OSLOMET

Introduction to artificial intelligence applied to organizations and industries

Core Business Data and Supervised Machine Learning

Ola Huse Ramstad

OSLO, 6 OF AUGUST 2020

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Core business data is the kind of data with the strongest ties to the value- generating engine of an organization. Therefore, it also holds the highest potential for impact by AI.

 Supervised learning is a family of machine learning algorithms that allows computers to learn how to map inputs (features) and outputs (labels), given enough examples.

 Machine learning can tackle many hard problems where conventional software engineering fails, because it's based on historical data rather than mathematical understanding.

 Even if ML-based models can never be 100% accurate, numerical metrics (such as accuracy) can help track their performance.

Artificial Intelligence is any trait exhibited by machines that is considered similar to traits of human intelligence.
(STRONG/ GENERAL AI)

Any device that perceives its environment and takes actions that maximize its chances to successfully achieve a goal.
(WEAK/ NARROW AI)

We are departing from the realm of general AI. Too much hype and high expectations generated mistrust on the idea that machines can actually be as intelligent as us in so many different fields.
Instead, let's focus into something that can deliver more immediate value and is easier to understand. The weak AI.

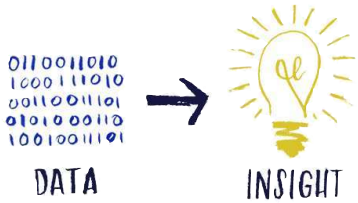
GENERAL AI is Occupied with the greater goal of reproducing and understanding intelligence. A kind of psychology (in Sylico) from the engineering perspective. Nowadays it is mostly relevant for basic research.
NARROW AI are systems able to perceive a narrow segment of the world, perform operations and deploy

decisions or actions in response to them. Essentially, it is automation of decisions. Narrow AI is more relevant for applied research and business, as it is part of almost everything we do today, even in things you do not think as AI.

There are 3 broad use-case areas for weak Artificial Intelligence in business

Task automation

- Assembly lines
- Expert systems
- Fraud detection
- Anti-spam
- Image classification
- Self-driving vehicles
- Fast trading



Insight generation

- Business Insight
- Data analytics
- Pattern recognition
- Sciences
- Traffic control
- Churn analysis

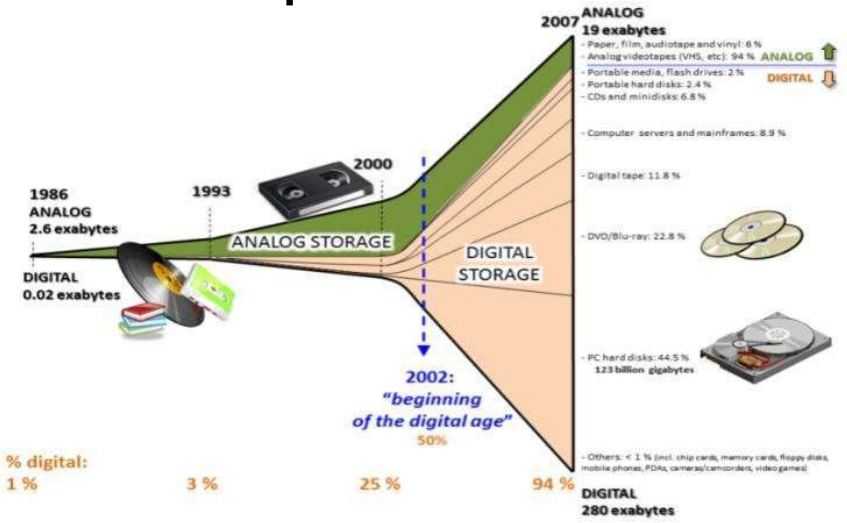
People engagement

- Entertainment
- Intuitive user interface
- Social reward
- Click baiting/ phishing
- Chat-bots
- Voice assistants
- Face recognition

As you saw


BIG
DATA
ALSO
TOOLS
DISK
TARGET
APPLIED
SENSOR
MAY
EXAMPLES
EVERY
SHARED
DEFINITION
CURRENT
MOVING
WIRELESS
MANAGE
ARCHIVES
WITHIN
BIOGEOCHEMICAL
TECHNOLOGIES
USING
INTERNET
SETS
DISTRIBUTED
BUSINESS
GENOMICS
CAPTURE
PRACTITIONERS
RECONSIDER
ZETTABYTES
THOUGHT
NEEDED
REQUIRING
TIME
ONE
LARGE
INFORMATION
PROCESS
MANAGEMENT
TERABYTES
CASE
ABILITY
SOFTWARE
PETABYTES
INCLUDE
ELAPSED
SYSTEMS
TOLERABLE
PERFORMANCE
STORE
SIZE
RESEARCH
STORAGE
LOGS
COST
MPP
SIGNIFICANT
DEFINING
RELATED
COMPUTING
COMBAT
PARALLEL
SAN
QUALITIES
SINCE
MASSIVELY
GROW
SEARCH
DIFFICULTY
GARTNER
WORKING
OPPORTUNITIES
AMOUNT
CURRENTLY
SOLID
TYPES
DESKTOP
RECORDS
NETWORKS
DATABASES
HUNDREDS
CAPACITY
BURIED
WORLD'S
BIOLOGICAL
PROCESSING
NOW
USED
CREATED
UBIQUITOUS
TENS
COMPLEX
CITATION
INDEXING
LARGER
CONTINUES
SOCIAL
USE
SET
RELATIONAL
ORGANIZATIONS
CONNECTOMICS
ANALYTICS

