

Advance Analytics in a day

WELCOME



Introduction to Machine Learning

WELCOME



Facundo Santiago

Cloud Solution Architect

Advance Analytics and Artificial Intelligence

Microsoft

medium.com/@santiagof

fasantia@microsoft.com





Advance Analytics in a day Roadmap

Introducción a Machine Learning .1



Presentaremos una introducción a la teoría que está detrás de Machine Learning, sus conceptos y fundamentos, cuáles son los requerimientos para emplearla y construiremos finalmente nuestro primer pipeline básico de Machine Learning utilizando los conceptos que se presentaron.

Análisis predictivo en tiempo real .3



Daremos nuestros primeros pasos en la implementación de un modelo de Machine Learning que sea consumido en tiempo real. Este modelo utilizará los datos que se ingestan en tiempo real y realizará predicciones como por ejemplo detectar anomalías o predecir un resultado puntual. Este tipo de despliegues se los conoce también como hot-path ya que los datos son procesados tan rápido como son generados. Implementaremos un caso real de una solución de análisis de sentimiento de los Tweets que incluyen un hashtag que nos es de interés..



2. Modelado predictivo: clasificación y regresión

Nos adentraremos en lo que implica realizar un modelo predictivo, desde la adquisición de los datos, su transformación y preparación, elección de un algoritmo de aprendizaje y su posterior publicación para que sea consumido. Aprenderemos como hacer que nuestro modelo sea consumible a través de servicios web tipo REST y como podemos hacer uso del mismo de forma rápida incluso desde Excel, sin escribir una sola línea de código.



4. Manipulación de datos utilizando Python y Jupyter Notebooks

Comenzaremos a familiarizarlos con el lenguaje de programación Python. Utilizaremos un SDK para manipular y transformar los datos que tenemos antes de proveerlos a nuestros modelos de Machine Learning. Una de las tareas más comunes en Data Science. Nos familiarizaremos también con una herramienta clave para cualquier científico de datos: Jupyter Notebooks y veremos como podemos correrlos fácilmente en Azure..

Natural Language Processing con servicios cognitivos .5

Microsoft Azure provee un conjunto de modelos de Machine Learning listos para utilizar empaquetados como servicios REST que cualquier desarrollador puede utilizar rápidamente.

En esta sesión aprenderemos sobre estos servicios, como consumirlos, y como construir aplicaciones inteligentes rápidamente sin necesidad de construir nuestro propio modelo..

Operacionalizando modelos de Deep Learning .7

Veremos lo que implica llevar un modelo de Machine Learning a producción. Veremos como podemos medir la performance de nuestros modelos, evaluarlos, publicarlos como servicios y monitorearlos. Veremos también como tecnologías tales como Containers y Kubernetes serán de gran utilidad al momento de operacionalizar nuestros modelos junto con sus dependencias. Implementaremos un pipeline de Machine Learning completo que incluirá versionamiento de código, de modelos y de resultados...



6. Introducción a Deep Learning

Presentaremos uno de los conceptos más utilizados en Machine Learning: Deep Learning. Se trata de una técnica de Machine Learning en donde nuestros modelos son entrenados utilizando datos provenientes de un espacio altamente dimensional (veremos también que significa esto). Implementaremos un modelo de Machine Learning para Computer Vision en Python utilizando una de las librerías mas conocidas hoy en día: PyTorch (Creada por Facebook y open-source). Veremos como podemos realizar el entrenamiento de este tipo de soluciones utilizando Azure Machine Learning Services, tanto en la nube como en nuestro propio hardware.

8. Procesamiento en tiempo real a escala

Avanzaremos un paso más en nuestro camino de procesamiento en tiempo real, ahora utilizando Databricks. Haremos una introducción al procesamiento distribuido utilizando clusters Spark, en este caso Databricks. Haremos doble clic en una tecnología particular conocida como structure streaming, que nos permite procesar datos en tiempo real tan pronto llegan a la plataforma

