Intro to AI

- [x] ~~https://github.com/DAVE3625-22H/Lab-1~~
- [x] ~~https://github.com/DAVE3625-22H/Lab-2~~
- [x] ~~https://github.com/DAVE3625-22H/Lab-3~~
- [ ] https://github.com/tintun2602/Lab-4
- [ ] https://github.com/DAVE3625-22H/Lab-5

▼ **Different kinds of Artificial Intelligence**

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 pattern from experience, trial and error, reward policy

DL: Deep Learning

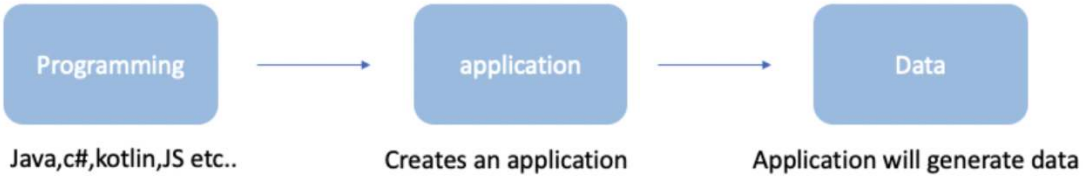
Generative Algorithm: Exploit mutation and selection to find a good solution to the problem at hand

▼ Data

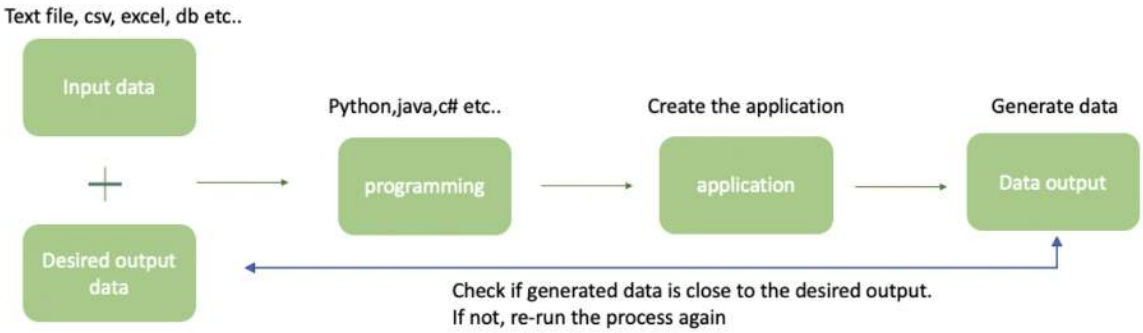
Steps to design an AI system

1. Identify the problem
2. Prepare the data
3. Choose the algorithms
4. Train the algorithms with the data
5. Run on a selected platform

General software development



AI based software development



> **Benchmark**: A standard or point of reference against which things may be compared.

Data pitfalls (problems which can occur with data)

Assuming the data is clean

- e.g spelling mistakes

Outliers

- Excluding outliers
- Including outliers

Ignoring seasonality

- Easter vacations, summer holidays, black Friday etc.

Context is critical

- Ignoring size when reporting growth

1.Data labeling / annotation

In machine learning, data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it.

2.Data annonymization

Data anonymization is the process of protecting private or sensitive information by erasing or encrypting identifiers that connect an individual to stored data.

3.Synthetic data

Synthetic data is artificial data that is generated from original data and a model that is trained to reproduce the characteristics and structure of the original data. This means that synthetic data and original data should deliver very similar results when undergoing the same statistical analysis.

4.Data preperation (cleansing + feature engineering)

Feature engineering, also known as feature creation, is the process of constructing new features from existing data to train a machine learning model.

Character recognition

- Features may include histograms counting the number of black pixels along horizontal and vertical directions, number of internal holes, stroke detection and many others

Speech recognition

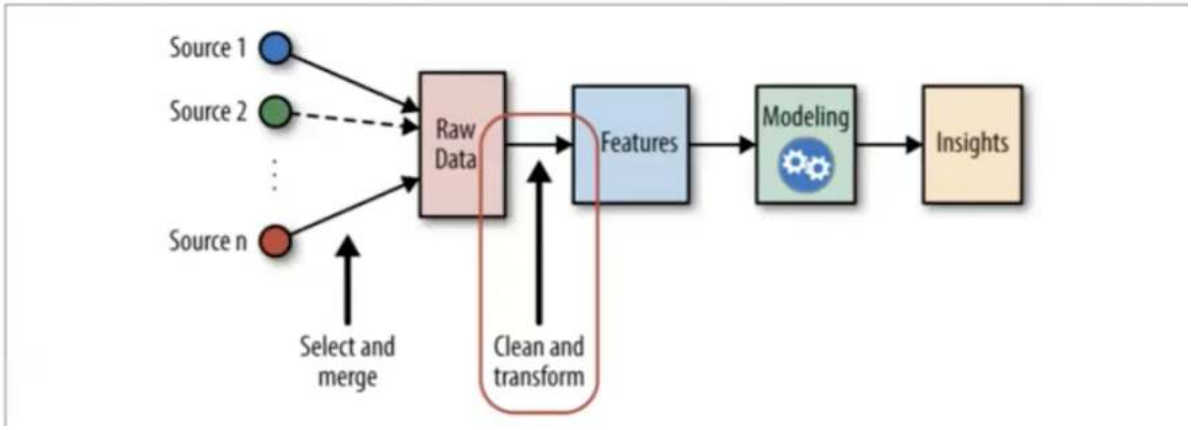
- features for recognizing phonemes can include noise ratios, length of sounds, relative power, filter matches and many others.

Spam detection

- features may include the presence or absence of certain email headers, the email structure, the language, the frequency of specific terms, the grammatical correctness of the text.

Computer vision

- there are a large number of possible features, such as edges and objects.