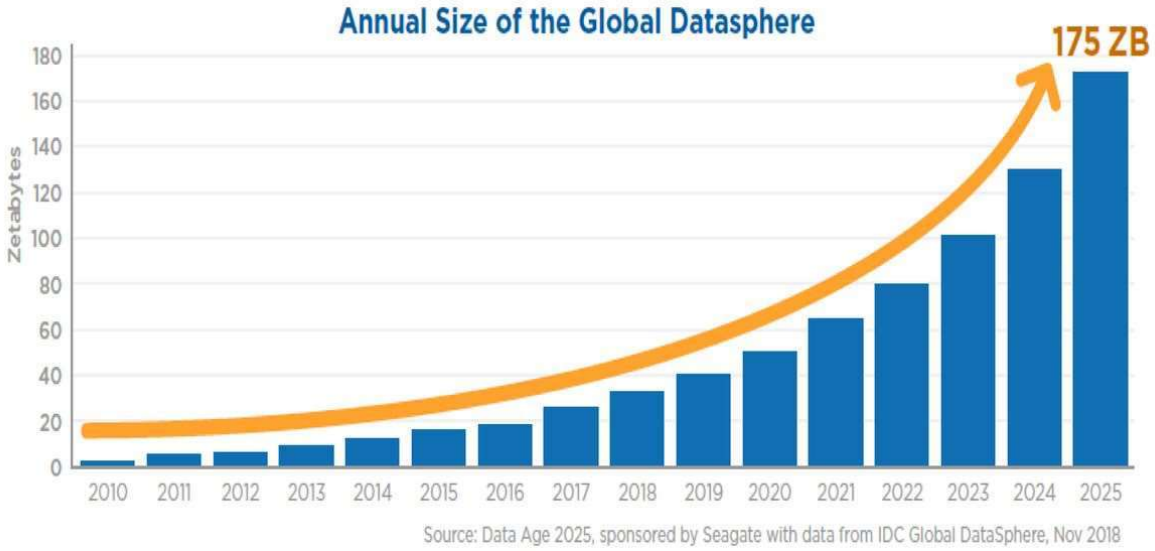
It became cheap to store data



Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, *332*(6025), 60-65.

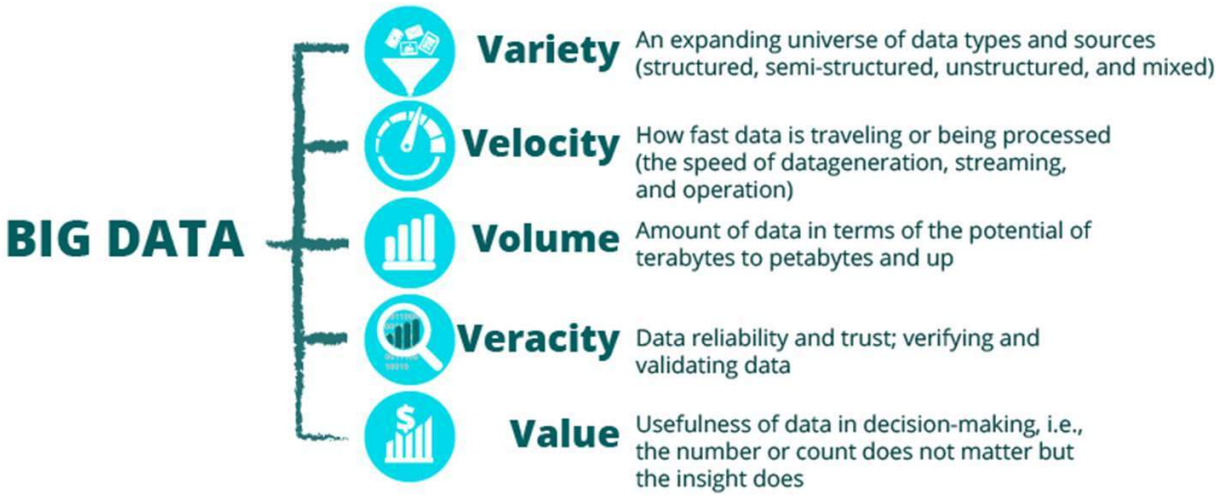
One of the previous limitations for machine learning was the lack of methods to store data. Machine learning is very data dependent and its only in the past 20 years that we've really had the technology and data to fully utilize it.

Data is everywhere and it is growing



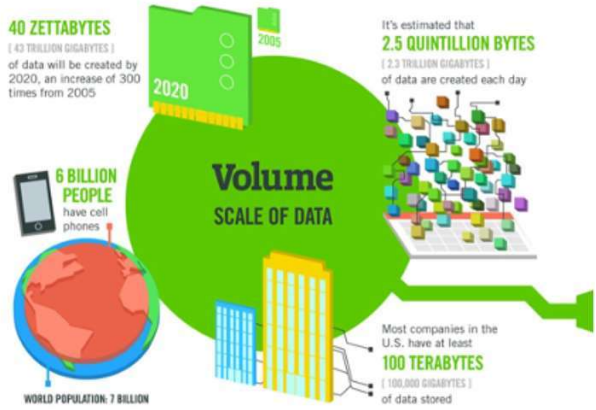
Data is growing exponentially, in part due to the internet

OSLOMET

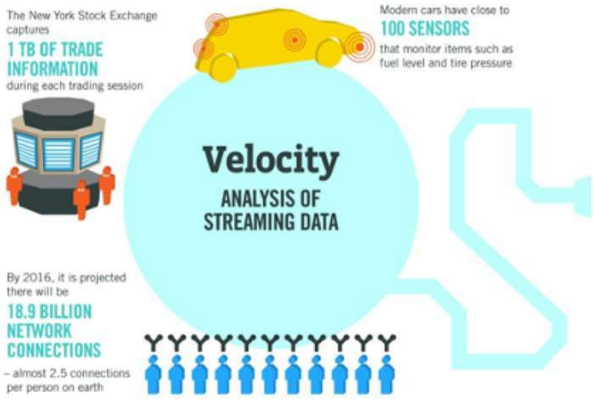


What is big data in the context of learning?