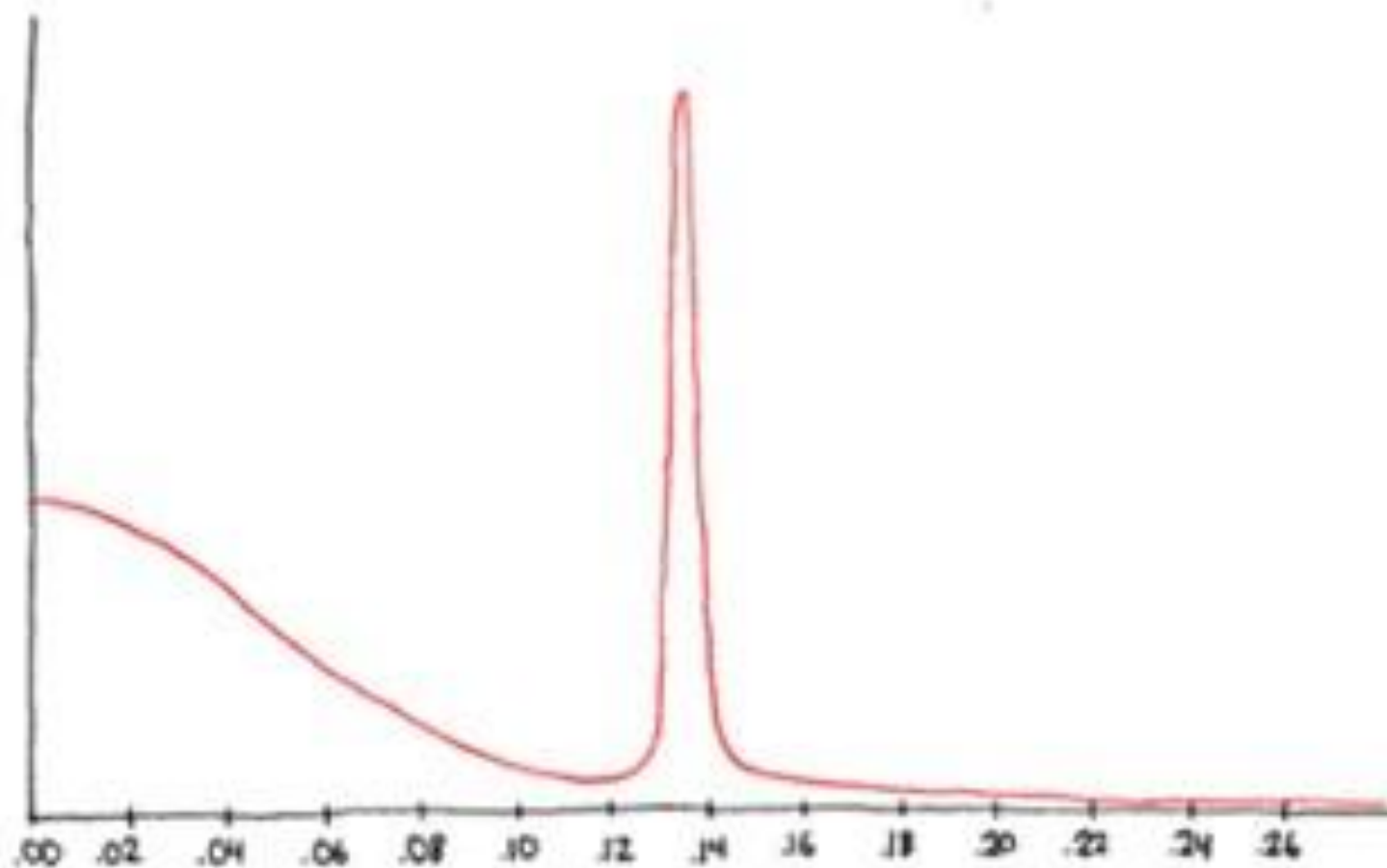
Introduction to Machine Learning



What is Machine Learning?



PROGRAMMING
SKILL



BLOOD ALCOHOL CONCENTRATION (%)

What about artificial intelligence?

*Machine
Learning*

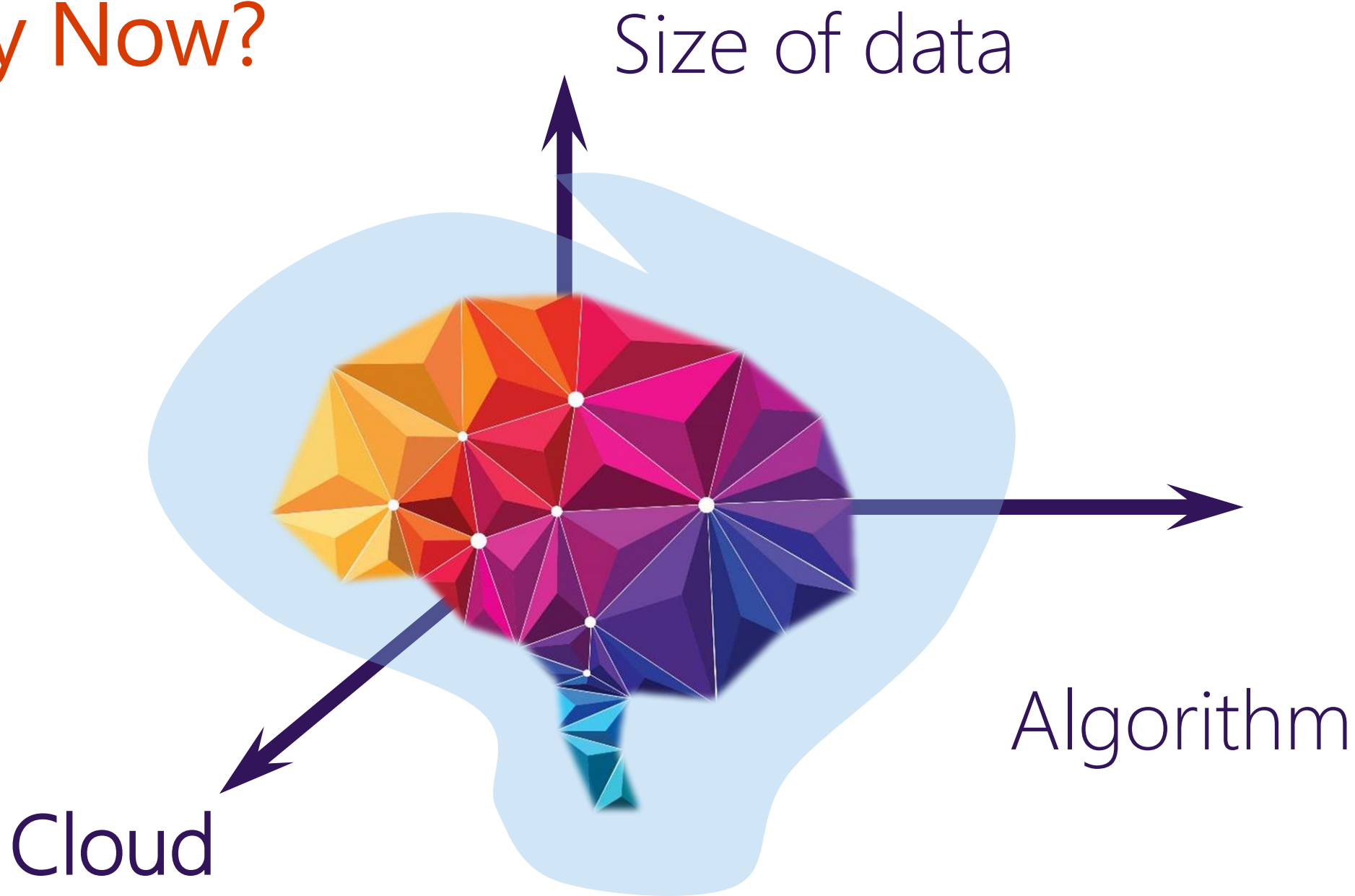
```
graph TD; AI[Artificial Intelligence] --- ML[Machine Learning];
```

Artificial Intelligence is defined as the ability to acquire and apply knowledge.