▼ **Definitions**

> ***Clustering*** *is a Machine Learning technique whose aim is to group the data points having similar properties and/or features, while data points in different groups should have highly offbeat properties and/or features*

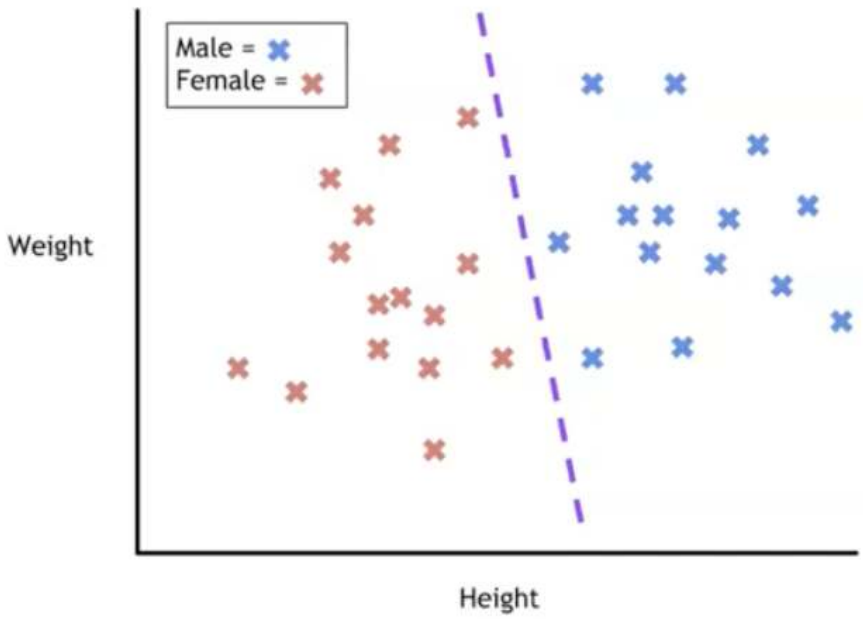
> ***Capacity prediction*** *is the 'class of the problems related to the prediction of when in the future the capacity of an existing system will become insufficient to process the installation's workload with a given level of performance'*

> ***Explainable AI (XAI)*** *is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms.*

▼ Supervised Machine Learning

There are two types of supervised learning:

1. Classification: Divided the data in classes / categories



1. Support vector machines
2. Decision trees
3. K-Nearest Neighbor
4. Random Forest

5. Logistic regression
6. Naive Bayes Classifier

Use cases:

- Spam detection (classification: spam or normal)
- Analysis of the customer data to predict wether they will buy computer accessories (Classification: Yes or No)
- Classifying fruits from features like colour, taste, size, weight (classification: Apple, Orange, Cherry banana)
- Gender classification from hair length (Classification: Male or female)
- Stock market price prediction)

2. Regression
    a. Linear regression
    b. Polynomial regression

▼ **In-depth of supervised machine learning algorithms**

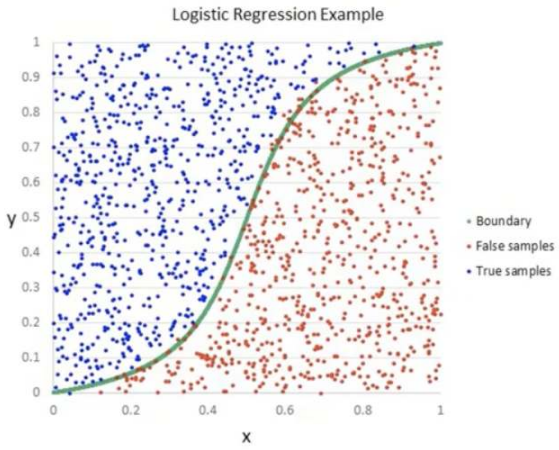
Classification vs Clustering algorithms

Clustering groups similar kind og things together. However, classification predicts similar kind of things together based on the historic trained data.

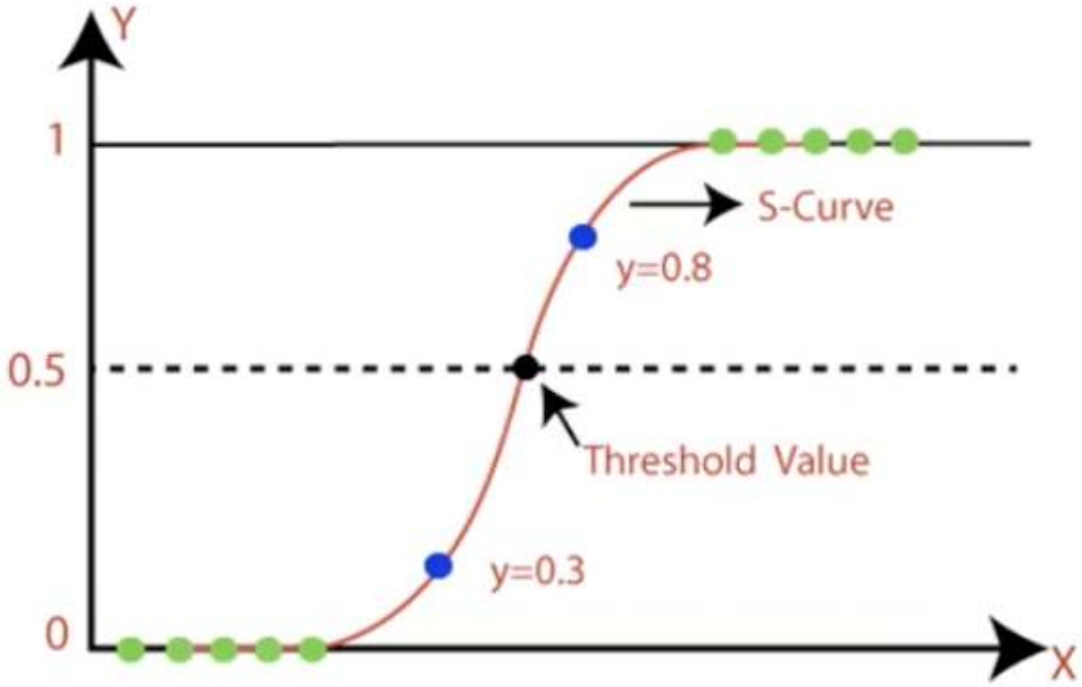
Classification algorithm: Logistic regression

Logistic regression is used when you have a classification problem - [yes / no, pass / fail, win / lose … etc.]