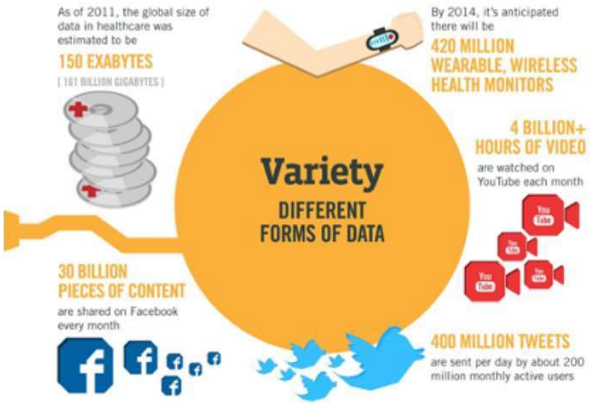
More data than conventional methods can handle



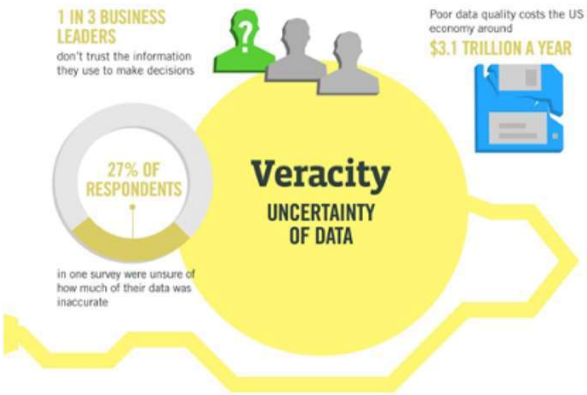
Data is generated faster than we are able to store



Unstructured data with multiple forms of encoding



The quality of the data limits the quality of the insight



https://en.wikipedia.org/wiki/Data_cleansing

OSLOMET

It is hard to identify and extract the “core business data”

"Core Business Data" *is the data closest to the value proposition of the business.*

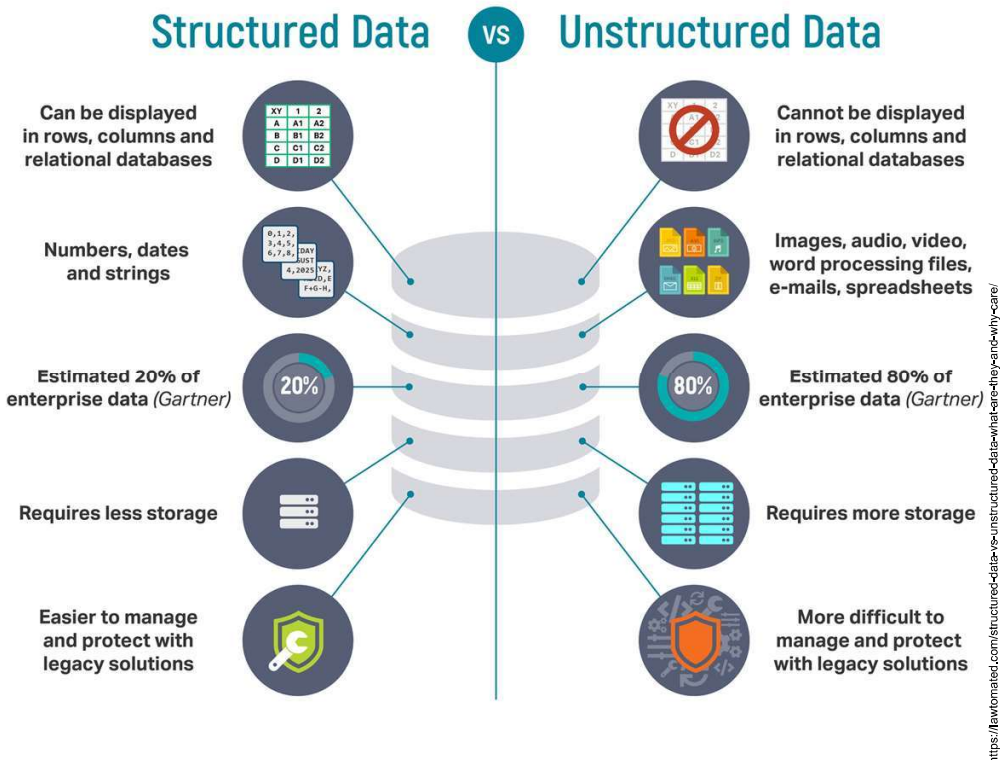
It describes events and patterns that have a direct impact on the organization's performance, and it is easy to attach monetary value to it.

Machine learning emphasizes on learning from data.
But what does it mean?

Money density of data estimates how much data influences the top or bottom line of the organization

- Data that is actionable and can lead to reduction of costs or increase in profit.
- **Core business data** has **high money density** value
- Changes are that there is a lot of core business data available as:
  - transactions
  - reports, reviews, complaints, opinions
  - tables
  - videos and images
  - phone calls
- Behaviors talk louder than opinions.
- It is the **prime candidate for AI applications** as any improvement in performance is almost guaranteed to make a dramatic impact

Dollar density of data: how much the data influences the top or bottom line of the organization. Core business data has a high dollar density:
each e-commerce order, job lead, or financial transaction has a direct impact on your top or bottom line.



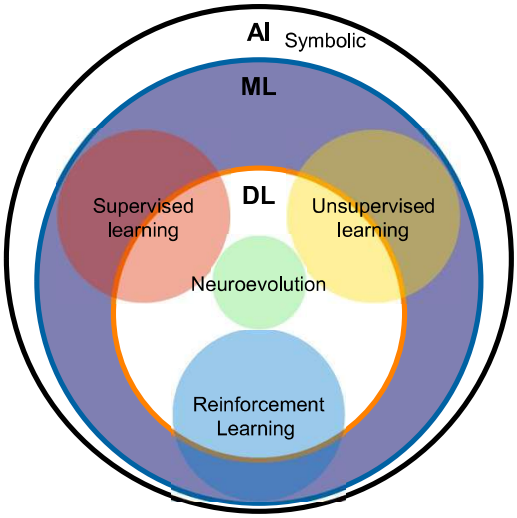
Structured data is essentially data that was properly organized for data analysis. Which is essentially to put it into some form of table based database
Unstructured data is pretty much everything else. Videos, texts, audios. Things that contain information but do not have a clear way to decode it.
Although much value can be taken from structured data, most of the information is in the unstructured form. And we should find a way of extracting value from it.