Actuators

```
graph TD; AI[Artificial Intelligence] --- A[Actuators];
```

AI: Why Now?

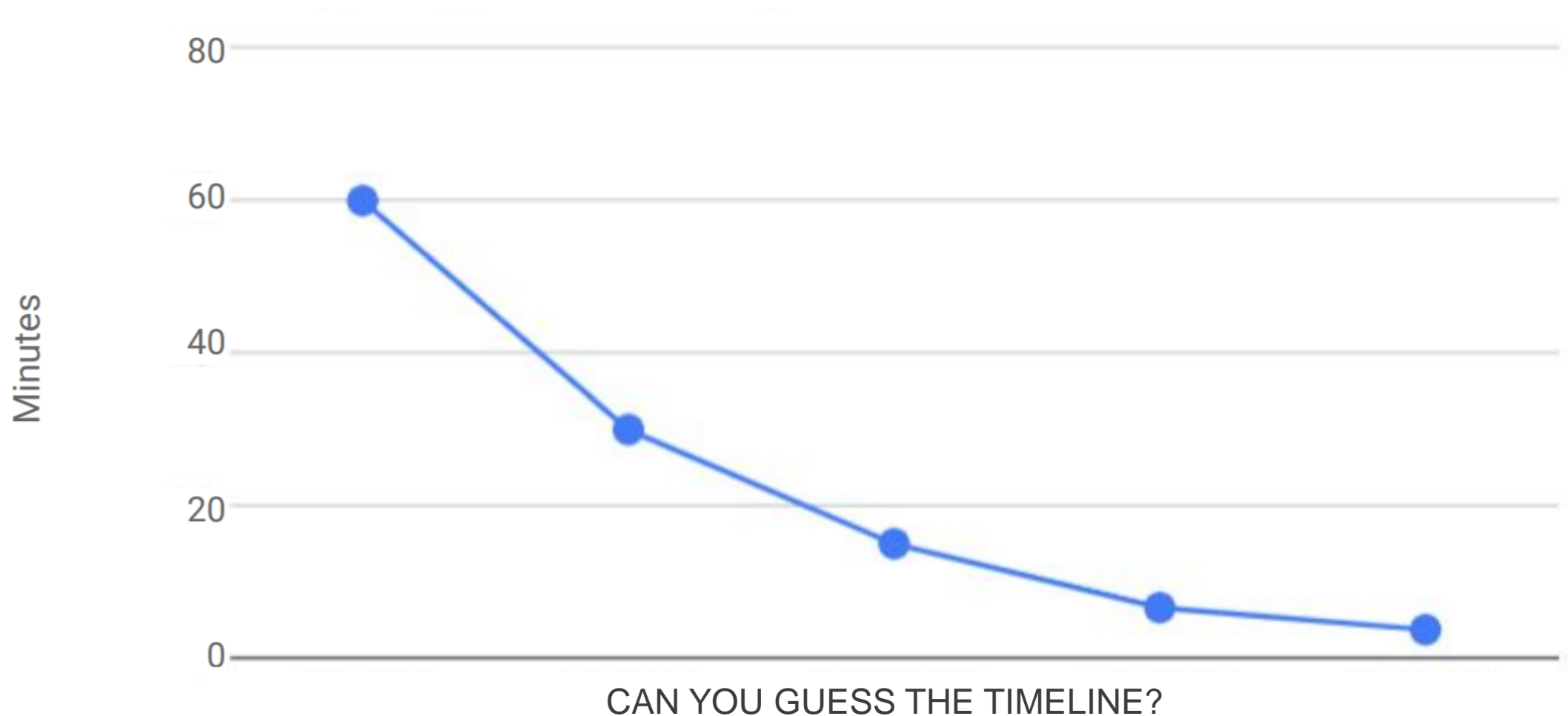


AI: Why Now?

ImageNet training time

Source: arXiv.org; see appendix for authors

14 MILLION
IMAGES



The modern data scientist

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



What can machine learning do?