It is the go-to method for binary classification problems. The logistic function looks like a big S and will transform any value into the fange 0 to 1.



Sigmoid function: Allows to put a threshold value. e.g 0,5 (use case spam detection)



USE-CASES

Logistic regression is used to predict the occurance of some event, e.g.

- Predict wether it will rain or not
- Spam detection, diabetes prediction, cancer detection

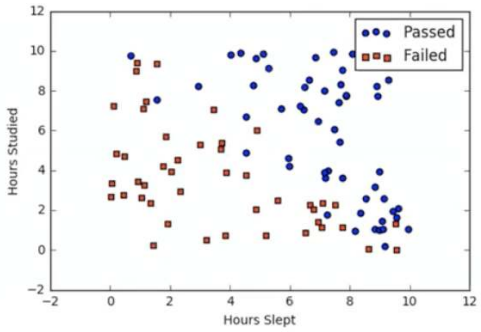
Types of Logistic regression:

- Binary logistic regression (e.g pass / fail)
- Multiclass logistic regression (e.g cats, dogs, sheep)
- Ordinal (low, medium, high)

Step 1: Visualize

Step1 : Visualize

Studied	Slept	Passed
4.85	9.63	1
8.62	3.23	0
5.43	8.23	1
9.21	6.34	0

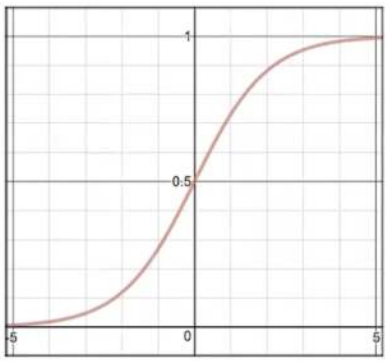


Step 2: Sigmoid function

$$S(z) = \frac{1}{1 + e^{-z}}$$

Code

```
def sigmoid(z):
  return 1.0 / (1 + np.exp(-z))
```

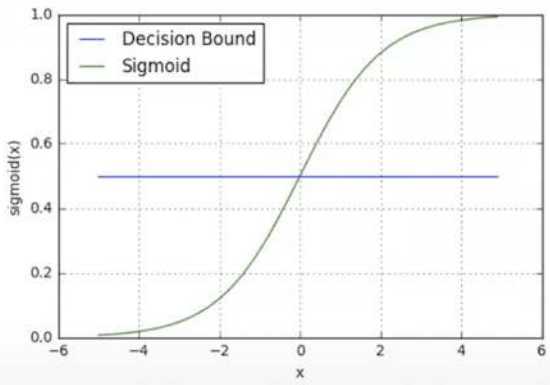


$s(z)$ = output between 0 and 1 (probability estimate)
z = input to the function (your algorithm's prediction e.g. mx + b)
e = base of natural log

Step 3: Decision boundary

- We need a threshold value

$$p \geq 0.5, class = 1$$
$$p < 0.5, class = 0$$



Step 4: Making predictions

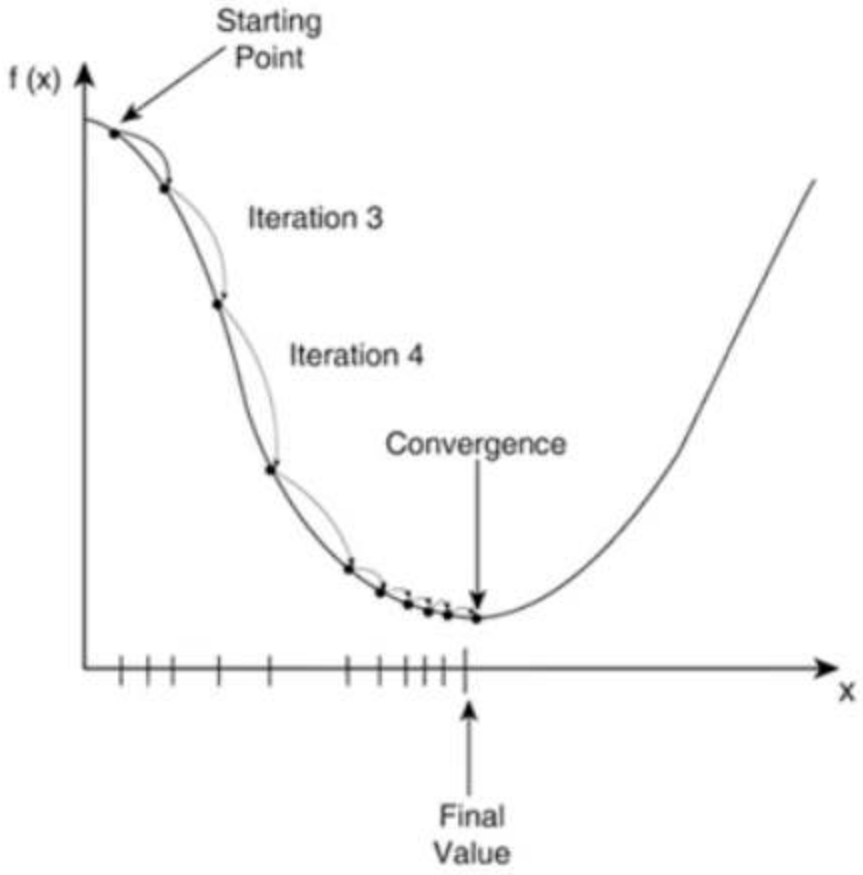
- Using our knowledge of sigmoid functions and decision boundaries, we can now write a prediction function.
- A prediction function in logistic regression returns the probability of our observation being positive, True, or "Yes".
- We call this class 1 and its notation is $P(class=1)$.