So how AI can help?

A machine has artificial intelligence if it can interpret data, ***potentially learn from the data*** *and use that knowledge to take decisions, adapt to its environment, potentially generate novel behavior and achieve specific goals.*

Let’s review our definition of Artificial Intelligence. …

Different kinds of AI



- **AI:** Artificial Intelligence
- **Symbolic AI (GOFAI):** Separation between logic and data, programmed rules, ***Fuzzy logics***
- **ML:** Machine Learning
- **Supervised Learning:** Learn from examples, requires structured and labeled data
- **Unsupervised Learning:** Learn patterns and groups in unlabeled data
- **Reinforcement Learning:** Learn from experience, trial and error, reward policy
- **DL:** Deep Learning
- **Generative Algorithms:** Exploit mutation and selection to find a good solution to the problem at hand

It seems that the solution for an explosion of data is not symbolic ai, but some technique that allows the machine to learn from the data without being explicitly programmed.
Enters…
Machine learning We also mentioned many different techniques and types of Artificial Intelligences to solve different types of tasks.
Each of them has different characteristics that must be understood in order to apply them to the adequate problem.
To understand them means:

- To know what they do
- Their advantages and disadvantages
- What problems they solve

In this diagram we have an overview of the methods discussed
In the outer circle with have all that may be considered AI
In that outer circle we have the Symbolic AI or Good Old fashion AI, which is present in almost everything digital today
The blue circle represents all the AI systems that are able to learn . Machine learning
Inside of it we have the 4 main kinds of machine learning

- Supervised learning - or learning from examples
- Unsupervised learning - or learning to identify groups and patterns
- Reinforcement learning - or learning from experience, trial and error
- Genetic algorithms - adapt with mutation and selection

Machine learning can be done with classical method or with Neural Networks, now known as deep learning
All the kinds of machine learning can leverage deep learning and its ability to extract features from unstructured data.
If these words do not mean much, do not worry, we will eventually explain them. Just be aware that deep learning is just one special way for doing machine learning.

-

Symbolic AI based on high level symbolic, representation of problems, logic and search. It does not require massive amount of data or symbols. Instead, the world and objects

in the world are represented as symbols. These symbols are then manipulated using logic to search for solutions.

Logic is the problem solving technique and symbols are how we are going to represent the problem in the computer.
Symbols can be numbers, texts objects, etc.
A relation can be an adjective that describes a symbol, and we write it in front of the symbol that is inside the parenthesis
A relation can also be a verb that describe how a symbol interacts with other symbols EAT(GUSTAVO, DONUT)

Symbols = nouns
Relations = Adjectives or verbs
Logic = AND, OR, NOT

The collection of all symbols in a software (the representation of the entire world) is called a knowledge base (HOW SIRI WORKS)
Knowledge base can be combined to build propositions
We can test if these propositions are true or not
With tools of propositional logic called truth table
It converts each statement as 0 or 1
AND as multiplication and OR as addition
If the proposition evaluates true, than it is true

IMPLICATIONS connects one proposition to the next (IF/THEN)
AN implication is true if the THEN side is true or the IF side

is false

We can start build a knowledge base of all the propositions that are true in our knowledge base. With that we can use the AI to answer questions and discover new things.
You can provide a set of true propositions and use logic operations for the computer to search for new true propositions without the human giving them explicitly.
This is INFERENCE
(sandwitch discount example)
Expert systems
https://www.youtube.com/watch?v=WHCo4m2VOws
Expert systems can make conclusions based on logic and reason, not just trial and error guesses like neural network
Expert systems can explain its decision by showing which parts were evaluated as true or false. GOOD OLD FASHION AI

PROBLEMS:
DIfficult to scale
Not easy to represent all forms of data
Some things are not clear-cut deterministic such as true or false
Symbols are not static, in time, symbols change as well as their relations.

FUZZY LOGIC way of dealing with uncertainty
Not 0 or 1 but values between 0 and 1
Hot cold

Car break
Uncertainty

THERE IS A SET where things within it are true, outside of it are false

THEY ARE USED EVERYWHERE, they are just now embedded in what we use in normal devices
Expert systems Hot shower, washing machine
The Sendai Subway 1000N series (仙台市交通局1000N系電車) is a rapid transit electric multiple unit (EMU) train type operated on the Sendai Subway Namboku Line in Sendai, Japan.
https://www.youtube.com/watch?v=r804UF8Ia4c

The 1000 series was the world's first train type to use fuzzy logic to control its speed, and this system developed by Hitachi[1] accounts for the relative smoothness of the starts and stops when compared to other trains, and is 10% more energy efficient than human-controlled acceleration.[2] It was the recipient of the 28th Laurel Prize award presented by the Japan Railfan Club

What is learning?

Machine learning emphasizes on learning from data.
But what it means?
What is learning?

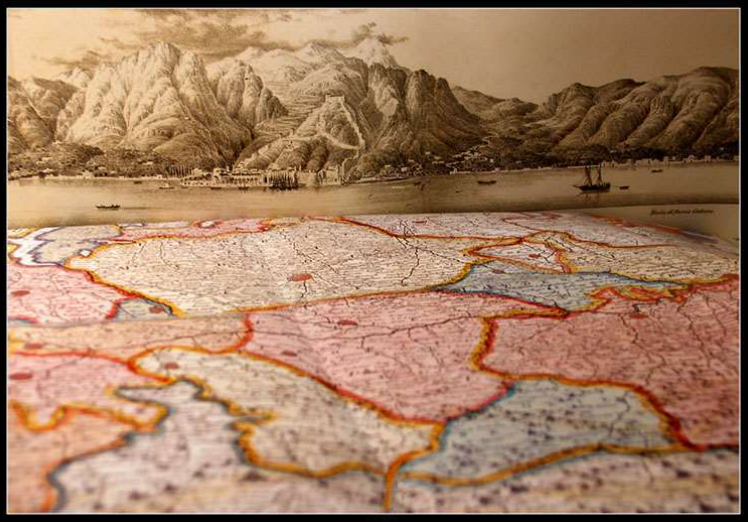
Learning are changes in behavior as result of experience

- Changes in behavior that do result from lesion are not learning
- Learned behaviors are functionally connected to the experience they were acquired
- Good learning leads to increase of performance towards achieving a goal
- Not all learning increases performance, some reduce performance (dysfunctional or pathologic behavior)
- Machines learn by creating models from data

Machines learn by creating models from the data.