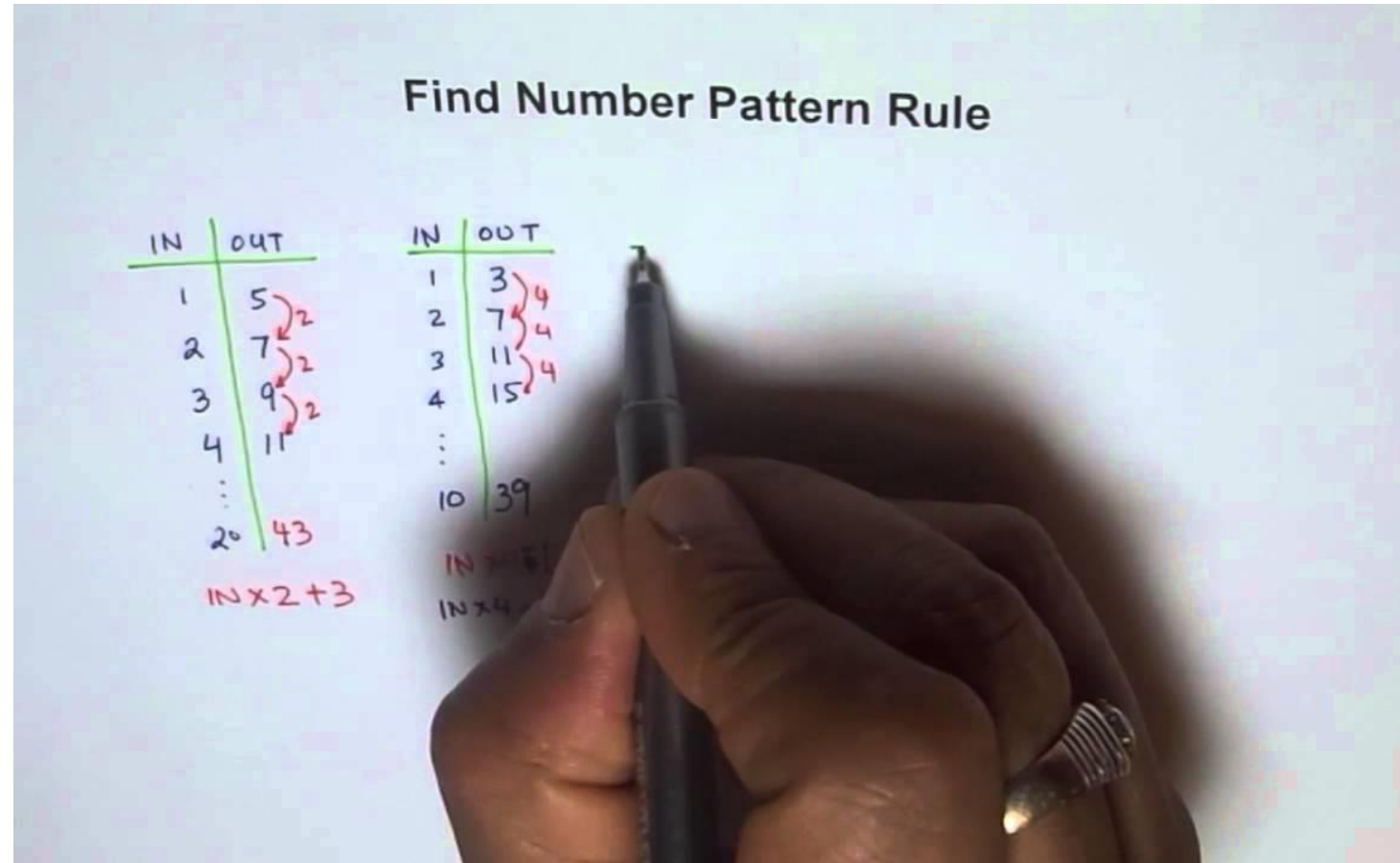
What Machine Learning does?

Finds patterns in data

Then uses those patterns to predict the future

Examples:

- Detecting credit card fraud
- Determining whether a customer is likely to switch to a competitor
- Deciding when to do preventive maintenance on a factory robot



Finding Patterns: A Simple Example

<i>Name</i>	<i>Amount</i>	<i>Fraudulent</i>
Smith	\$2,600.45	No
Potter	\$2,294.58	Yes
Peters	\$1,003.30	Yes
Adams	\$8,488.32	No

What's the pattern for fraudulent transactions?

Finding Patterns: Another Example

<i>Name</i>	<i>Amount</i>	<i>Where Issued</i>	<i>Where Used</i>	<i>Age</i>	<i>Fraudulent</i>
Smith	\$2,600.45	USA	USA	22	No
Potter	\$2,294.58	USA	RUS	29	Yes
Peters	\$1,003.30	USA	RUS	25	Yes
Adams	\$8,488.32	FRA	USA	64	No
Pali	\$200.12	AUS	JAP	58	No
Jones	\$3,250.11	USA	RUS	43	No
Hanford	\$8,156.20	USA	RUS	27	Yes
Marx	\$7,475.11	UK	GER	32	No
Norse	\$540.00	USA	RUS	27	No
Edson	\$7,475.11	USA	RUS	20	Yes

*What's the
pattern for
fraudulent
transactions?*

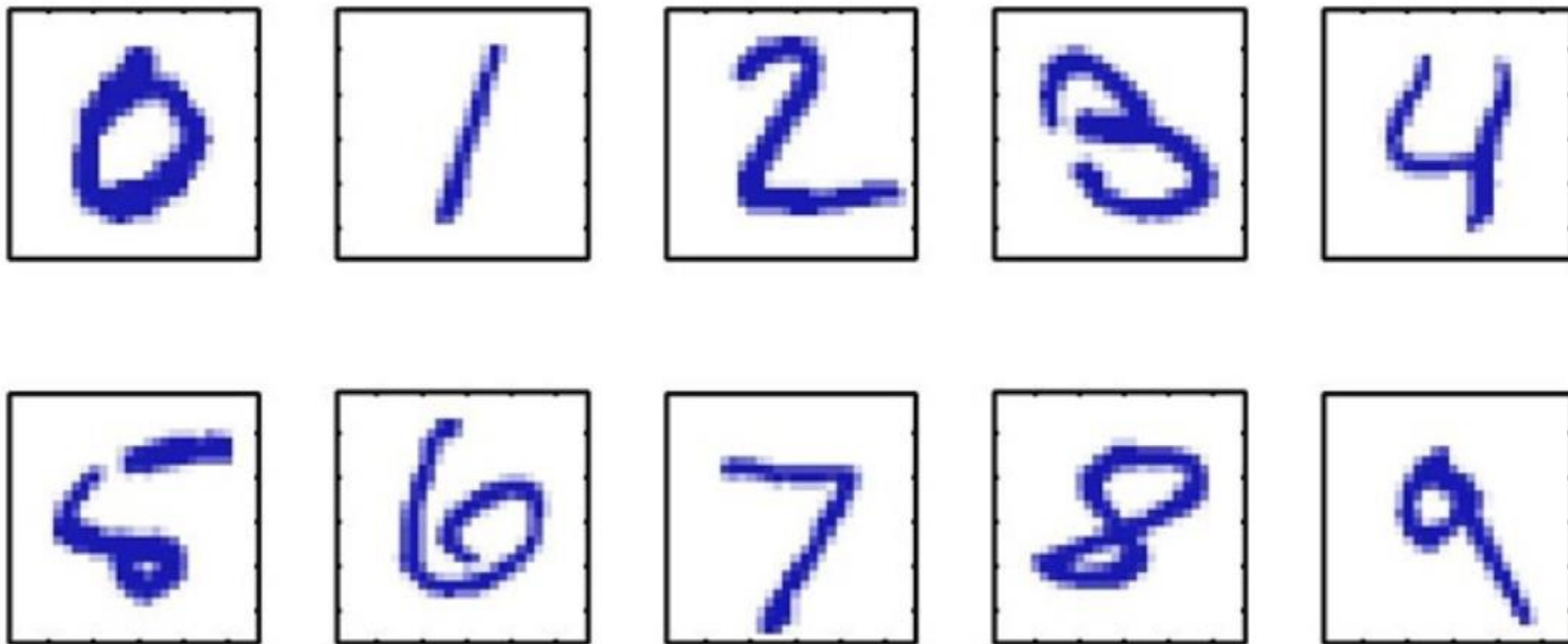
What is Machine Learning (revised)?

