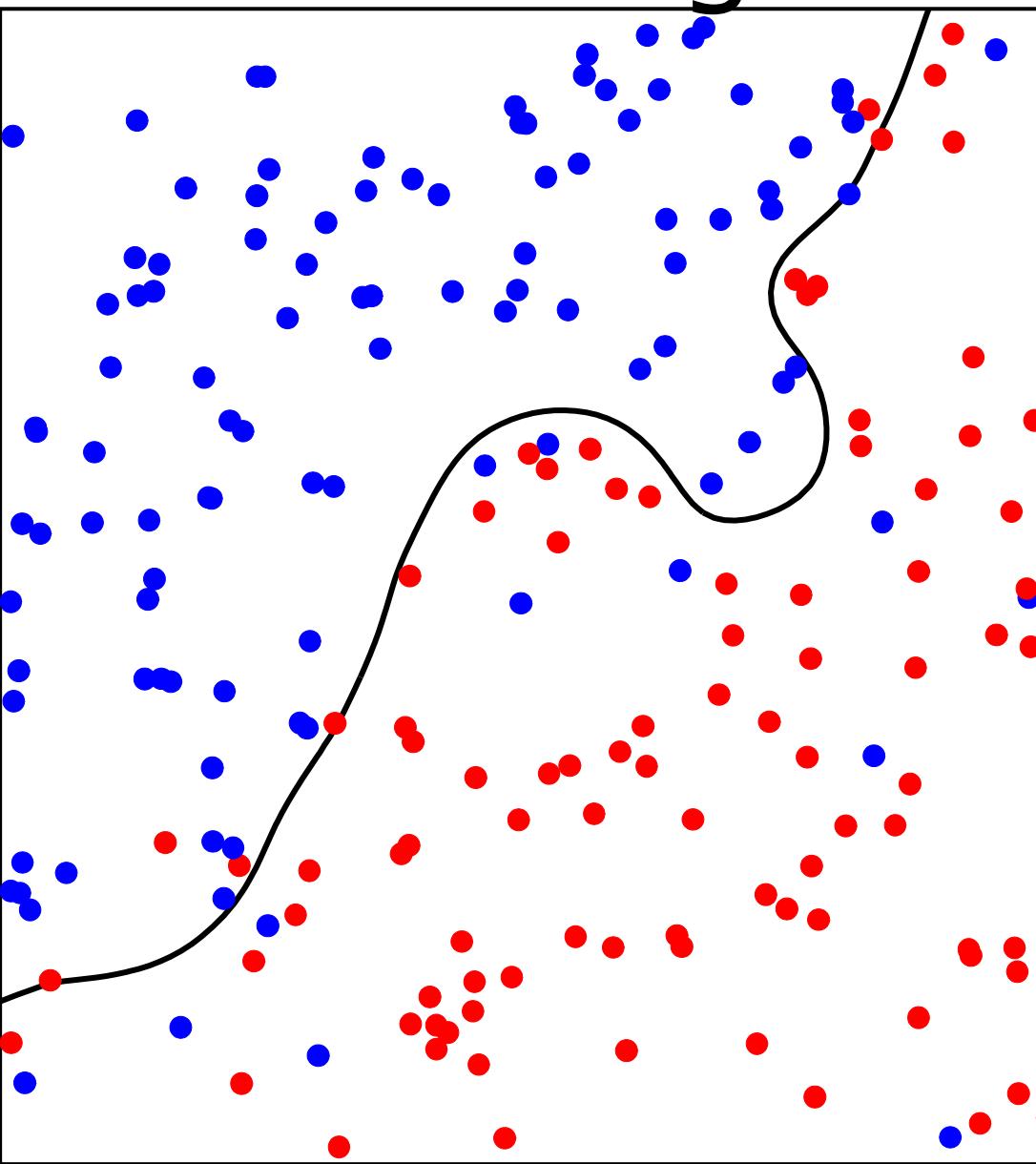
Socio-demo: Age*, Gender, Country, etc.

Behavior: R*FM, Turnover per Year, month, product group, etc.