$$P(class = 1) = \frac{1}{1 + e^{-z}}$$

- If the model returns .4 it believes there is only a 40% chance of passing. If our decision boundary was .5, we would categorize this observation as "Fail.""
- Code it

Step 5: Gradient Descent

Gradient Descent is an optimization algorithm used to tweak parameters iteratively to minimize the cost function. The algorithm enables a model to learn the gradient or direction that the model should take in order to reduce errors (differences between actual Y and predicted Y)



Step 6: Map probabilities to classes

The final step is assign class labels (0 or 1) to our predicted probabilities

Decision boundary

```
def decision_boundary(prob):
  return 1 if prob >= .5 else 0
```

Convert probabilities to classes

```
def classify(predictions):
  '''
  input  - N element array of predictions between 0 and 1
  output - N element array of 0s (False) and 1s (True)
  '''
  decision_boundary = np.vectorize(decision_boundary)
  return decision_boundary(predictions).flatten()
```

Example output

```
Probabilities = [ 0.967, 0.448, 0.015, 0.780, 0.978, 0.004]
Classifications = [1, 0, 0, 1, 1, 0]
```

Step 7: Finish it up

- Train the model
- Evaluate the model
  - Minimize the cost with repeated iterations
- Measure the accuracy of your outputs
- Measure the probability score

Naive Bayes Classification Algorithm

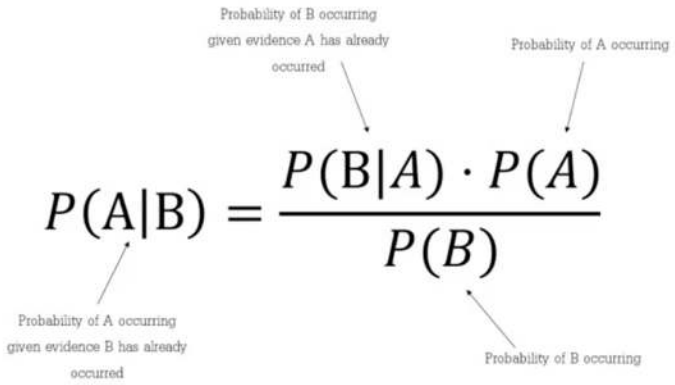
A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

- e.g. fruit may be considered to be an apple if it is red, round and about 3 inches in diameter
- Its easy to build algorithm and a very powerful one
- Does not take into account the relationship between features

(1) Bayes Theorem

Bayes`s theorem descibes the probability of an event, based on prior knowledge of conditions that might be related to the event.

Bayes's theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event



Natural Language Processing

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

(2) Naive Bayes use in Natural Language Proccessing

Naive bayes will count 2 occurances for mood category and 0 for not mood category.

Hence classifying the text as mood.

TEXT	CATEGORY
I am so angry.	mood
I feel like superstar.	mood
It is going to rain.	not mood
I want to cry.	mood
They will come together.	not mood

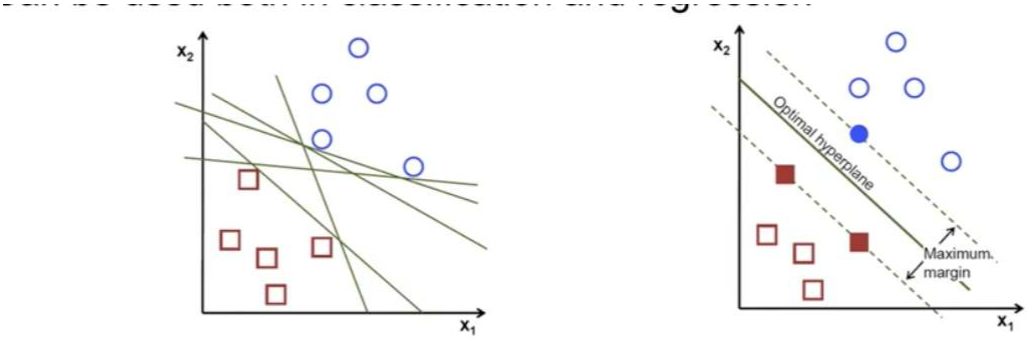
or

I	mood
feel	mood
so	
terrible	

Support vector machines (classification)

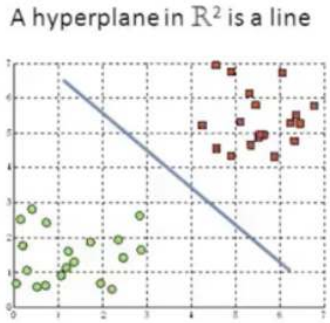
Main use is to classify unseen data. Support vector machines classifies the data by finding a clear seperation between the data points → it looks for hyperplane → can be used both in classification and regression

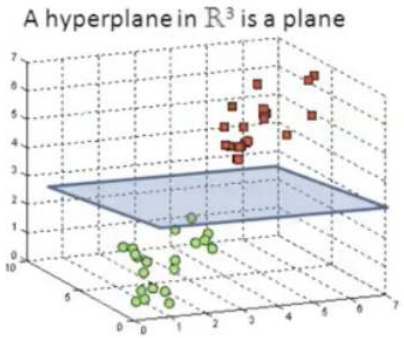
> **Hyperplanes are decision boundaries that help classify the data points**



We can use a 2 dimensional hyperplane (if number of features is 2) → or a 3 dimensional hyperplane (if number of features is 3). We can also have n dimensional hyperplane (for more features)

POLITAN UNIVERSITY
RSITETET





Use case of support vector machines

- Used to detect cancerous cells
- Used to predict driving routes
- Face detection
- Image classification
- Handwriting detection

OSLOMET

Pros and cons

- Pros
  - Effective on data sets with multiple features
  - Effective in cases where number of features is greater than the data points
  - Memory efficient
- Cons
  - They donot provide probability estimates. Those are calculated using an expensive five fold cross vsalidation
  - Works best on small sample sets because of its high training time

Regression

Regression models are used to predict a continuous value: e.g

- Predicting prices of a house given a set of features (size, price, location etc…)
- Predicting sales revenue of a company based on previous sales figures

.