A computer program is said to learn from experience E with respect to some class of task T and performance measure P if P at T improves with E .



This is new!

Example: Handwriting digit recognition



Example: Handwriting digit recognition

- Non trivial problem due to variability
- What about using handcrafted rules or heuristics for distinguishing the digits based on strokes and shapes?
- Exceptions?
- How many rules?
- Solution? ML!

Example: Handwriting digit recognition

Task T:

- Recognizing and classifying handwriting digit within the image

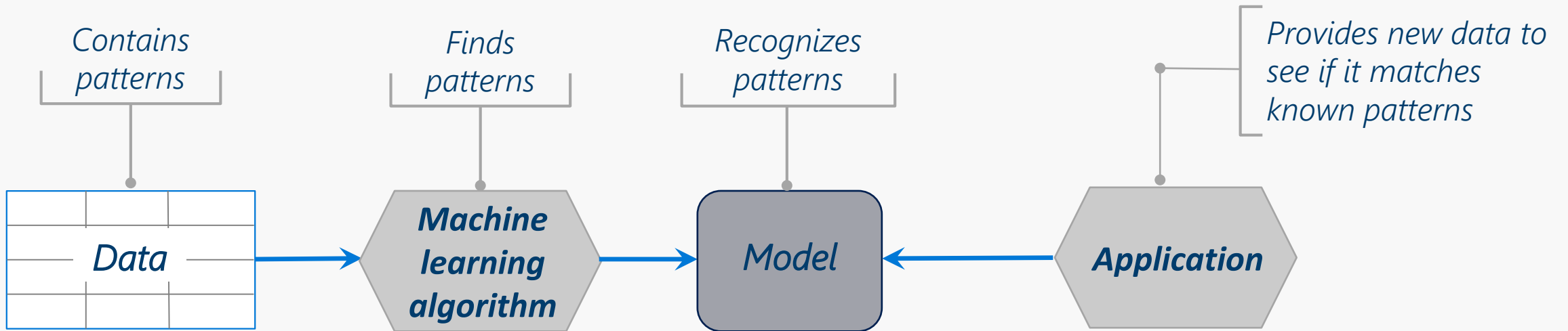
Performance measure P:

- Percent of digits correctly classified.

Training experience E:

- Database of handwritten digits with the correct number

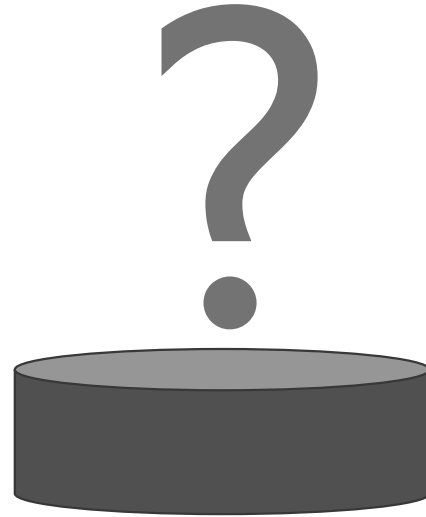
Machine Learning in a Nutshell



The First Problem: Asking the right question



Choosing what question to ask is the most important part of the process



Ask yourself: Do you have the right data to answer this question?



Ask yourself: What assumptions can I make about the data and the problem?

Types of Machine Learning

Types of **** Learning

- Supervised
- Unsupervised
- Semi-supervised
- Reinforcement
- Deep (Representation)
- Active
- Online
- Zero-shot
- Curriculum
- Q-Learning

Types of Machine Learning

Supervised learning

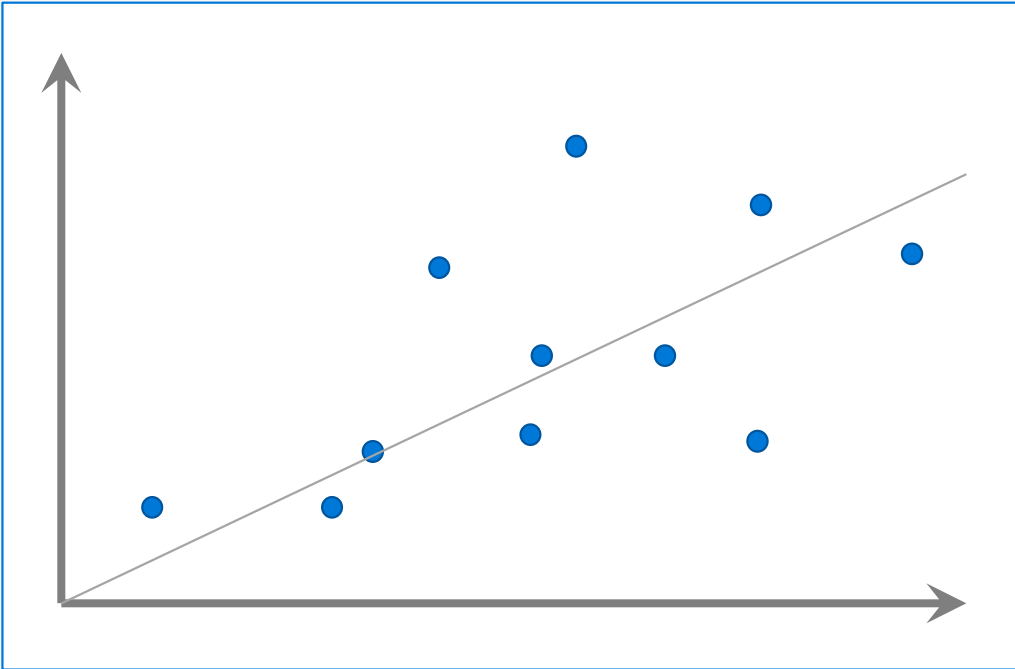
The value you want to predict is in the training data