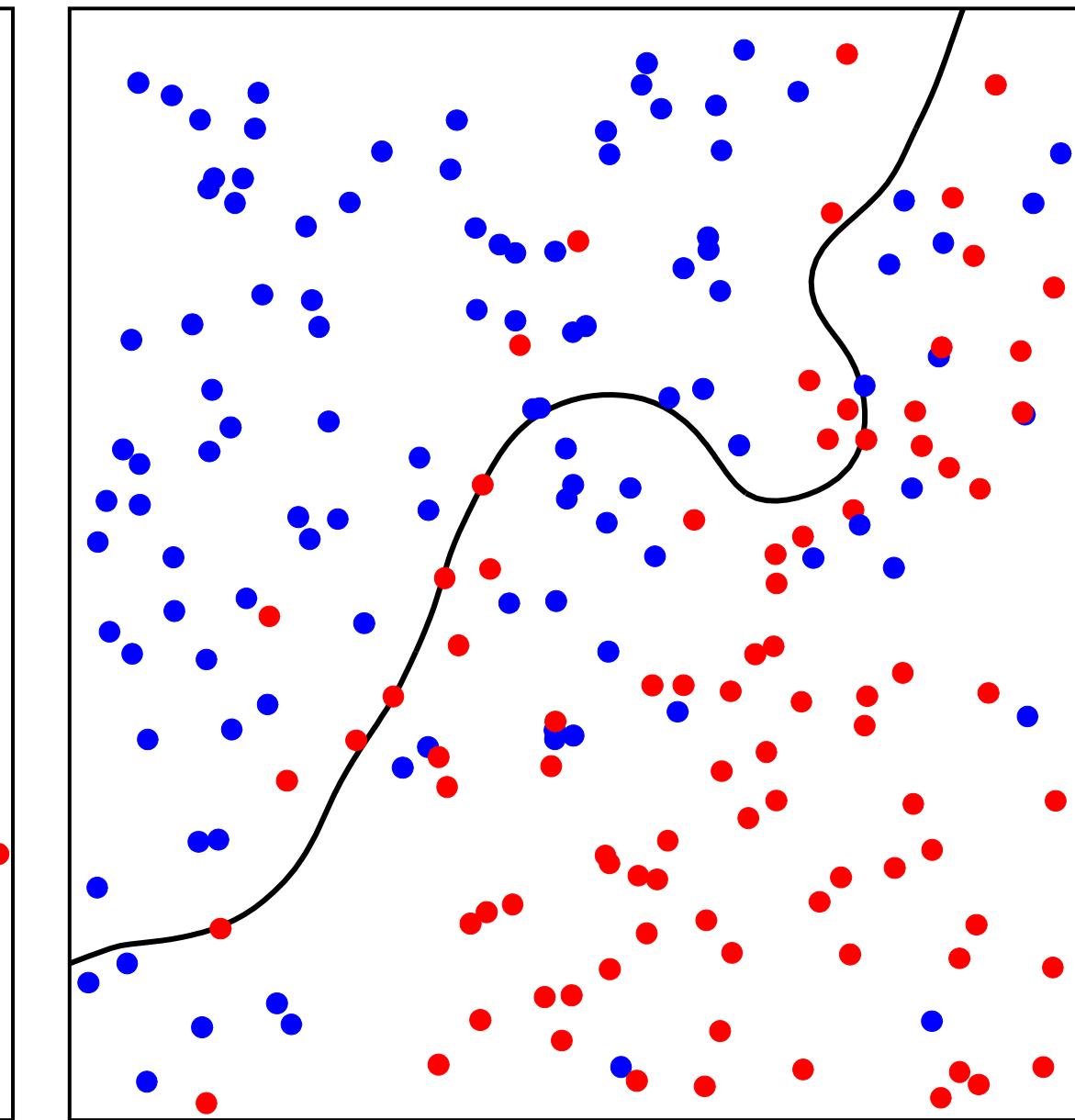


Overfitting

Training Set



Test Set

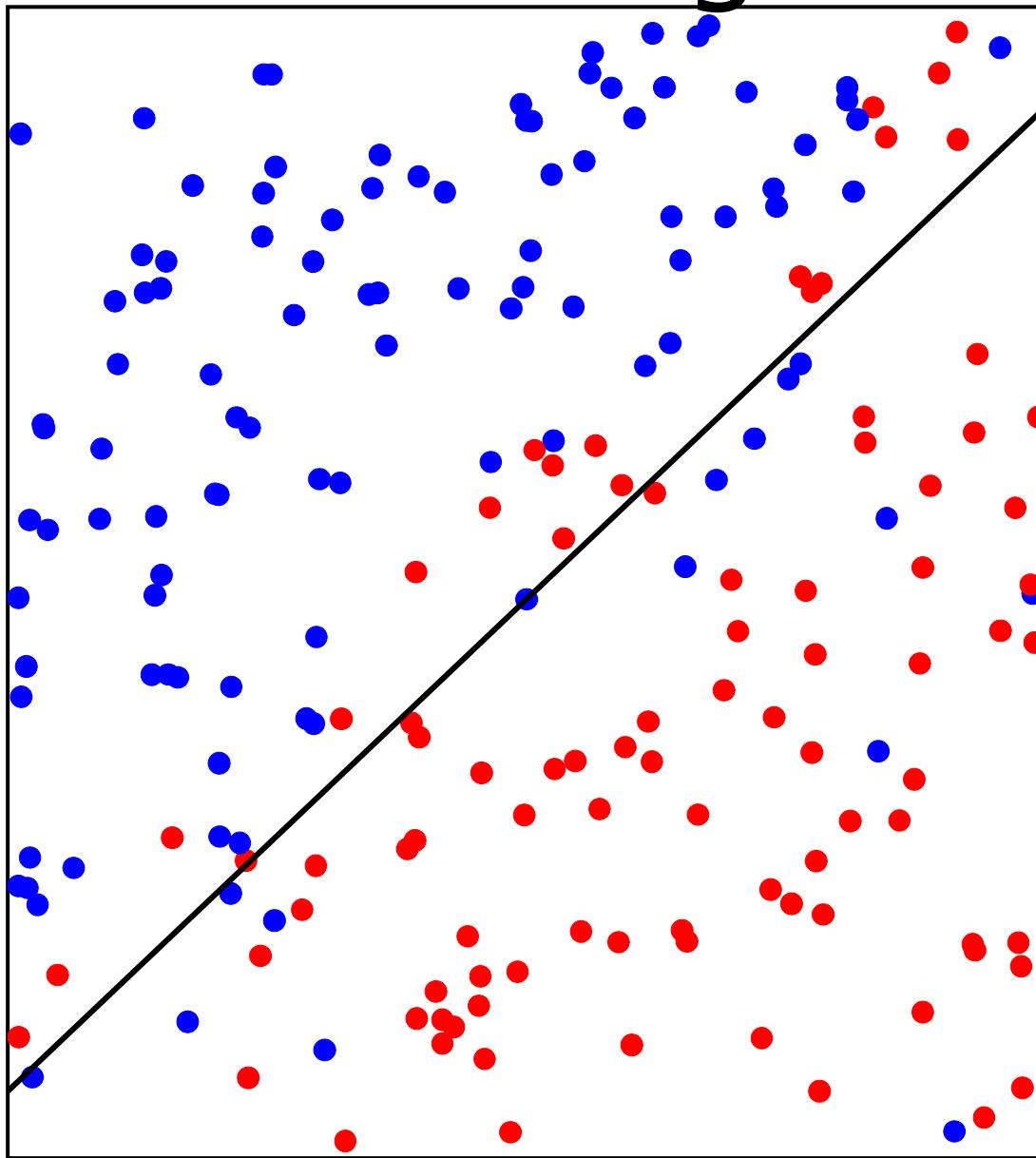


19 e = 90 %

49 e = 75 %

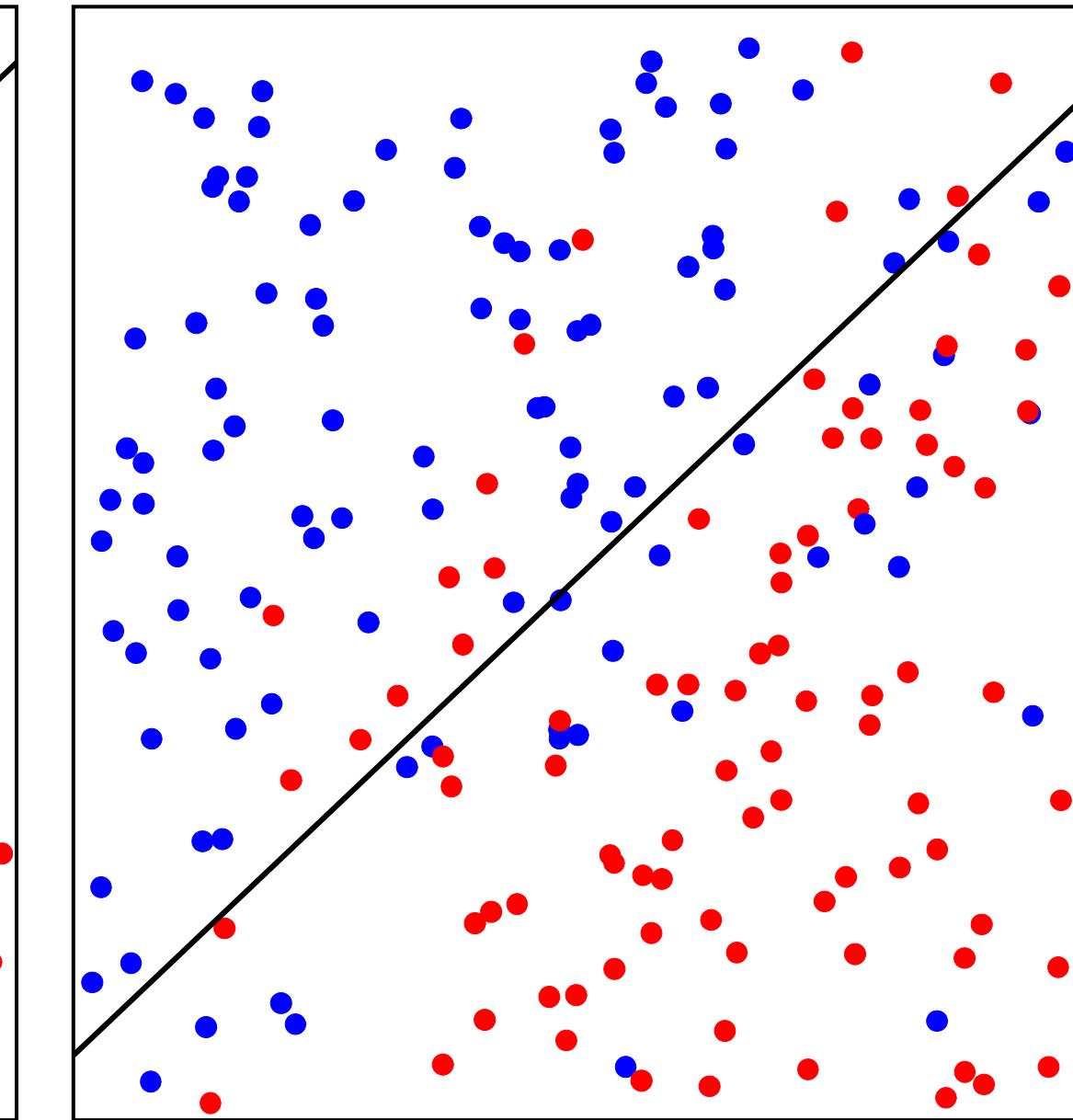
Better Fitting

Training Set