Use cases:

Churn: Classification

Stock market price prediction: Classification

Spam email: Classification

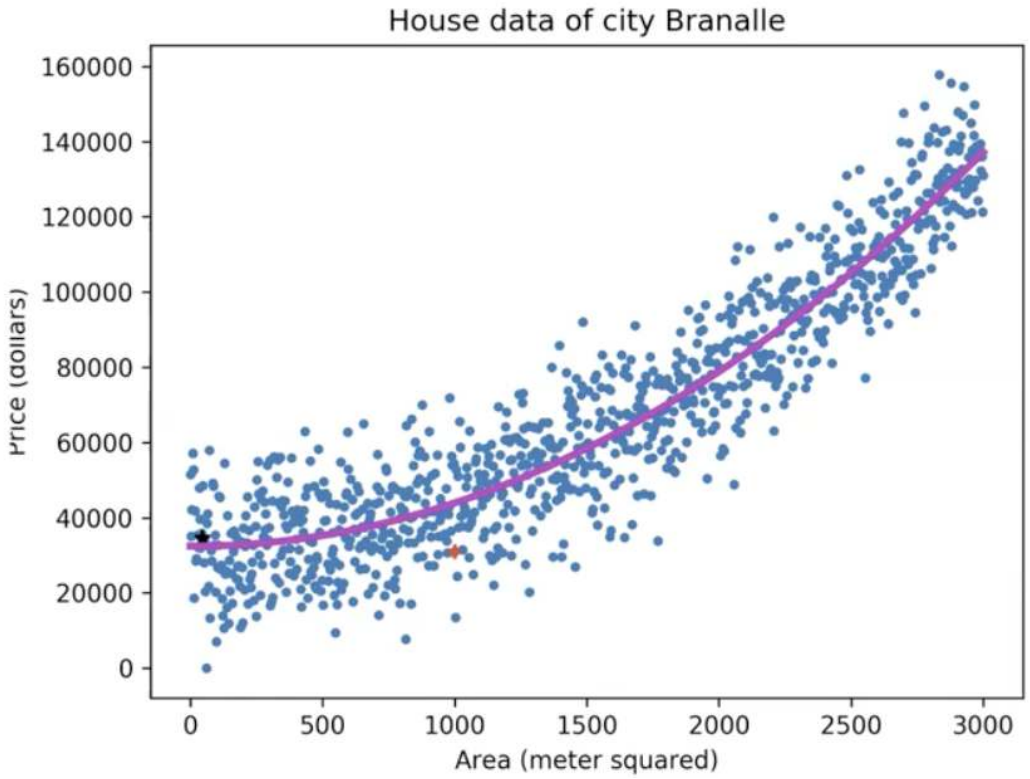
Prediction of price of an oil: Regression

Salary prediction: Regression

Age: Regression

Gender: Classification

E.g Regression

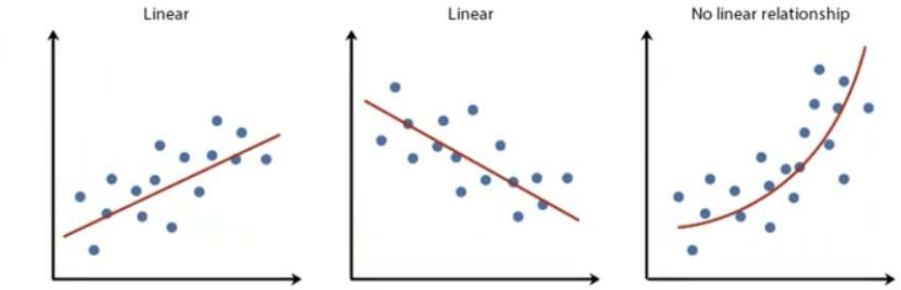


Difference between regression and classification algorithm

Regression Algorithm	Classification Algorithm
In Regression, the output variable must be of continuous nature or real value.	In Classification, the output variable must be a discrete value.
The task of the regression algorithm is to map the input value (x) with the continuous output variable(y).	The task of the classification algorithm is to map the input value(x) with the discrete output variable(y).
Regression Algorithms are used with continuous data.	Classification Algorithms are used with discrete data.
In Regression, we try to find the best fit line, which can predict the output more accurately.	In Classification, we try to find the decision boundary, which can divide the dataset into different classes.
Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc.	Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc.
The regression Algorithm can be further divided into Linear and Non-linear Regression.	The Classification algorithms can be divided into Binary Classifier and Multi-class Classifier.

Types of regression

- Linear regression
  - An equation that describes a relationship between two quantities that show constant rate of change → e.g the older i get, the wiser i will be → Attending all lectures in intro to ai course will result in passing the exam
  - There is always an input (X) and an output (Y) → Y can be calculated from a linear combination of input variables (X)
  - Example (1)

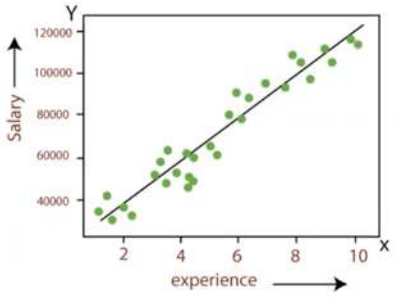


  - **Simple linear regression:** when there is a single input variable (x)
  - **Multiple linear regression:** multiple input variables
  - Example (2)