The data is *labeled*

Target is a value

Categorizing Machine Learning Problems:

Regression

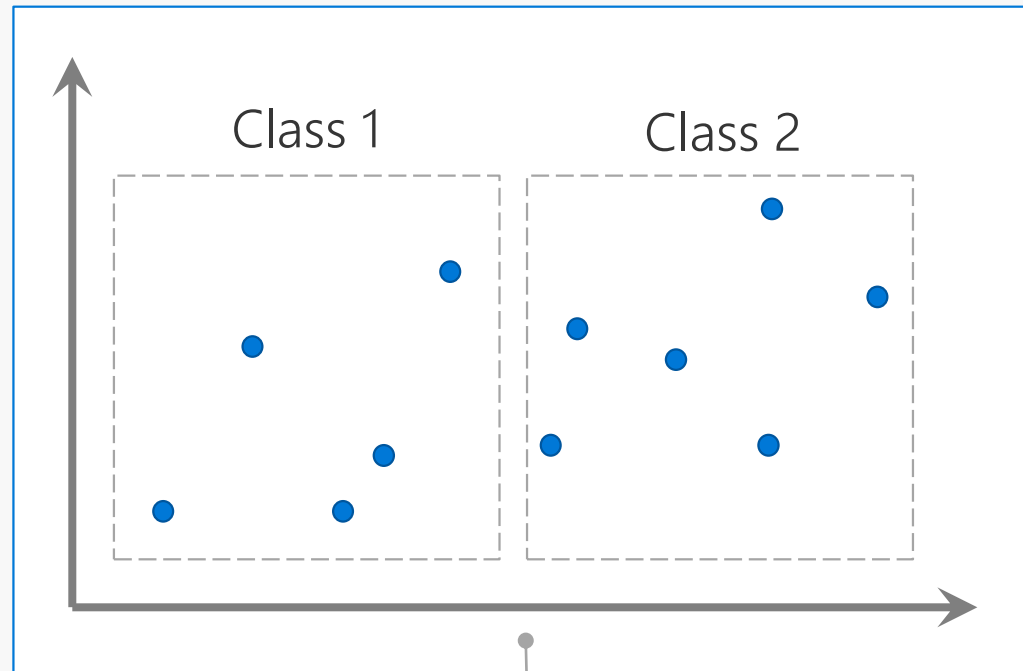


Goal: Predict a value
For *supervised* learning

Example question:
How many units of this product
will we sell next month?

Categorizing Machine Learning Problems:

Classification



Can be more than
two classes

Goal: Predict a class
For *supervised* learning

Example question:
Is this credit card transaction
fraudulent?

Types of Machine Learning

Supervised learning

The value you want to predict is in the training data

The data is *labeled*

Target is a value

Unsupervised learning

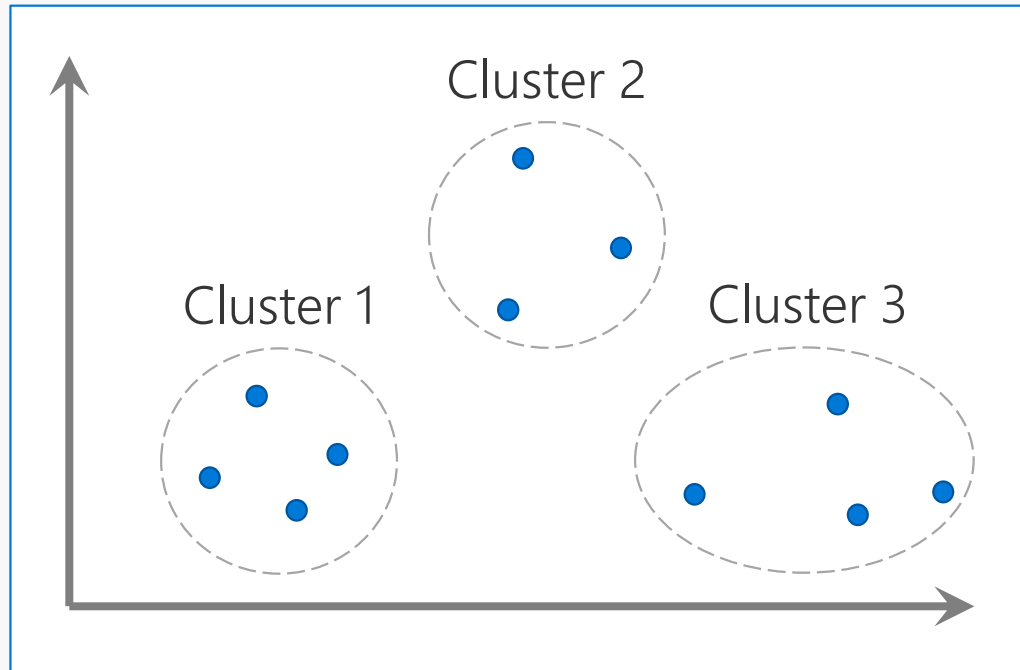
The value you want to predict is not in the training data