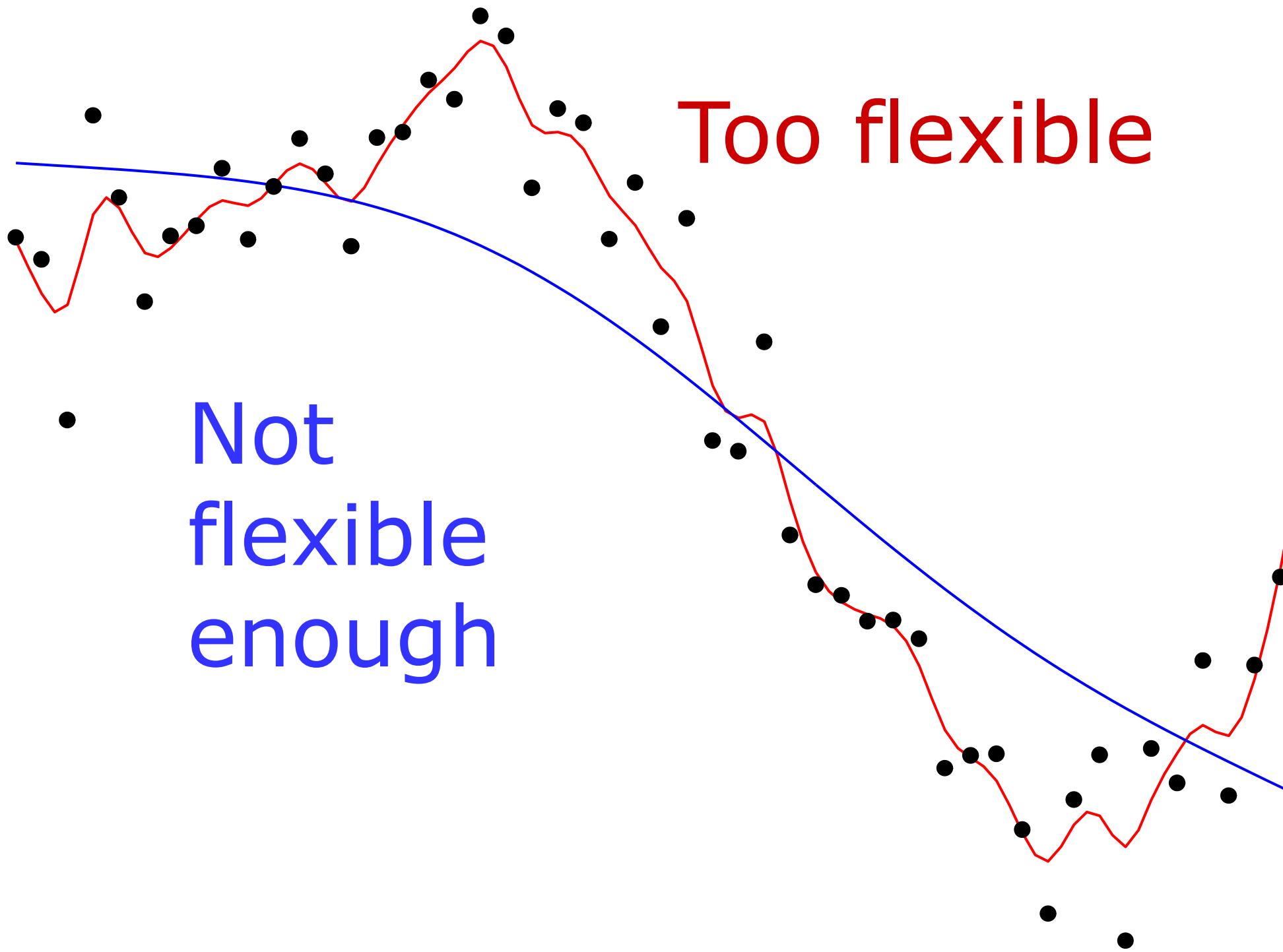
34 e = 83%

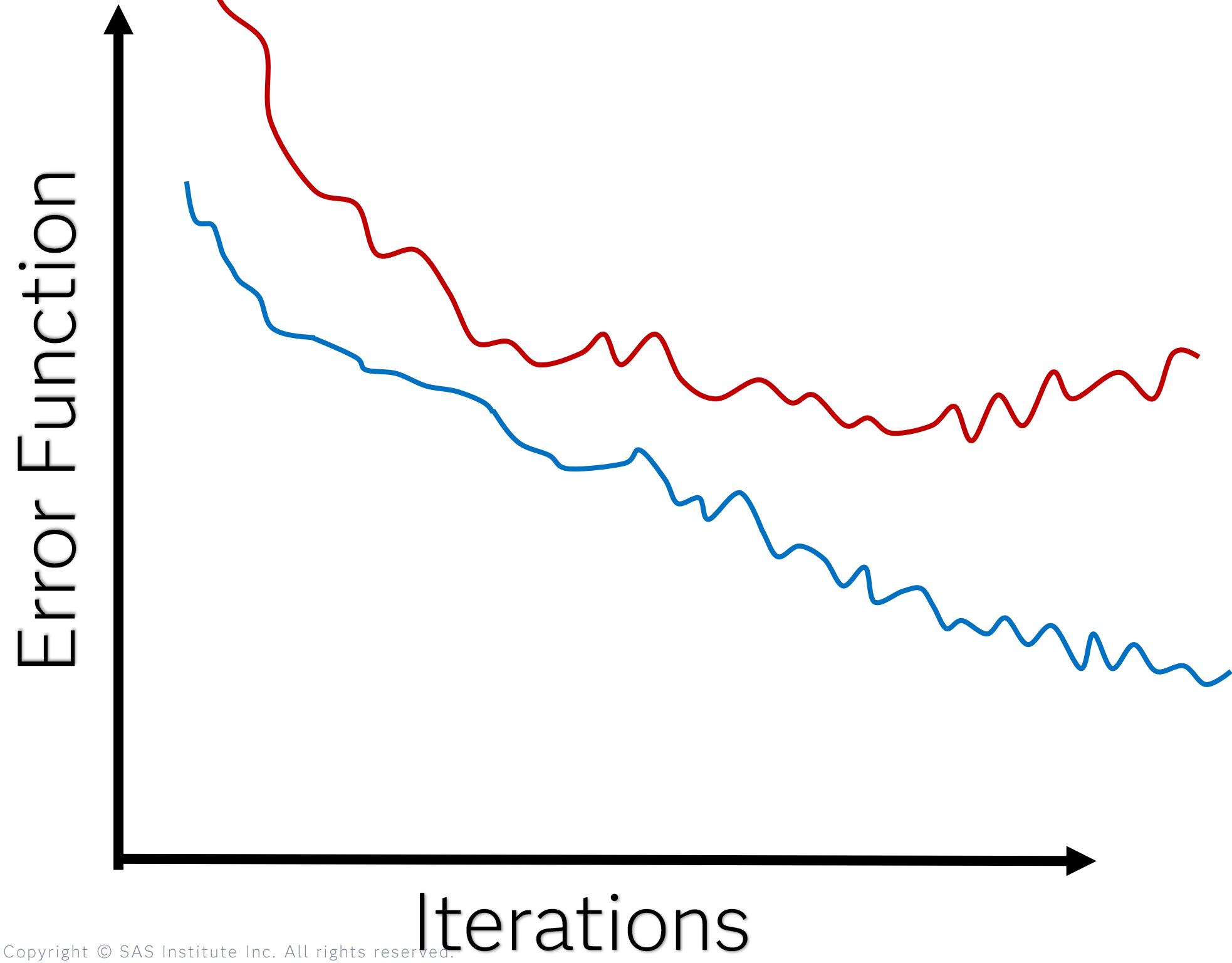
Test Set



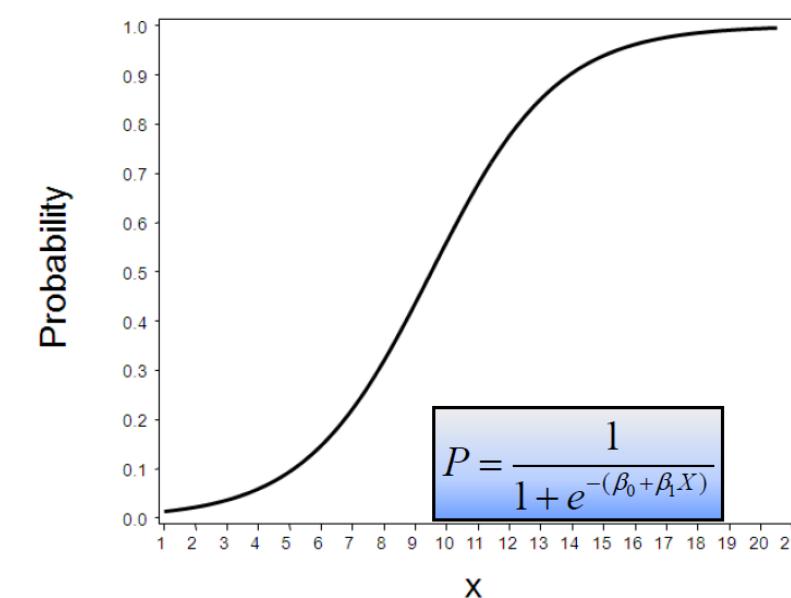
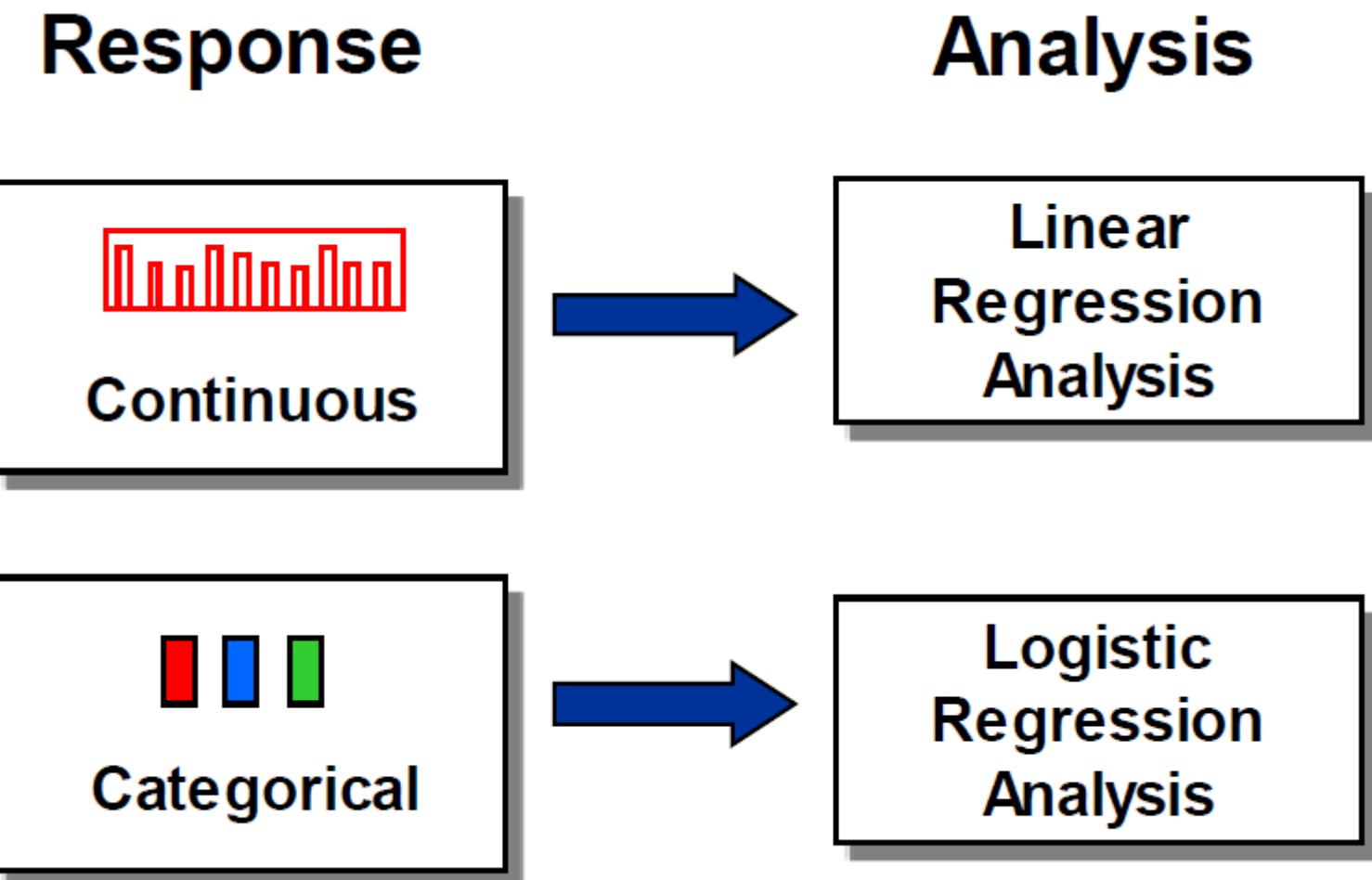
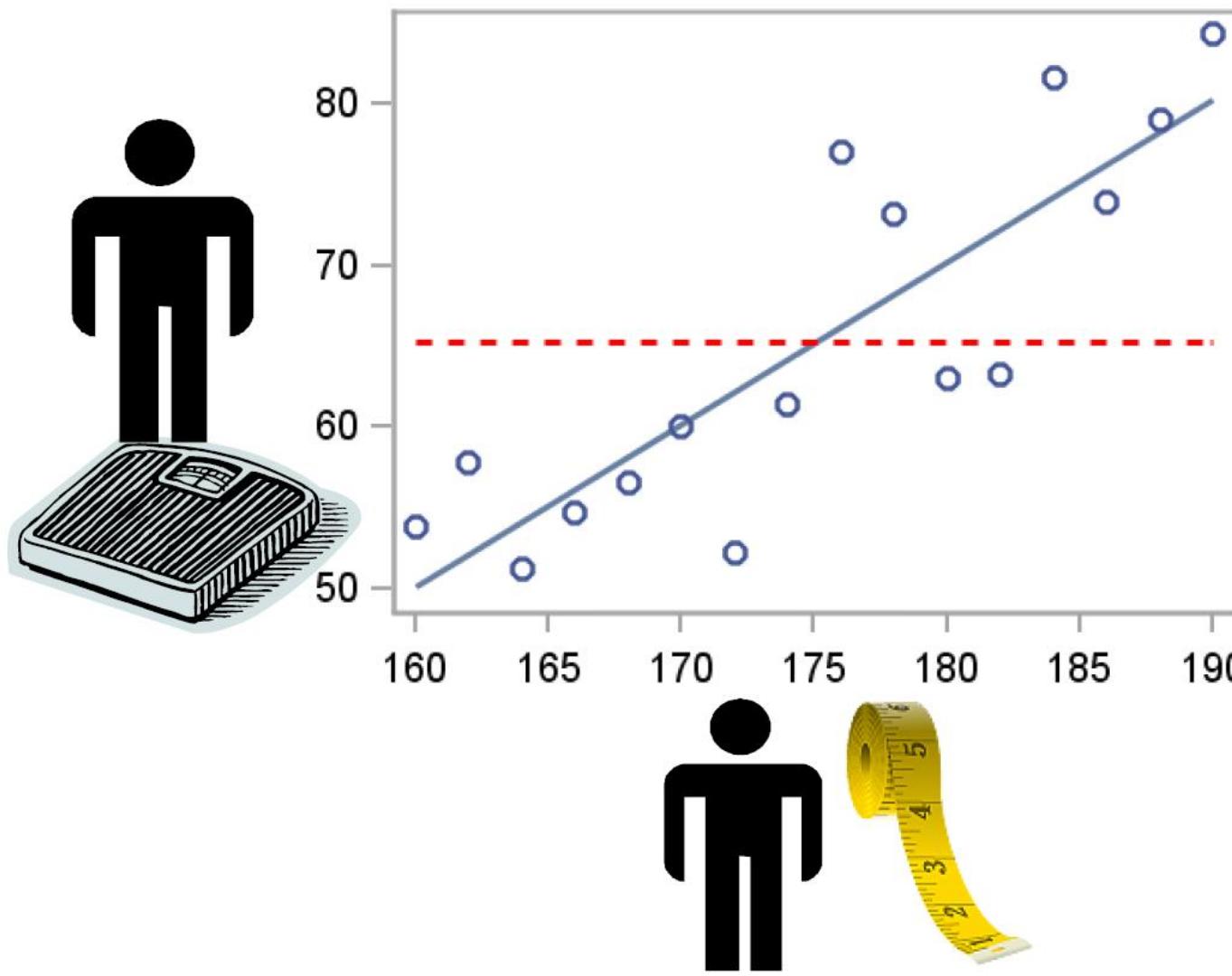
43 e = 78%