There are many kinds of learning

Machines learn by creating mathematical models of the world from data

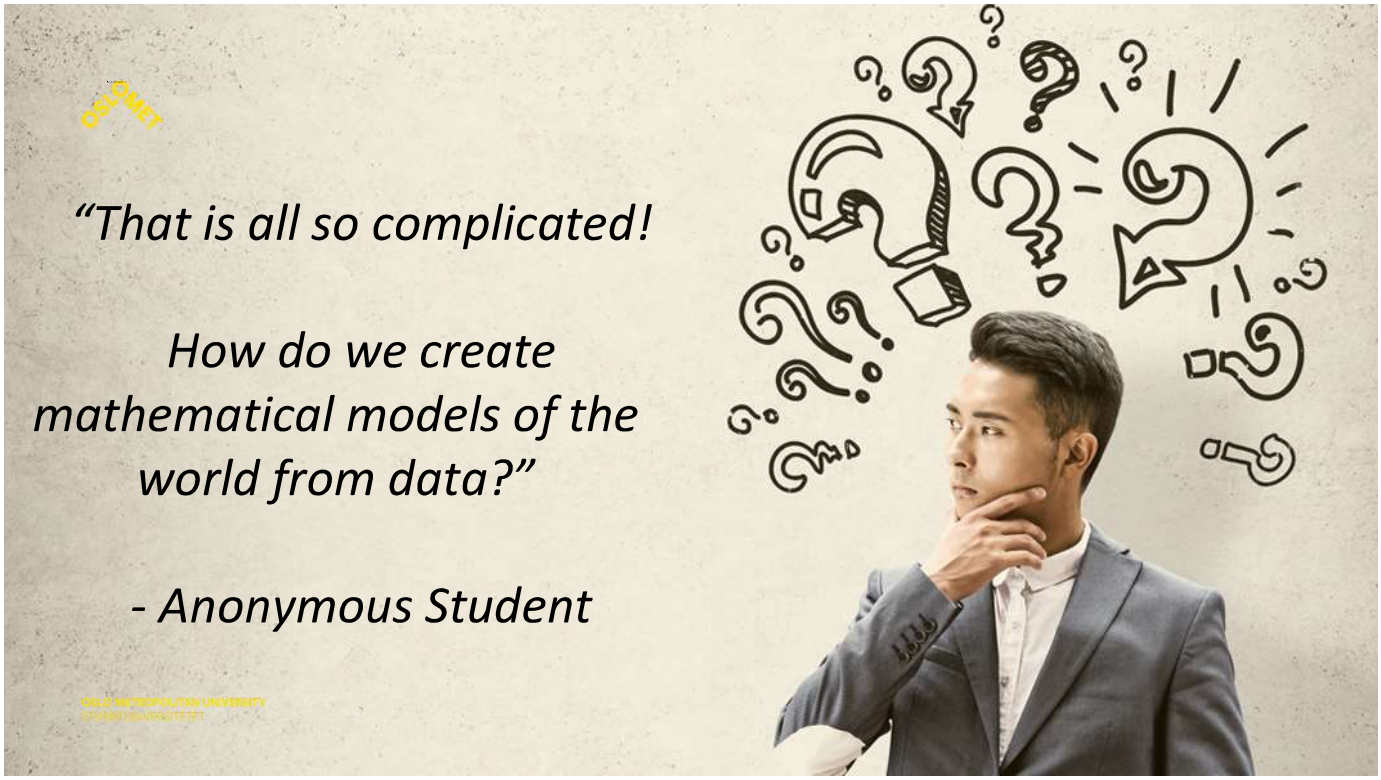


- Models are simplified descriptions of an observation or a process
- Like a map is to a territory
- It highlights important and relevant information
- It ignores most of irrelevant information and some of the important too
- There are many kinds of models:
  - Descriptive
  - Predictive / retrodictive
  - Mechanistic
  - Normative

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Descriptive models - describe what you observe now in a simple way. It is a form of removing complexity
Predictive models and retroactive models - look for correlations and trends in the data to point out how data would look like if the trend continues in the future or in the past
Mechanistic models - describes the process or phenomenon in terms of its parts and causes. This can only be developed through carefully designed experiments
Normative models - describe optimal strategies. In other words, given a problem, what is the optimal way to solve it?

“All models are wrong, but some are useful”
- George Edward Pelham Box



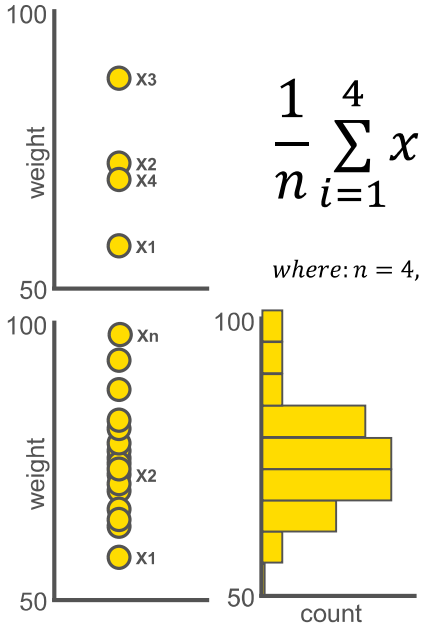
Building such a set of instructions is called programming.
To do so, you need some fundamental elements like
Values: that can describe the world for you. Using Names for categories of things and Numbers to express magnitudes and amounts.
Variables: to hold in memory the values while you perform the operations and actions
Actions: ways to transform values, like adding , subtracting…
Decisions: ways of comparing values and take one set of actions or the other depending on what is observed
Repetitions: ways to repeat a set of actions and value transformations many times
And Abstractions: That can encapsulate many small

detailed actions into a set of actions that can be understood as an action.