The data is *unlabeled*

Target is a function

Categorizing Machine Learning Problems:

Clustering

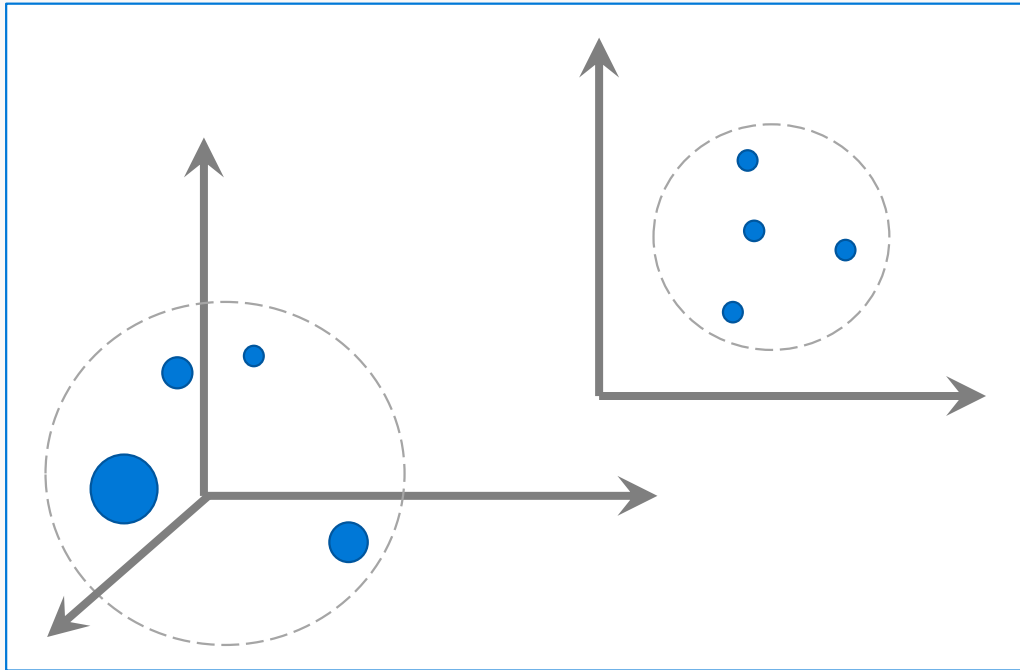


Goal: Discover structure
For unsupervised learning

Example question:
What are our customer segments?

Categorizing Machine Learning Problems:

Dimensionality reduction



Goal: Learn latent features
For *unsupervised* learning

Example question:
What information tell the most
about my customers?