Model Complexity



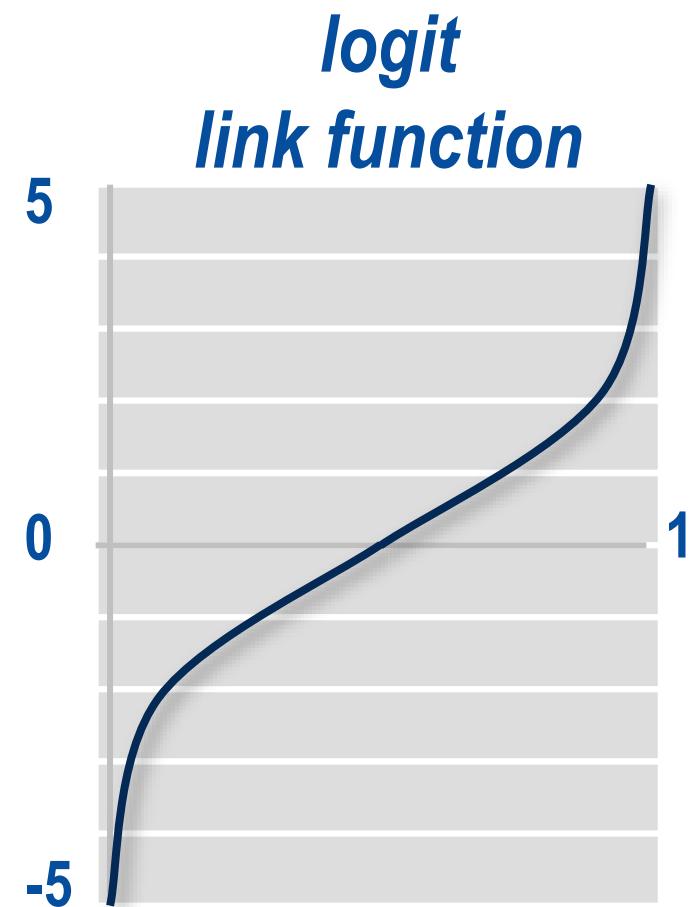


Régressions



Logistic Regression

$$\log\left(\frac{\hat{p}}{1 - \hat{p}}\right) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2 \quad \textit{logit scores}$$



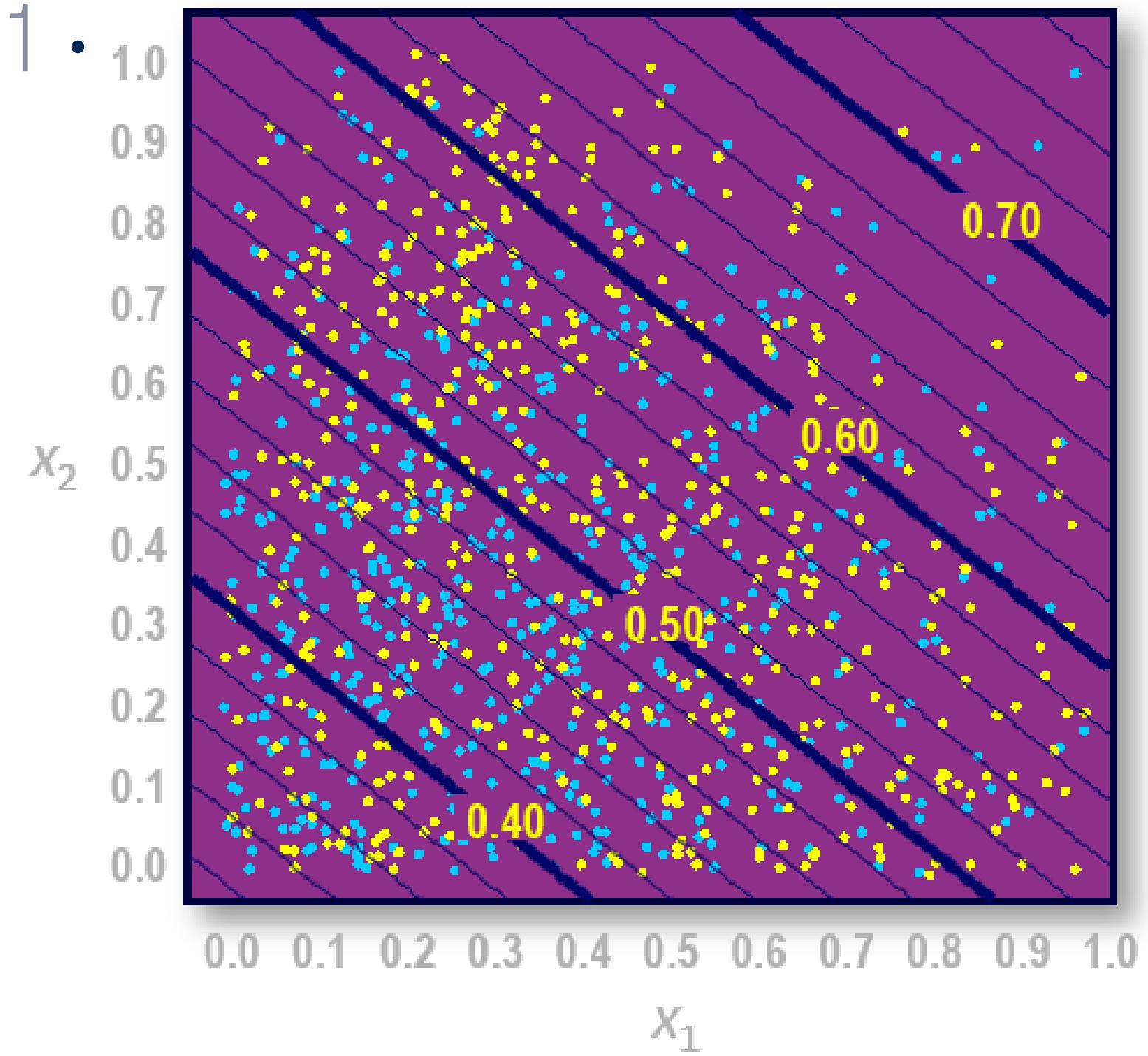
The logit link function transforms probabilities (between 0 and 1) to logit scores (between $-\infty$ and $+\infty$).

Logistic Regression Example

$$\text{logit}(\hat{p}) = -0.81 + 0.92 \cdot x_1 + 1.11 \cdot x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Using the maximum likelihood estimates, the prediction formula assigns a logit score to each x_1 and x_2 .