Mathematical models start from data

People	Weight
X1	65
X2	78
X3	89
X4	75

People	Weight
X1	65
X2	78
…	…
Xn	99



$$\frac{1}{n}\sum_{i=1}^{4} x_i \qquad \frac{65+78+89+75}{4} = 76{,}74$$

$where: n = 4, i = 1 \ldots 4, x = values, \sum = summation$

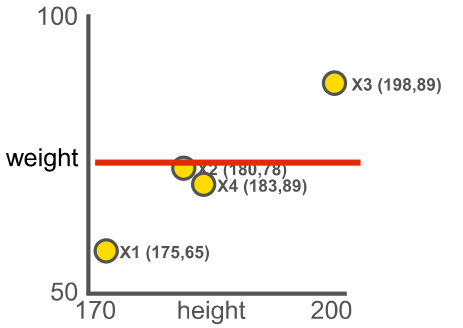
mean/average

median: middle value in the data set
mode: most frequent value

Lets say we have one variable Weight. Just so you know, a variable is any named feature or characteristics which value may vary, depending on the observation in the data set.
In a structured data set (or table) values are organized in this way typically. Observations as rows (that would be individuals, time, sales, etc) and the columns are the different variables or features. How you name your observations or your variables is irrelevant.
But for the sake of example, lets say we have 4 people, with their names properly anonymizes. And their respective weights.

Mathematical models can be used to express relationships between variables

People	Weight	Height
X1	65	175
X2	78	180
X3	89	198
X4	75	183

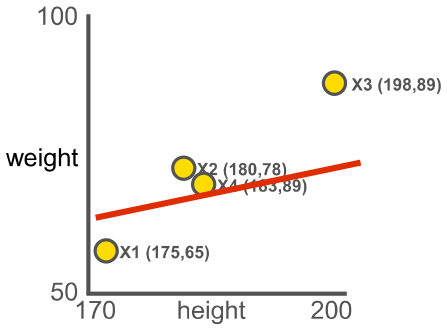


$$weight = b_1 * height + b_0$$

$Where: b_1 = slope, b_0 = bias/intercept$

Mathematical models can be used to express relationships between variables

People	Weight	Height
X1	65	175
X2	78	180
X3	89	198
X4	75	183



$$weight = b_1 * height + b_0$$

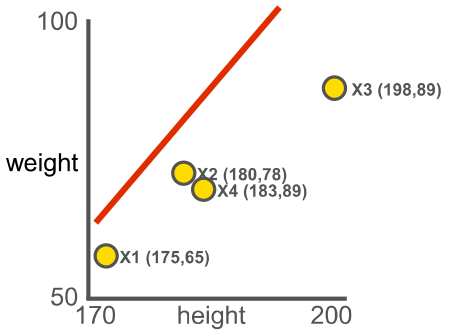
$Where: b_1 = slope, b_0 = bias/intercept$

$$avg(weight) = 0 * height + avera$$

$$underestimation = 0.02 * height + b_0$$

Mathematical models can be used to express relationships between variables

People	Weight	Height
X1	65	175
X2	78	180
X3	89	198
X4	75	183



$$weight = b_1 * height + b_0$$

$Where: b_1 = slope, b_0 = bias/intercept$

$$avg(weight) = 0 * height + avera$$

$$underestimation = 0.02 * height + b_0$$

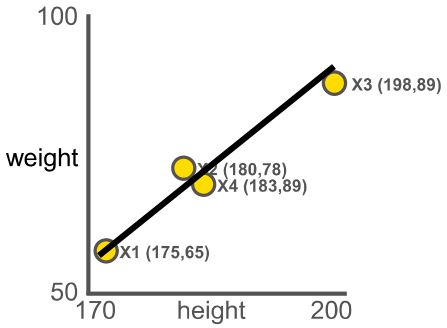
$$overestimation = 3.7 * height + b_0$$

When generating lines of best fit to our data, we can calculate the deviation, or error, between the line we produce and the data.

By taking the Mean error you can adjust the line to better fit the datapoints and provide a closer fit.

Mathematical models can be used to express relationships between variables

People	Weight	Height
X1	65	175
X2	78	180
X3	89	198
X4	75	183



$$weight = b_1 * height + b_0$$

$Where: b_1 = slope, b_0 = bias/intercept$

$$avg(weight) = 0 * height + avera$$

$$underestimation = 0.02 * height + b_0$$

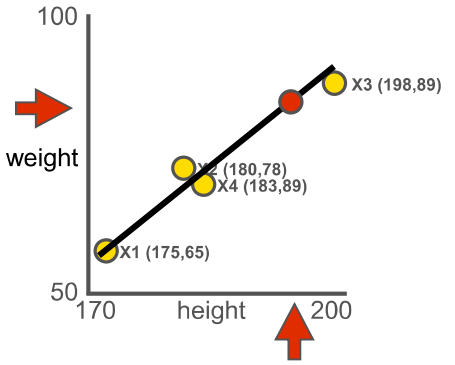
$$overestimation = 3.7 * height + b_0$$

$$closestestmiation = 0.63 * height + b_0$$

When using linear regression, the fit between the data and the generated model is given by the R squared, the closer this is to 1, the better the fit of the model.