Types of Machine Learning

Supervised learning

The value you want to predict is in the training data

The data is *labeled*

Target is a value

Semi-supervised learning

The value you want to predict is in the training data

The data is *unlabeled*

Target is a function

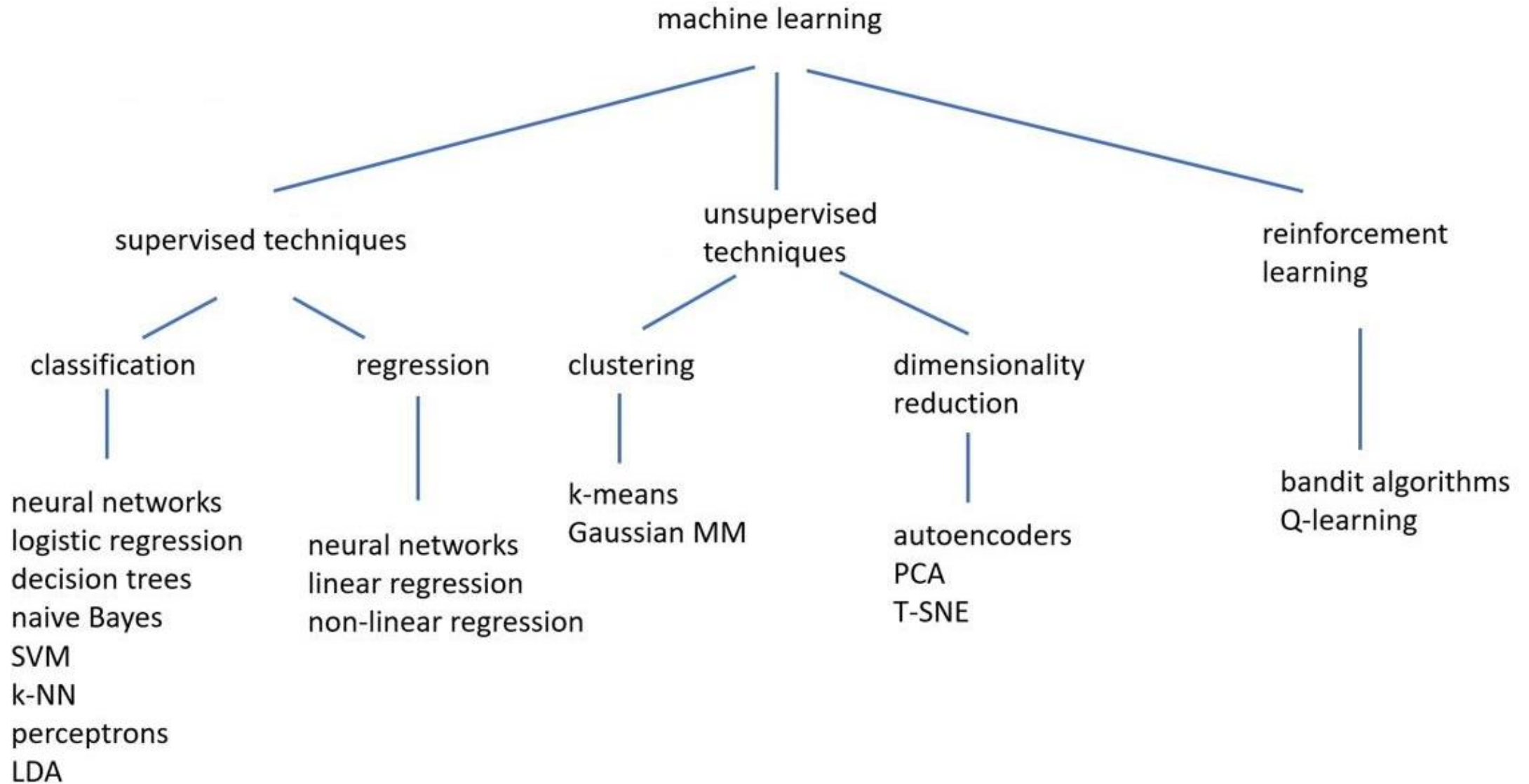
Unsupervised learning

The value you want to predict is not in the training data

The data is *unlabeled*

Target is a function

Types of Machine Learning





No free-lunch theorem

The theorem:

If you look at **all possible problems** that you might apply machine learning to, then, **on average any algorithm is as good or bad as any other.**

Why is so important?

You cannot learn just from data!

The First Problem: Asking the right question



Choosing what question to ask is the most important part of the process

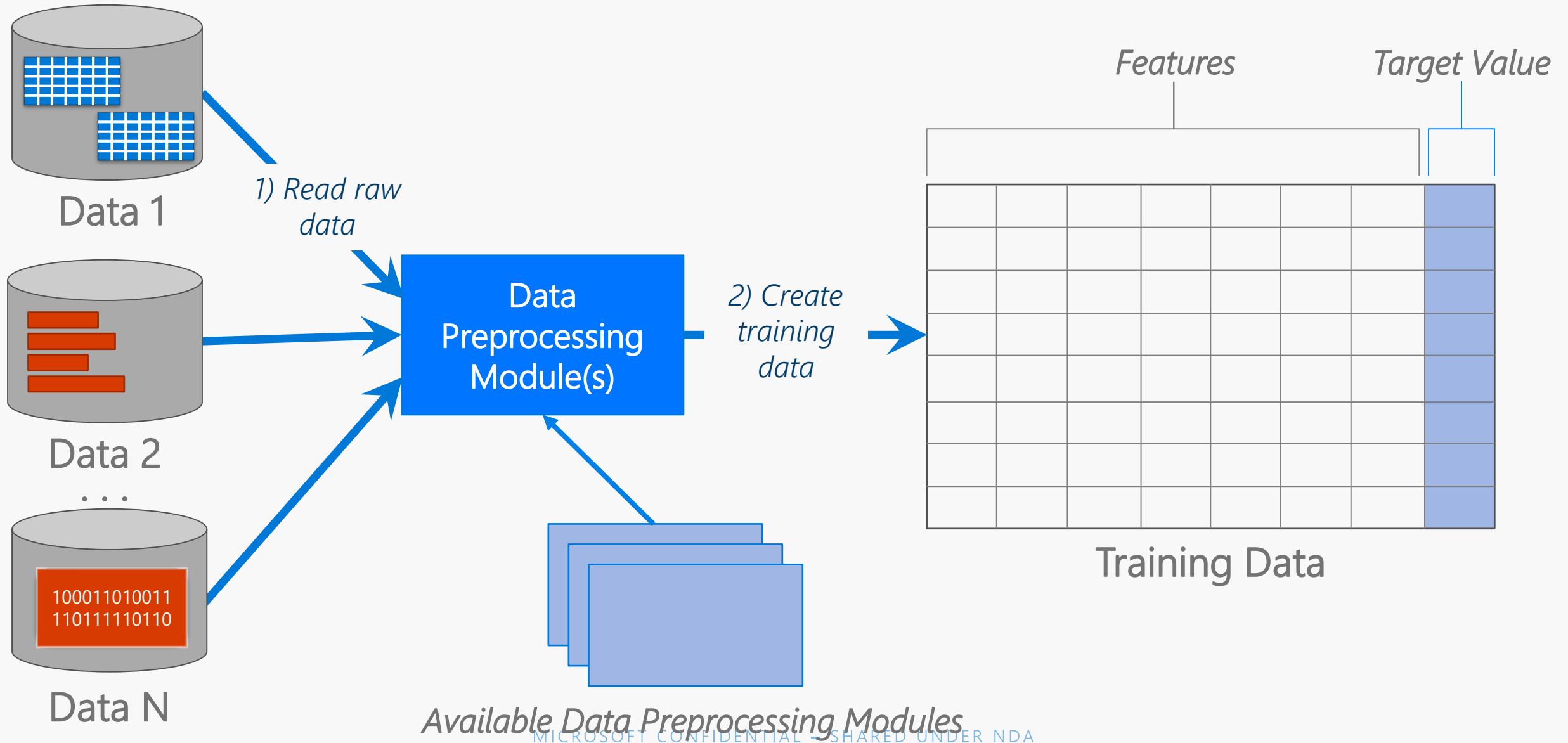


Ask yourself: Do you have the right data to answer this question?



Ask yourself: What assumptions can I make about the data and the problem?

Example with Supervised Learning



Generalization

- The ability to categorize correctly new examples that differ from those in the training set.
- Generalization is a central goal

Overfitting and Underfitting

	<i>Prediction Quality of Model with Training Data</i>	<i>Prediction Quality of Model with Real Data</i>	
Goal	<i>High</i>	<i>High</i>	
Underfitting	<i>Low</i>	<i>Low</i>	• [<i>Model too simple?</i>
Overfitting	<i>High</i>	<i>Low</i>	• [<i>Model too complex?</i>

Thank you

Coming next: Classification and regression