Forêt

ID	E	E	C
1			
2			
3			
4			
5			
6			

ID	E	E	C
1			
4			
6			
1			
1			
6			

ID	E	E	C
1			
2			
3			
4			
5			
6			

ID	E	E	C
2			
3			
5			
5			
5			
4			

ID	E	E	C
2			
3			
5			
5			
5			
4			

ID	E	E	C
6			
6			
2			
1			
3			
3			

ID	E	E	C
2			
3			
5			
5			
5			
4			

Tree 1

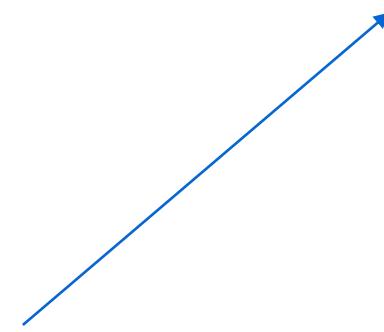
Tree 2

Tree n

μ

Boosting

ID	E	C	P
1			1/N
2			1/N
3			1/N
4			1/N
5			1/N
6			1/N

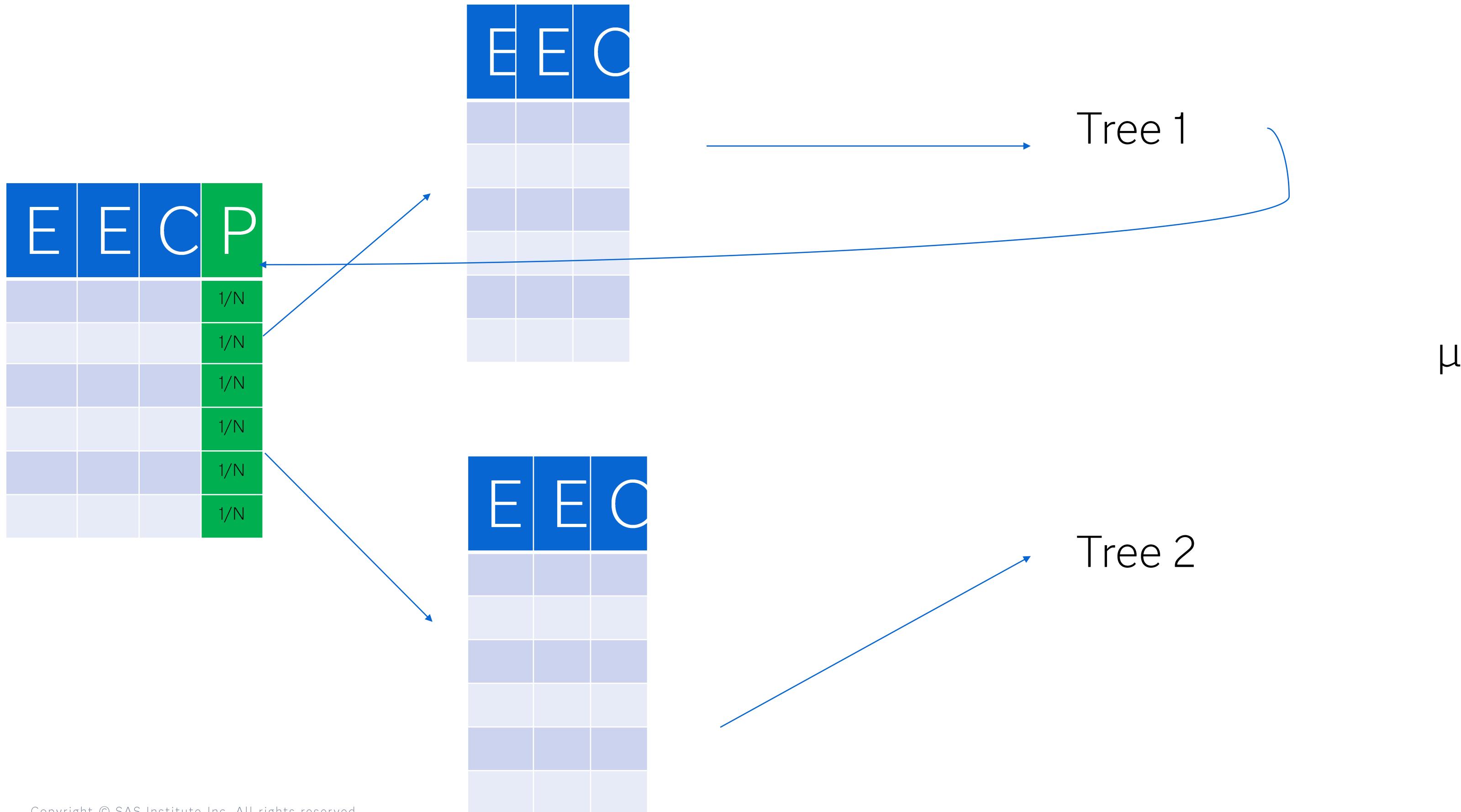


ID	E	C
1		
4		
6		
1		
1		
6		



Tree 1