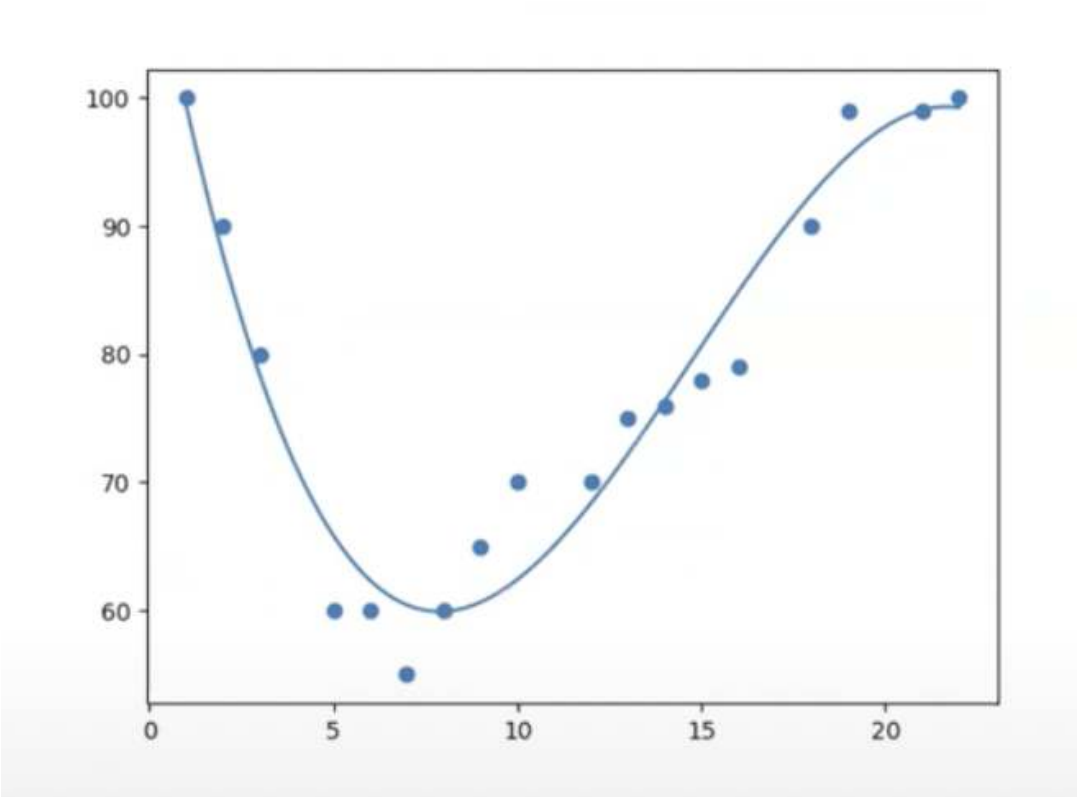
$$Y = a+bX$$

Y = Salary
X = Employees age
A,B: Coefficients of the equation

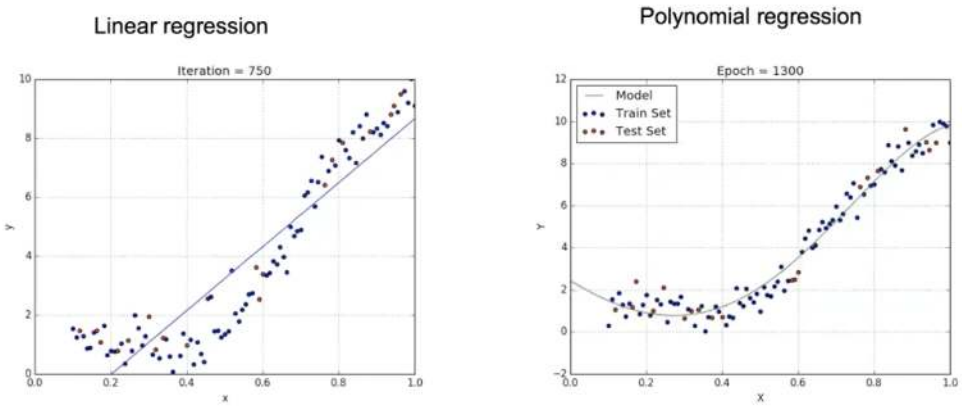
Coefficients are estimates of the unknown. These are calculated for a regression model and help us predict value of Y for each value of X



- Logistic regression
  - Also a linear model, however it is never a straight line
  - Example (1)



  - Example (2)



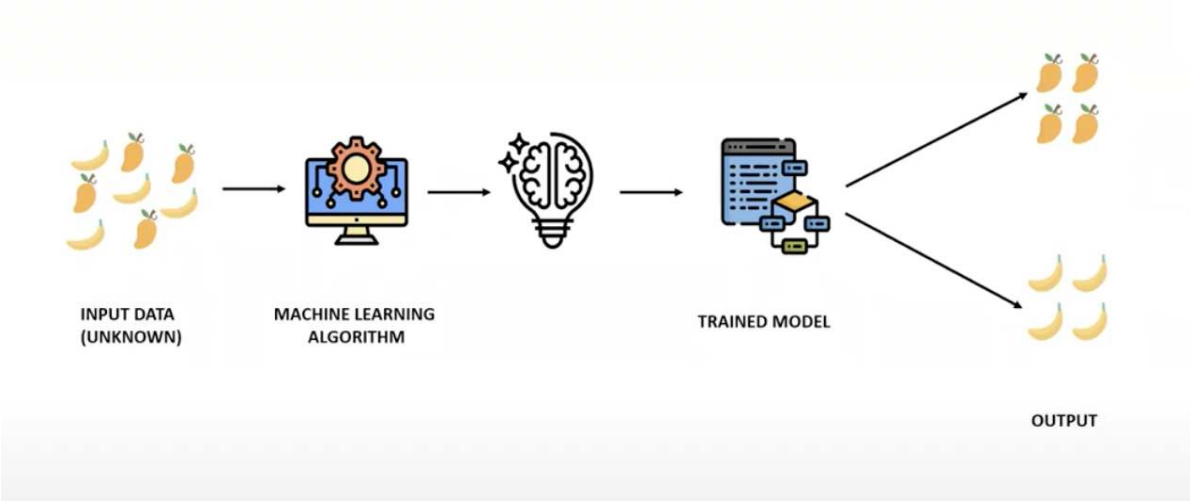
- Polynomial regression
- Support vector regression
- Decision tree regression
- Random forest regression
- Ridge regression
- Lasso regression

▼ **Unsupervised machine learning**

Definition

Unsupervised machine learning is a paradigm for problems where the available data consists of unlabelled examples, meaning that each data point contains features (covariates), only without an associated label. The goal of unsupervised learning algorithms is learning useful patterns or structural properties of the data.