Mathematical models can be used to express relationships between variables

People	Weight	Height
X1	65	175
X2	78	180
X3	89	198
X4	75	183



$$weight = b_1 * height + b_0$$

$$Where: b_1 = slope, b_0 = bias/intercept$$

$$avg(weight) = 0 * height + avera$$

$$underestimation = 0.02 * height + b_0$$

$$overestimation = 3.7 * height + b_0$$

$$closestestmiation = 0.63 * height + b_0$$

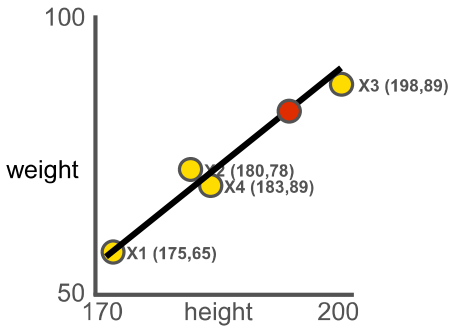
- The machine goes about finding the parameters that describe the function of a line that "**bets fits**" to the data
- This line is a linear function, and it is often referred as "**the model**" or "**the hypothesis**"
- The process of finding the line is called "**line fitting**", "**model fitting**", "**linear regression**" or more broadly "**model training**"
- Regressions are very useful to generate predictions (often referred as **inference**)
- Regressions are also useful in data interpolation (estimating missing values)

Mathematical models can be used to express relationships between variables

People	Weight	Height
X1	65	175
X2	78	180
X3	89	198
X4	75	183

Dependent variable Target

Independent variable



- Machines learn from known examples to create the model. This is called **Supervised Learning**
- Datasets with known target values are called "**labeled data**"
- Labeled data sets used to train the model are called **training set**
- Unlabeled data sets used in inference are called **test sets**

In regression we are often trying to estimate the value of one value from an other.
To do that, machines need to learn from actual examples. So we use some known data, with known relationships so machines learn from it
This is called the training set.

Regressions have a long list of applications

- Weather
- Finance
- Science
- Self-driving vehicles
- Intelligent health
- Computer vision
- Speech analysis
- … you name it

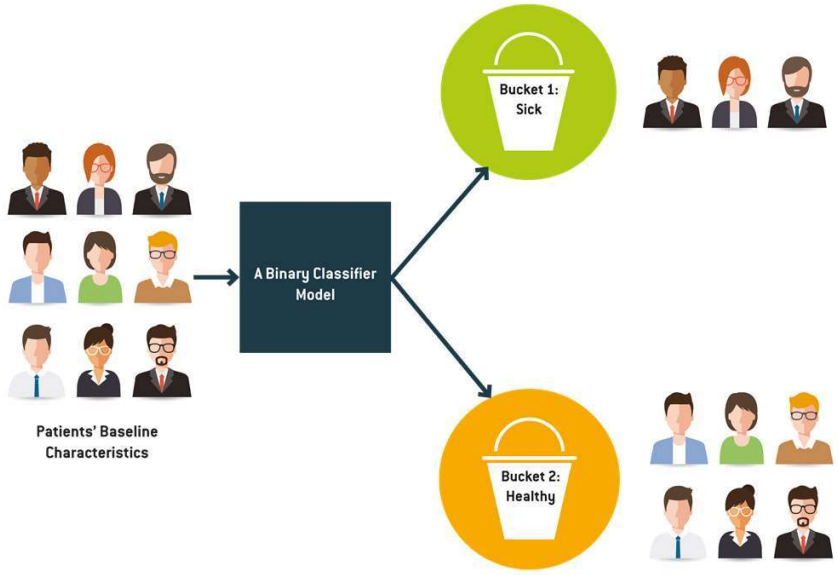
OSLOMET

DONT PANIC.

There is more…

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Machines can learn to classify observations into groups



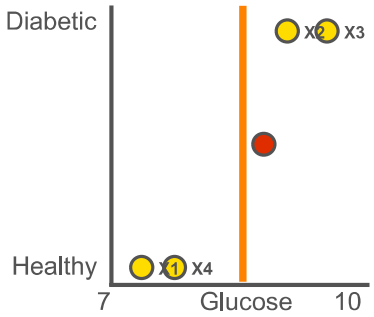
Collection of propositions is your knowledge base
The Knowledge base can be combined with data to produce answers. For instance, that is one way how expert systems and search engines may work.
They have a set of rules pre-programmed on how to operate a search.
Depending on the data (the database and the input keywords) the algorithm will apply the rules to give you the result.
This works well for established and well defined processes that do not change much.
This is the fundamental part of expert systems.

Building such a set of instructions is called programming.
To do so, you need some fundamental elements like
Values: that can describe the world for you. Using Names for categories of things and Numbers to express magnitudes and amounts.
Variables: to hold in memory the values while you perform the operations and actions
Actions: ways to transform values, like adding , subtracting…
Decisions: ways of comparing values and take one set of actions or the other depending on what is observed
Repetitions: ways to repeat a set of actions and value transformations many times
And Abstractions: That can encapsulate many small

detailed actions into a set of actions that can be understood as an action.

Classification requires labeled data. One label per category

People	Glucose	Diabetic	Diabetic
X1	7.6	No	0
X2	9.2	Yes	1
X3	8.5	Yes	1
X4	7.7	No	0



In a structured data set (or table) values are organized in this way typically. Observations as rows (that would be individuals, time, sales, etc) and the columns are the different variables or features. How you name your observations or your variables is irrelevant.

Suppose you wanted to target a promotion for a new digital camera and would like to know which of your customers are likely to buy the camera. Classification is a data mining technique that is useful for this application. Classification divides data into two or more well-defined classes. Unlike clustering in unsupervised learning, where you do not know which groups will be generated, in classification you know exactly what each group represents. In the previous

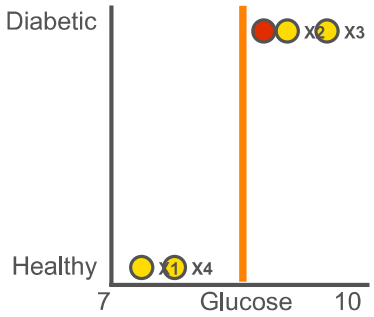
example, the two groups are: customers who are likely to buy a camera and customers who are not likely to buy a camera. This is an example of supervised learning.