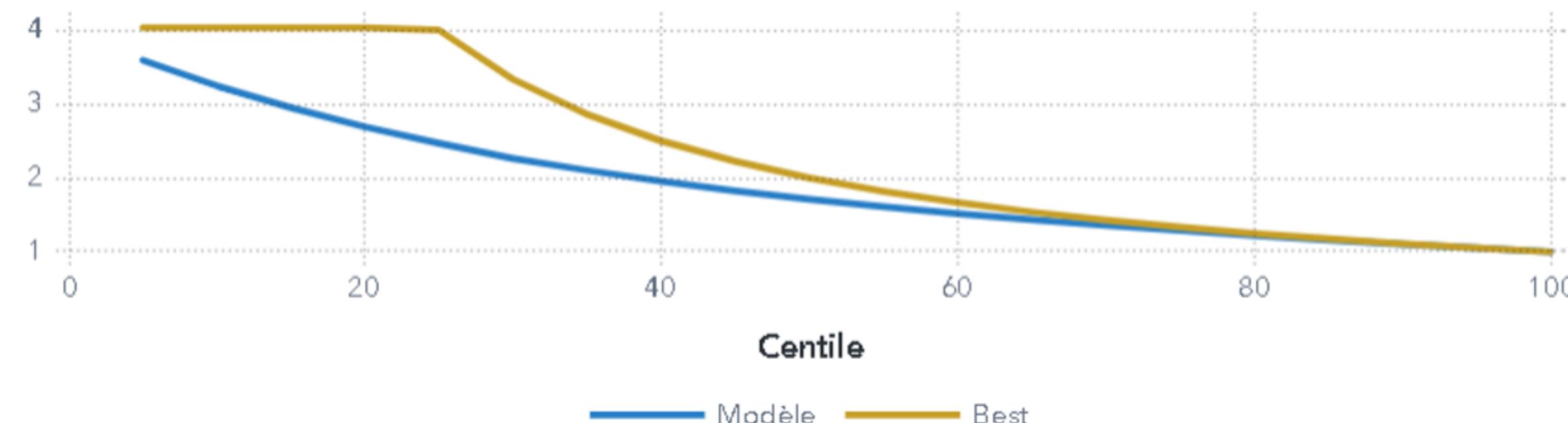
Boosting



Lift

Lift

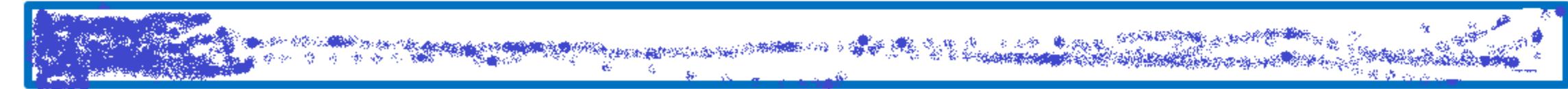
Lift cumulé



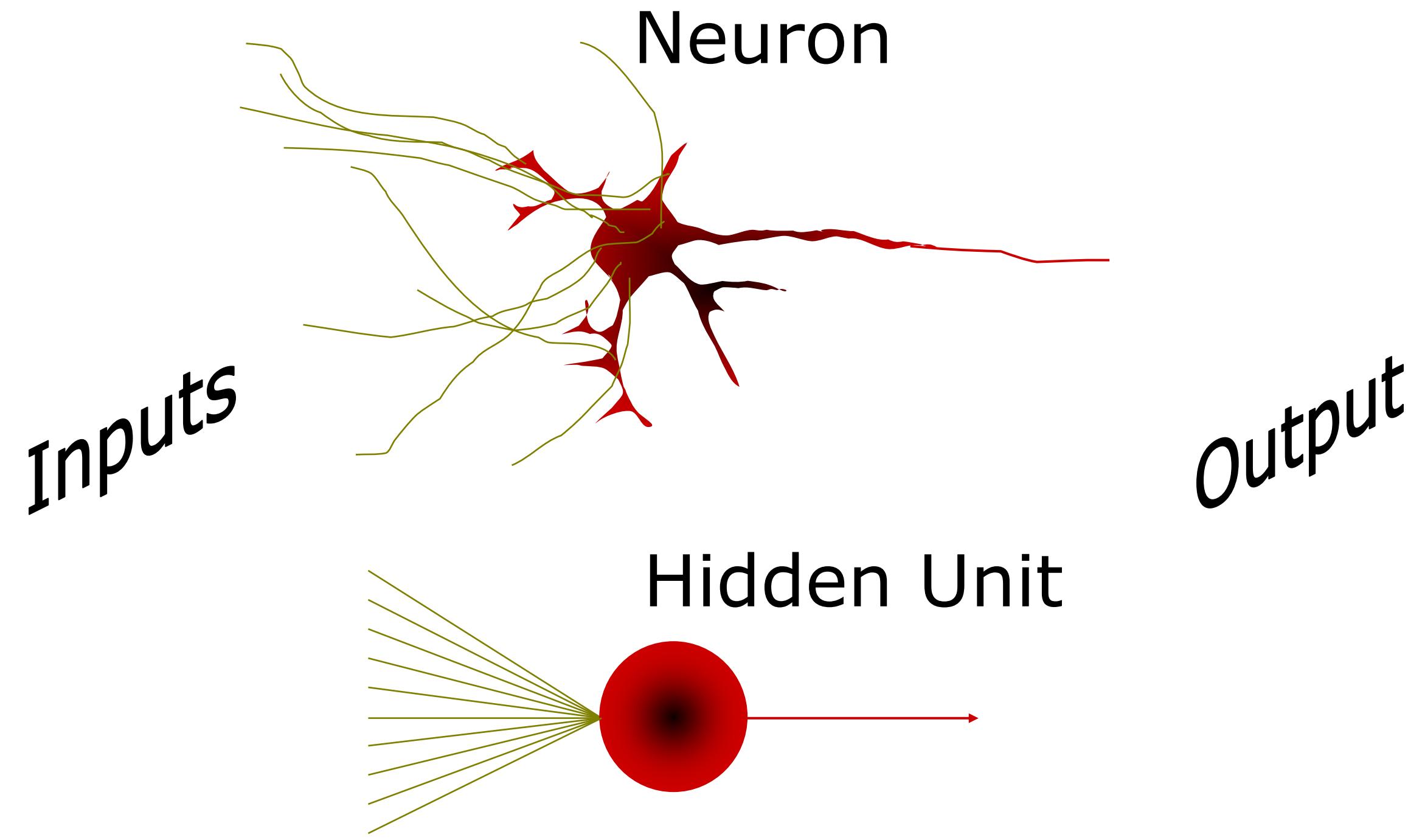
Best

Model

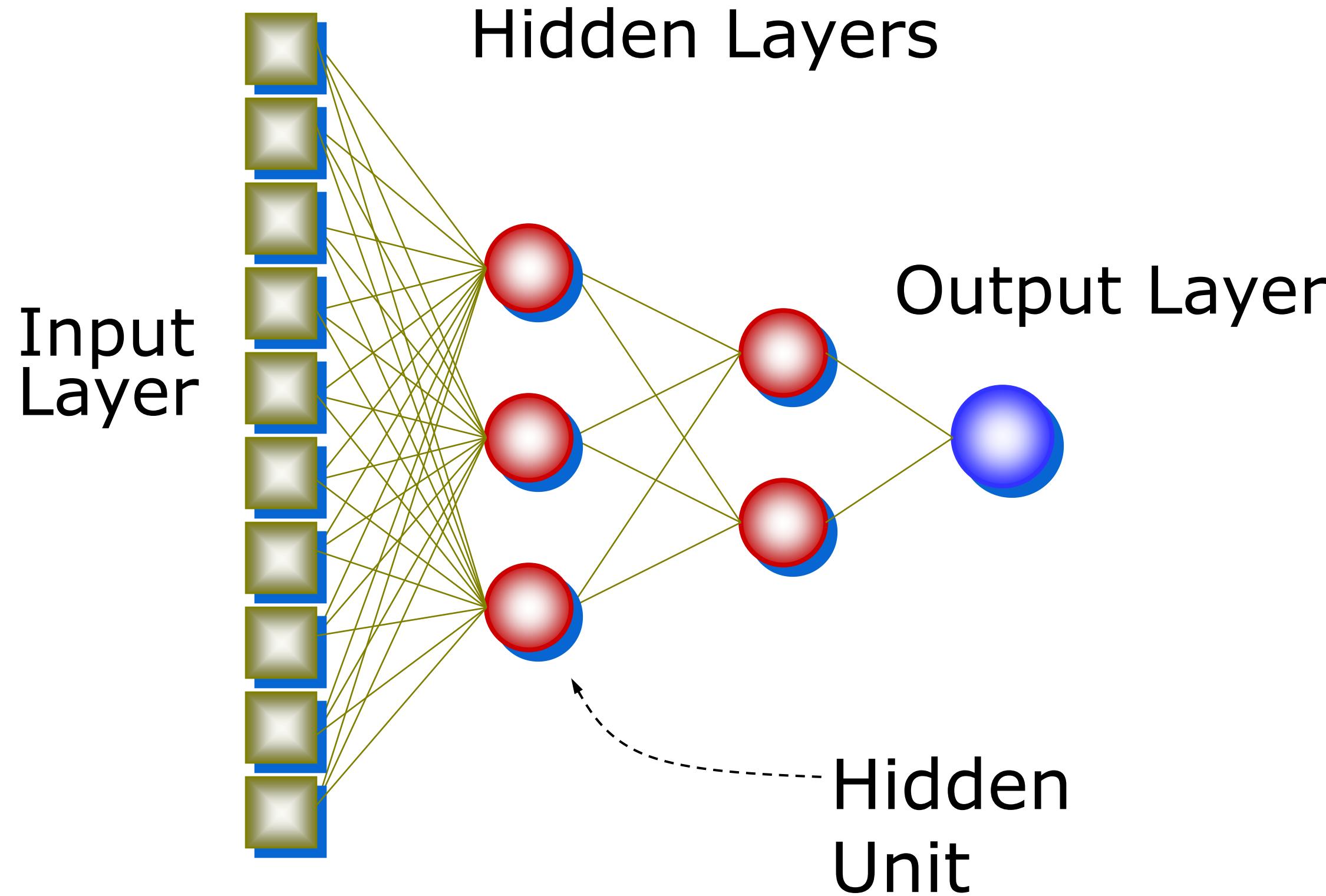
Random



Artificial Neural Networks



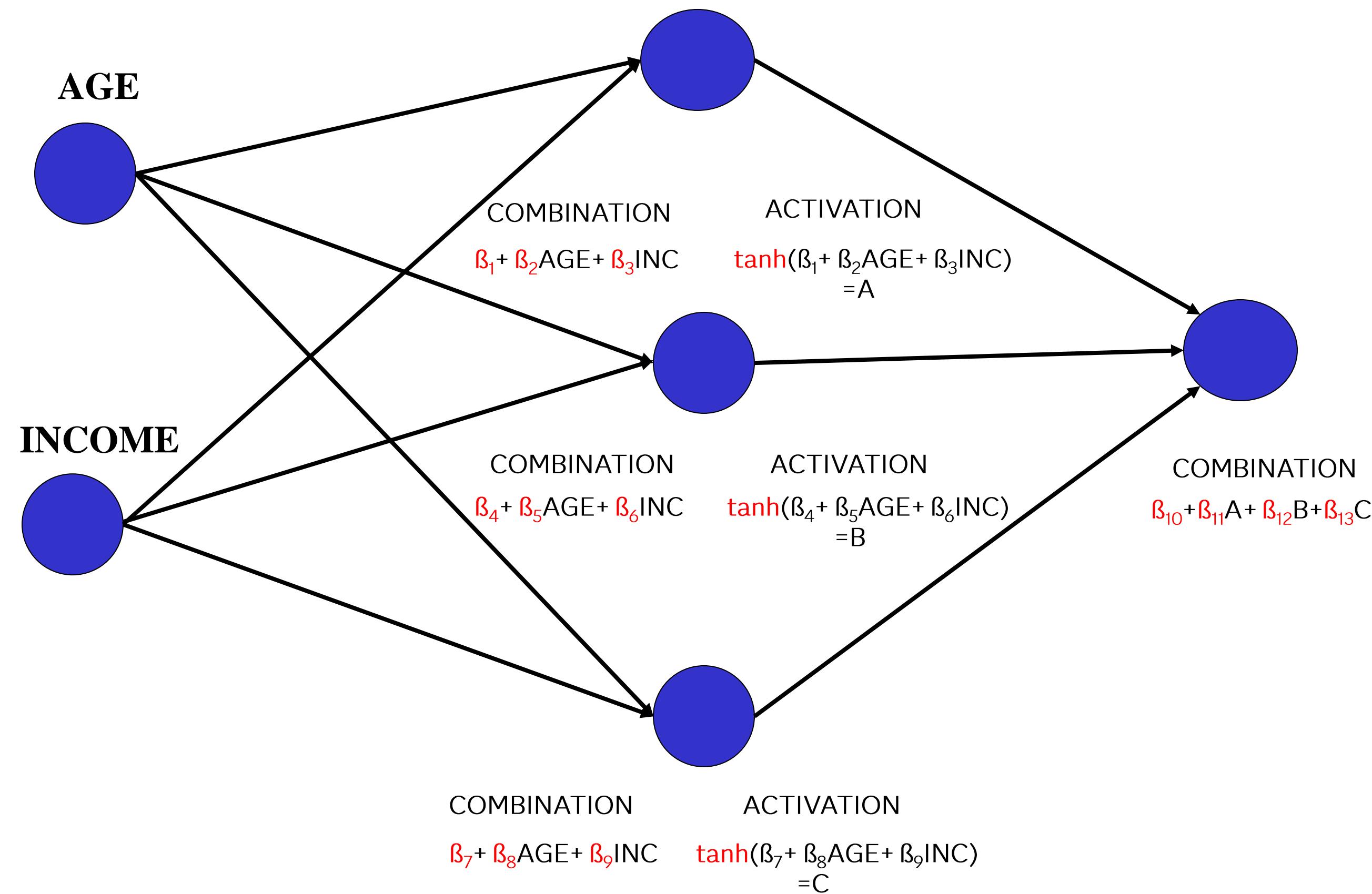
Multilayer Perceptron



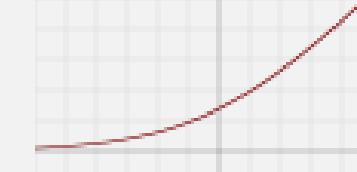
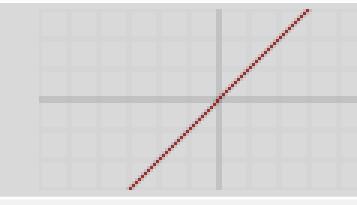
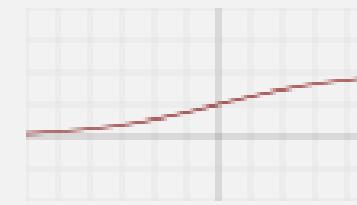
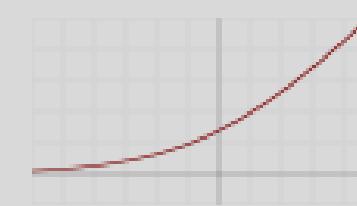
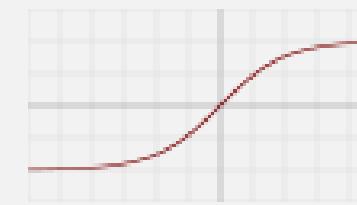
INPUT