Examples of unsupervised machine learning tasks are clustering, dimension reduction and density estimation.



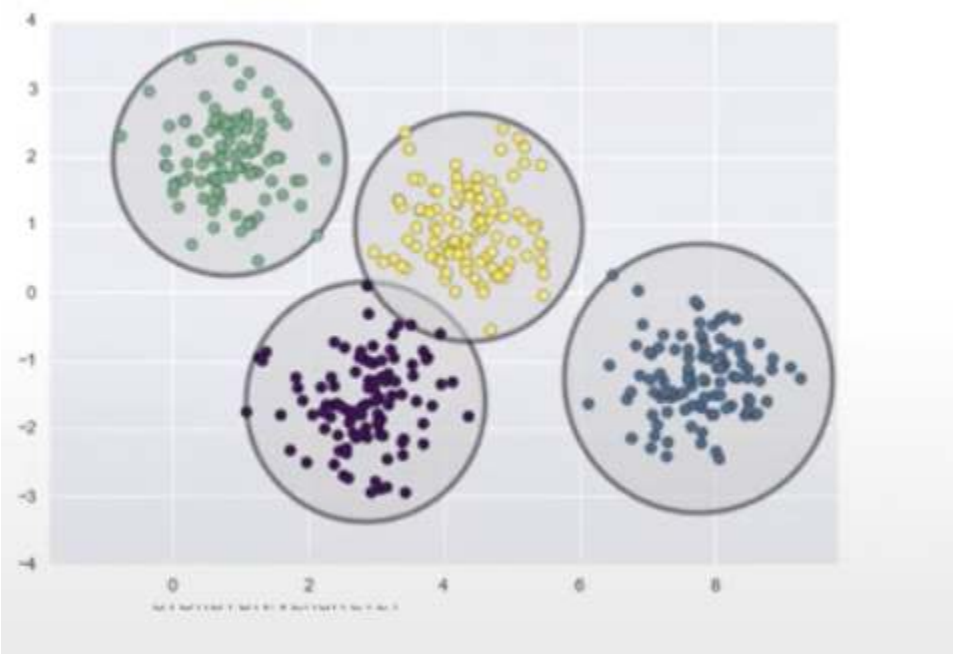
Types of unsupervised learning

Clustering

Clustering is when we want to discover groupings in data.

- **K-means clustering:** K-means is a distance based algorithm where we calculate distances between data points to assign to the cluster.
  - K-means clustering will stop when the centroids have been stabilized - there is no change in their values since clustering has been successful or the defined number of iterations have been reached → A centroid is the imaginary or real location representing the center of the cluster
  - Advantages of K-means: Very simple to run (choose k and run it a number of times) → “Most projects do not need quality sensitive clusters”
  - Uses of K-means:
    - Document classification
    - Customer segmentation
    - Fraud detection (insurance n bank)

- Ride share data analysis (uber etc.)
- Detection of anomalies
- Sorting sensor measurement



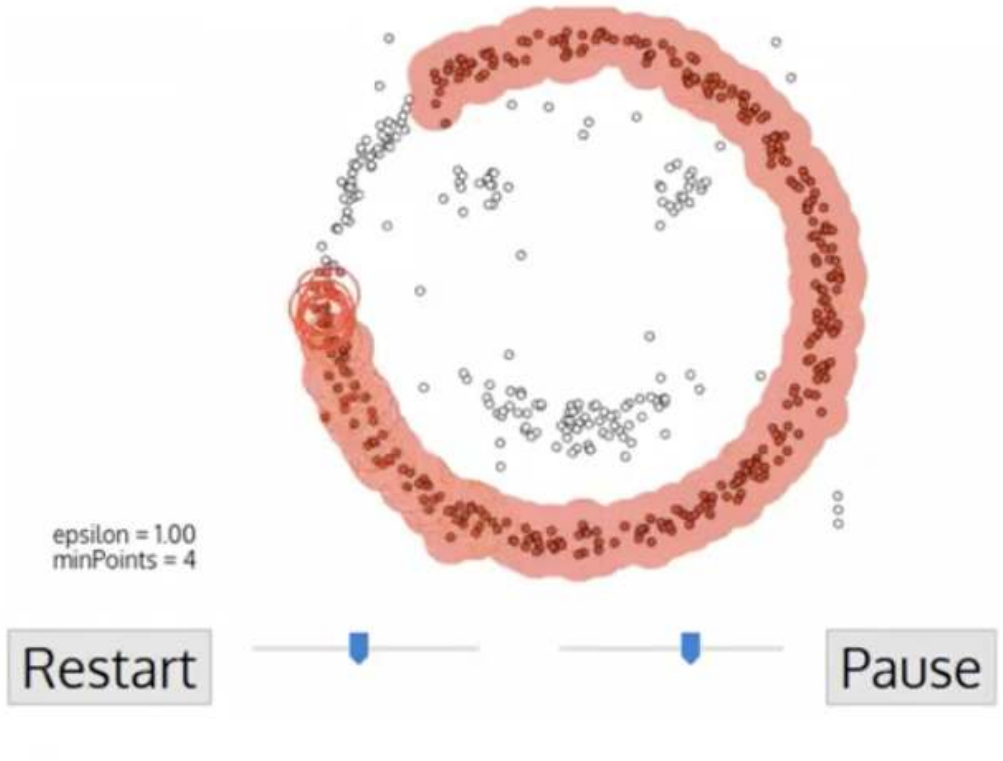
Mean shift clustering algorithm: The algorithm locates the heavy density clusters in a data → Traverses the datapoints and identifies the densest location

- Uses of Mean Shift Clustering:
  - Computer vision
  - Image processing

DBSCAN Algorithm

Stands for Density-Based Spatial Clustering of Applications with Noise. → This is also a density based algorithm.