In order for classification to work well, the build data must contain enough samples for each target category; otherwise, it may not be accurate. In other words, your build data must include enough people who have bought digital cameras in the past and enough who have not.

A labeled dataset allows you to classify different data points into groups, such as diabetic or healthy. Then based on the parameter you are assessing, here glucose levels, you can generate a model that distinguishes between the groups based on the parameter. Essentially this is drawing a line along the dimension of the parameter which divides the inputs into their labeled groups. This is the training step of the model, where you use data where you already know what groups the different parameters correspond to and train the model to generate this distinction.

Classification requires labeled data. One label per category

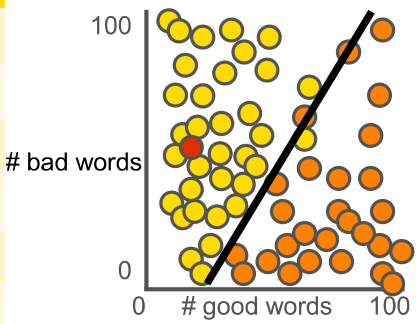
People	Glucose	Diabetic	Diabetic
X1	7.6	No	0
X2	9.2	Yes	1
X3	8.5	Yes	1
X4	7.7	No	0



Once the model is trained and you have a way to classify you can apply a test dataset which is also labeled to see how well the model performs. Based on the performance during the test set you can adjust the model to better distinguish between the groups.

Multiple features can be used to classify the data

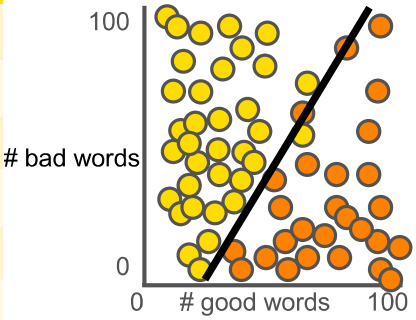
Movie	# good words	# bad words	Is a good movie
X1	7	93	0
X2	76	24	1
X3	49	50	1
X4	18	82	0
…	…	…	…
Xn	87	13	1



Only once the model is both trained and tested can you apply it to new data. As you can see this all requires a lot of structured and labeled data.

Multiple features can be used to classify the data

Movie	# good words	# bad words	Is a good movie
X1	7	93	0
X2	76	24	1
X3	49	50	1
X4	18	82	0
…	…	…	…
Xn	87	13	1



- Classification is about finding boundaries between categories
- Classifications are rarely perfect.
- It is important to compare the **performance** of the model with the performance of **experts** and **chance**
- For simplicity we used a line. But the **classification boundary** can be a curve or a plane

Data mining using classification usually involves a testing phase to check how good the model is. For this, data where the outcome is known is tested to see how well the model's predictions match it. For instance, you would take data for customers who have bought a digital camera in the past and check it against the predictions given by the model.

Testing a model involves computation of a structure known as the **confusion matrix**. A confusion matrix tells you how many times the model's prediction matched the actual data and how many times it did not. The columns correspond to the predicted values and the rows to the actual values. For instance, in Figure 16.4, the model was correct 555 + 45 = 600

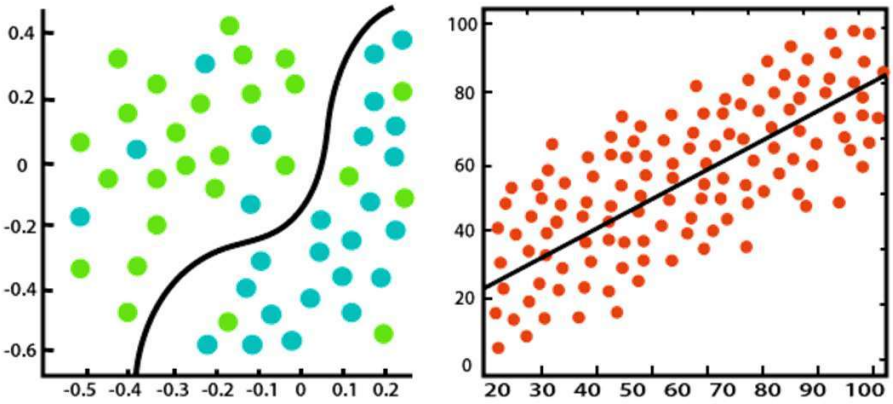
times and wrong 12 + 8 = 20 times. This shows that this model is a pretty good one.

In classification, you first analyze a small part of your data to build a model. For instance, you would analyze real data for people who have bought digital cameras and people who have not bought digital cameras, over a given time period. The data used to build a model is known as build data. The model will be built taking into account various factors, such as age, income, and occupation, that are known to influence people's buying habits. These factors are known as predictor attributes. The output that is predicted is called the target attribute and its values (whether the person will buy the camera or not) are known as categories or classes. Once the model has been generated, it can be applied to other data to come up with a prediction. This is known as model apply, or scoring. In our example, you would use the model to predict whether a certain customer is likely to buy a digital camera.

In the previous example, the target attribute has two values: will buy a digital camera and will not buy a digital camera. You can also use classification to predict attributes with more than two values—for example, whether the risk of a person defaulting on a payment is low, medium or high.

Classification is often used to create customer profiles. For instance, once you have determined which of your customers are likely to buy a digital camera, you can then

profile them by occupation, as shown in Figure 16.3. From this graph, you now know that most likely buyers are either engineers or executives. So you can now target your promotions more accurately toward these customers and reduce your costs.



Classification Regression

OSLOMET

Supervised learning is about finding functions that fit the values or the boundaries between groups in labeled data

EXERCISES

- Identify use-cases, activities, work processes or task that you believe that could employ supervised machine learning
- Describe the data that you have available:
  - What structure data you have there?
  - What unstructured data you have there?
- What kind of supervised machine learning methods would you use for every task? Classification or Regression?
- What is the insight you want to get from the data?
- What are the challenges you anticipate?