HIDDEN

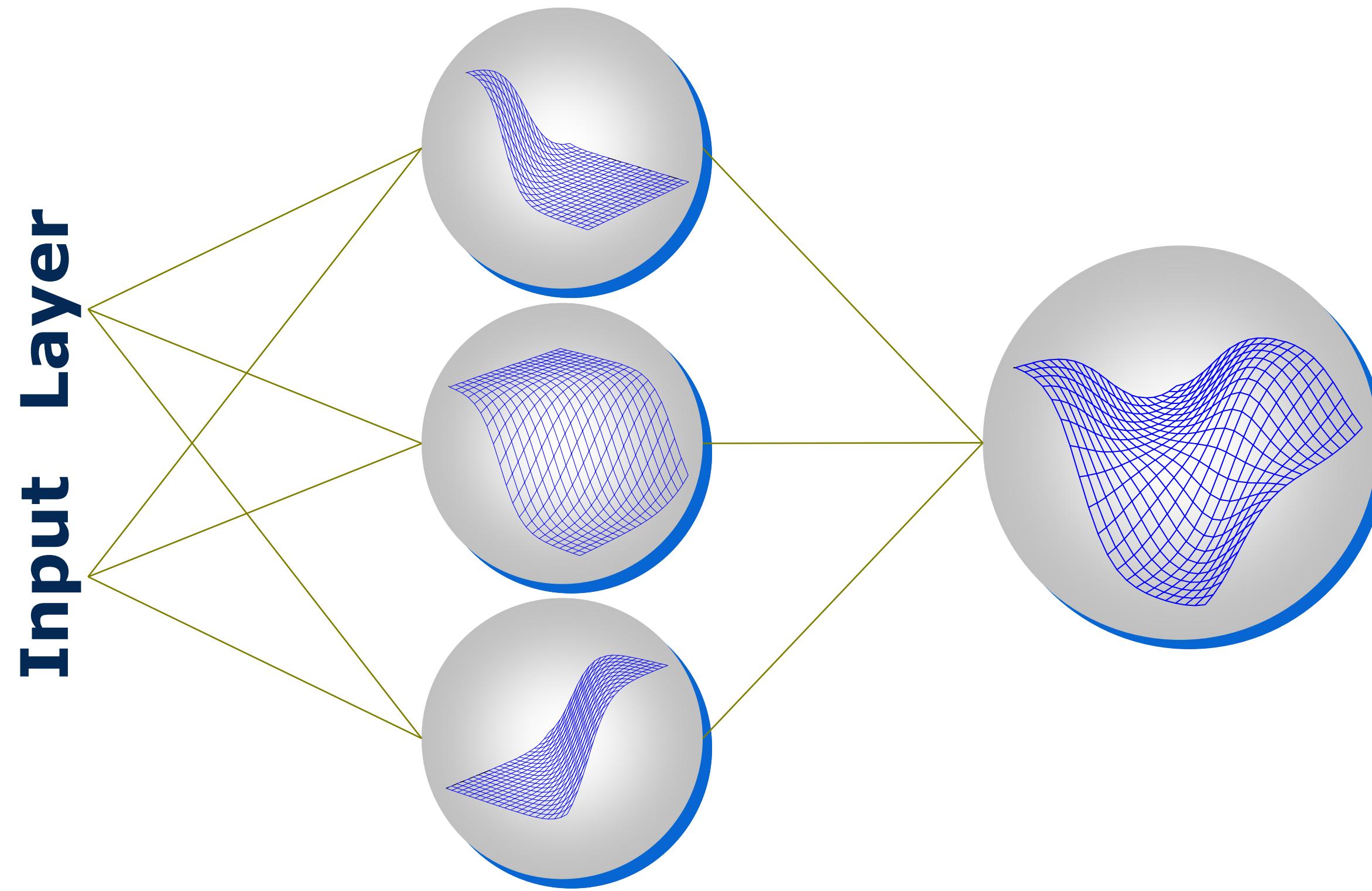
OUTPUT



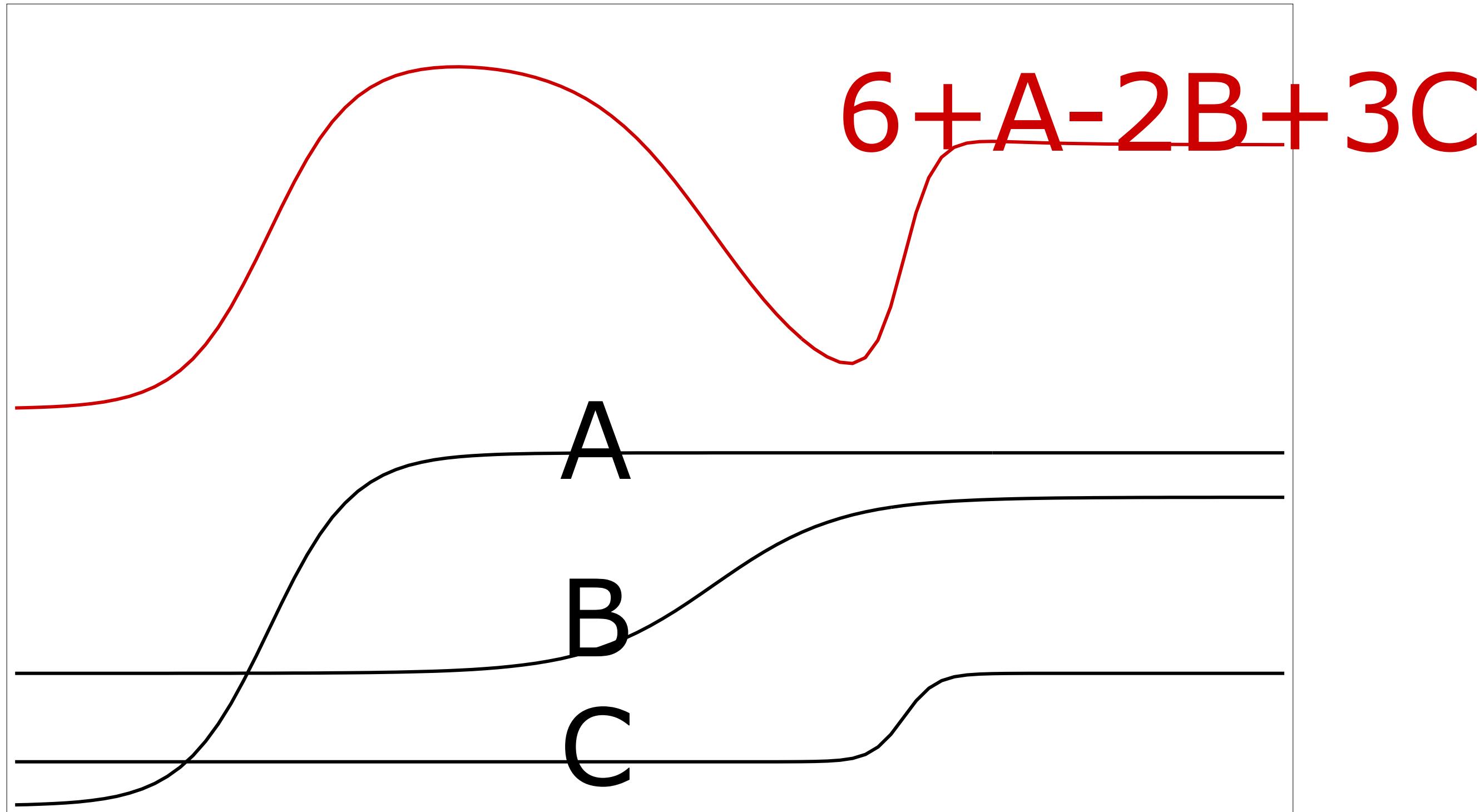
Activation Functions

Function	Plot	Equation	Range
Exponential		$f(x) = e^x$	$[0, \infty)$
Identity		$f(x) = x$	$(-\infty, \infty)$
Logistic		$f(x) = \frac{1}{1 + e^{-x}}$	$(0,1)$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$[0, \infty)$
Sine		$f(x) = \sin(x)$	$[-1,1]$
Softplus		$f(x) = \ln(1 + e^x)$	$[0, \infty)$
Hyperbolic Tangent (Tanh)		$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$	$(-1,1)$

Activation Function



Universal Approximator



Incremental Response Analysis

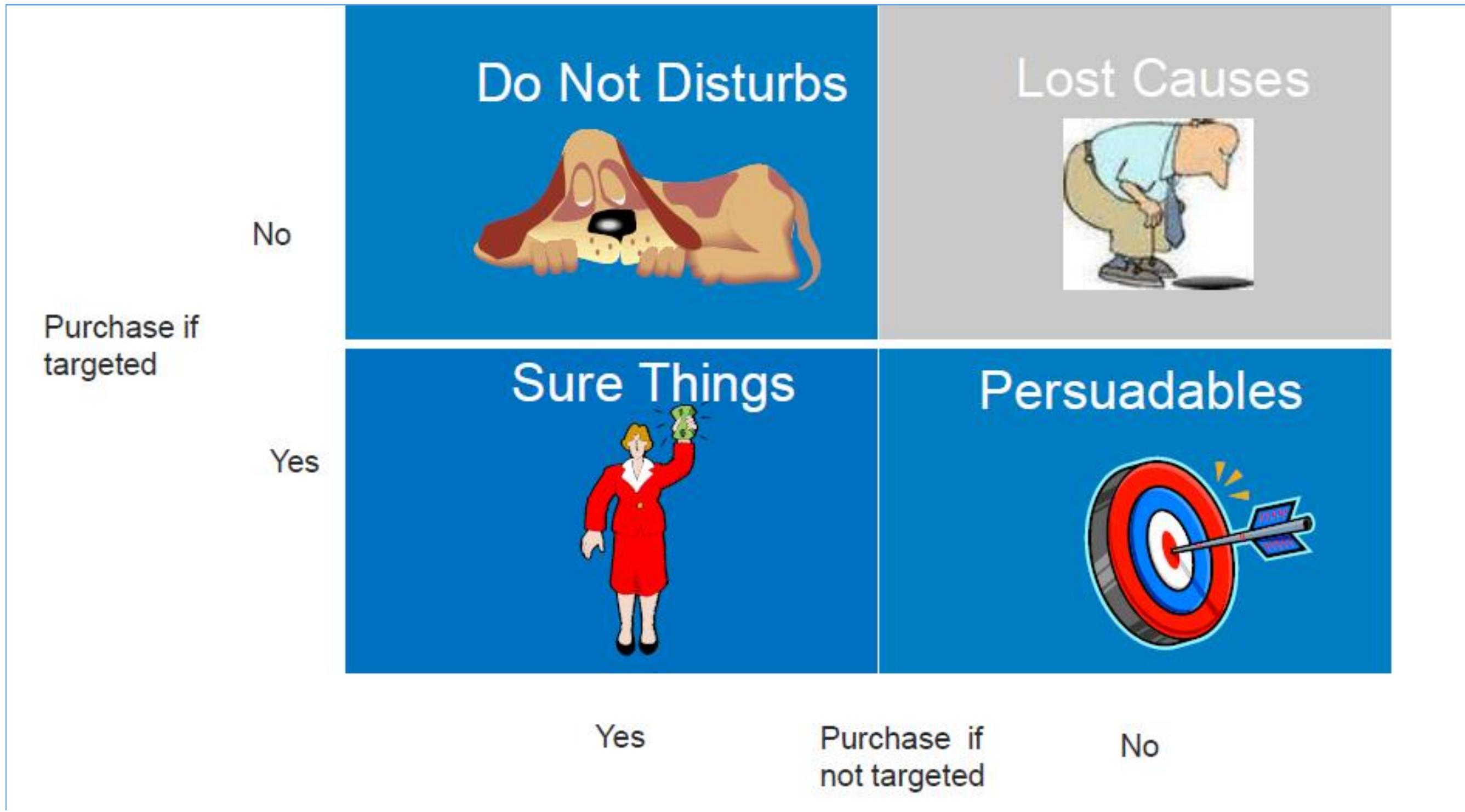
“I know half the money I spend on advertising is wasted, but I can never find out which half.”

— John Wanamaker

→ Which customers should be selected for contact/promotion so that the campaign/marketing can achieve the maximum net profit?

Customer Classes

→ Maximize impact by targeting the persuadables



Treatment Group (Promotion)

Incremental Response

Response = No

Control Group (No Promotion)

Response = Yes

Response = No

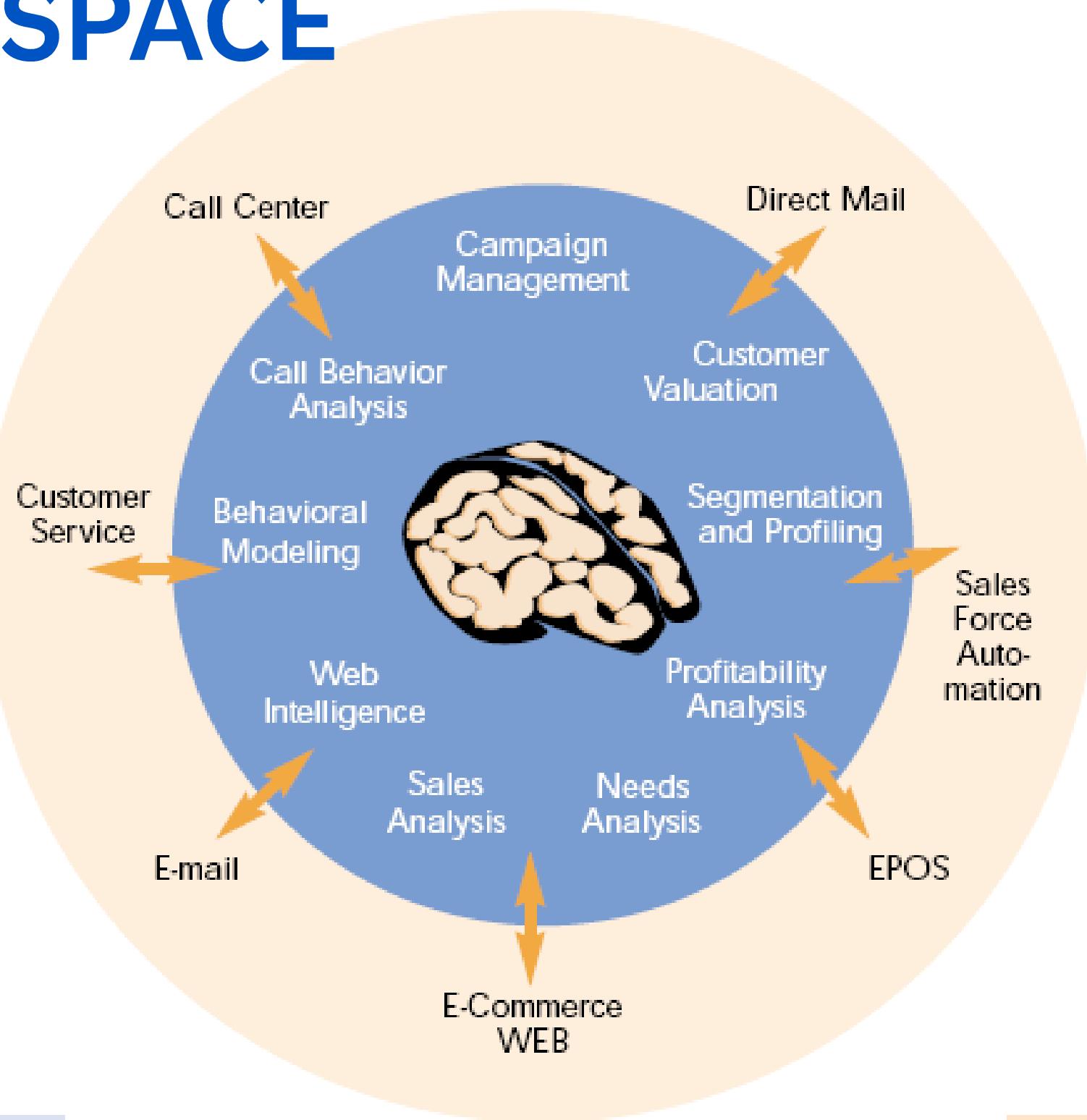
- **Treatment Group:** Assume incremental response exists, but do not know which observations have.
- **Control Group:** No incremental response (known).

Marketing evolution

Mass Marketing	Market segmentation	Relationship marketing
Product focused Anonymous Few campaigns Wide reach Little or no research Short term	Group focused General category profiles More campaigns Smaller reach Based on segment analysis of demographics Short term	Customer focused Targeted to individuals Many campaigns Discrete reach Based on retailed customer behavior and profiles Long term value

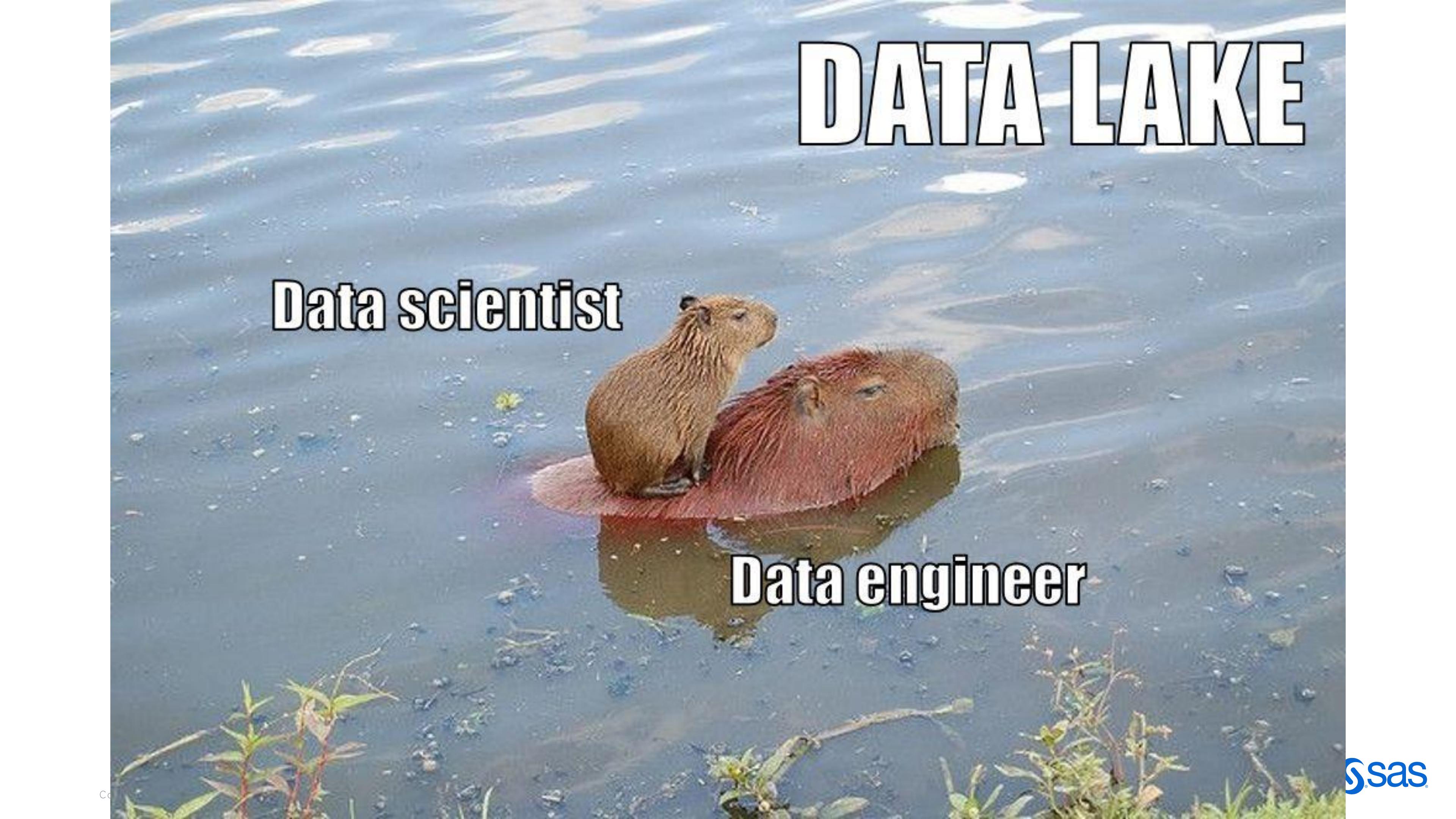


CRM SPACE



Analytic CRM

Operational CRM

A photograph of two capybaras in a body of water. One capybara is facing right, and the other is facing left. The water has ripples and some aquatic plants at the bottom. The word "DATA LAKE" is overlaid in large white letters at the top right. The words "Data scientist" and "Data engineer" are overlaid in large white letters on the capybaras.

DATA LAKE

Data scientist

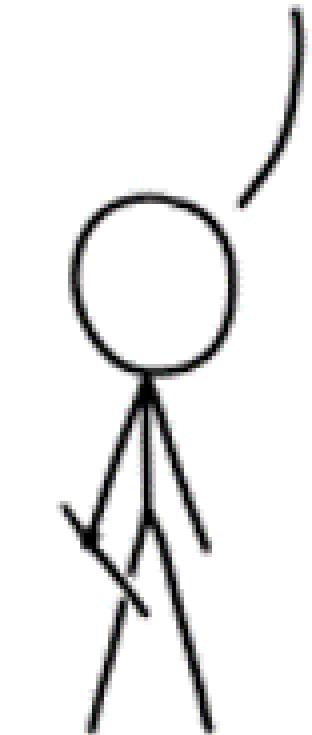
Data engineer

WE REALIZED
ALL OUR DATA
IS FLAWED.



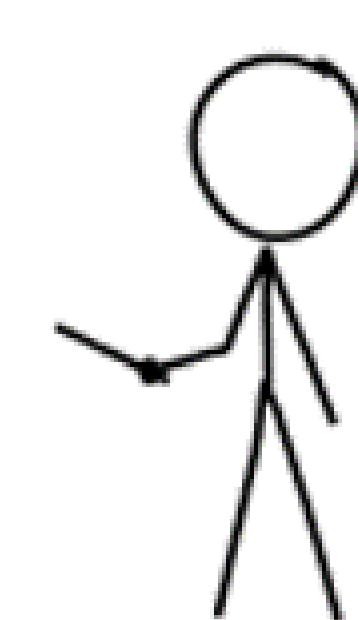
GOOD

...SO WE'RE NOT
SURE ABOUT OUR
CONCLUSIONS.



BAD

...SO WE DID LOTS
OF MATH AND THEN
DECIDED OUR DATA
WAS ACTUALLY FINE.



VERY BAD

...SO WE TRAINED
AN AI TO GENERATE
BETTER DATA.