It separates regions by areas of low-density so that it can detect outliers between the high-density clusters.

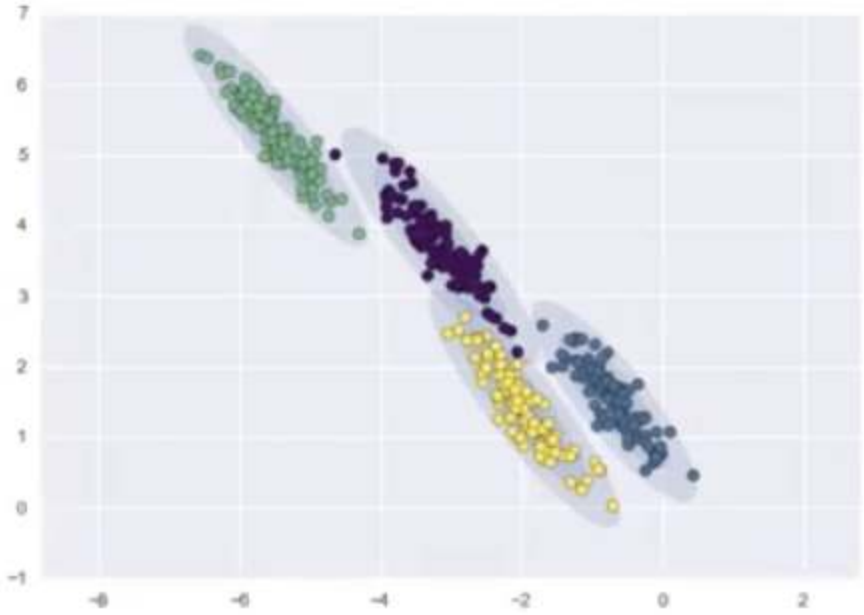


Uses two parameters:

- **minPts**: the minimum number of data points that need to be clustered together for an area to be considered high-density.
- **Eps**: the distance used to determine if a data is in the same area as other data points

Gaussian Mixture Model

This model is very similar to K-means however, K-means follows a circular format, meanwhile Gaussian can take on any format.

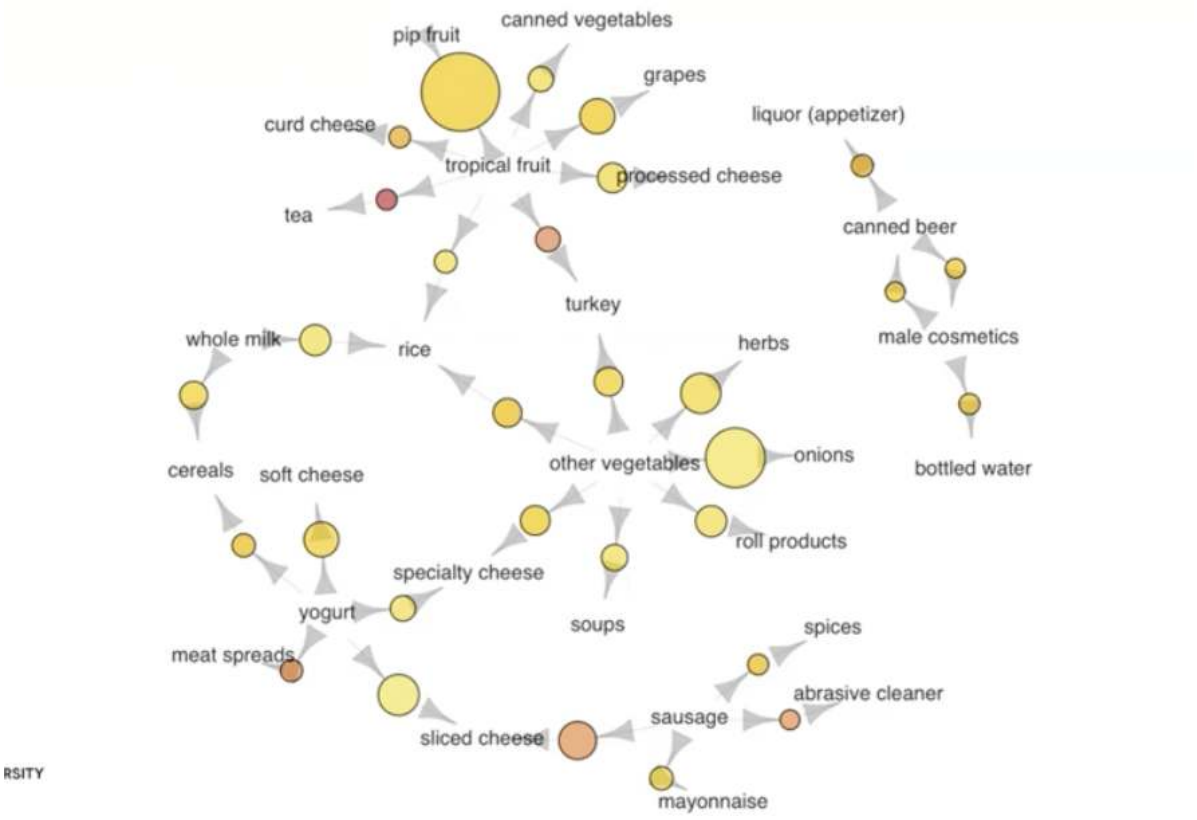


A few more clustering algorithms…

- BIRCH algorithm
- Affinity Propagation clustering algorithm
- OPTICS algorithm
- Agglomerative Hierarchy Clustering algorithm

Association

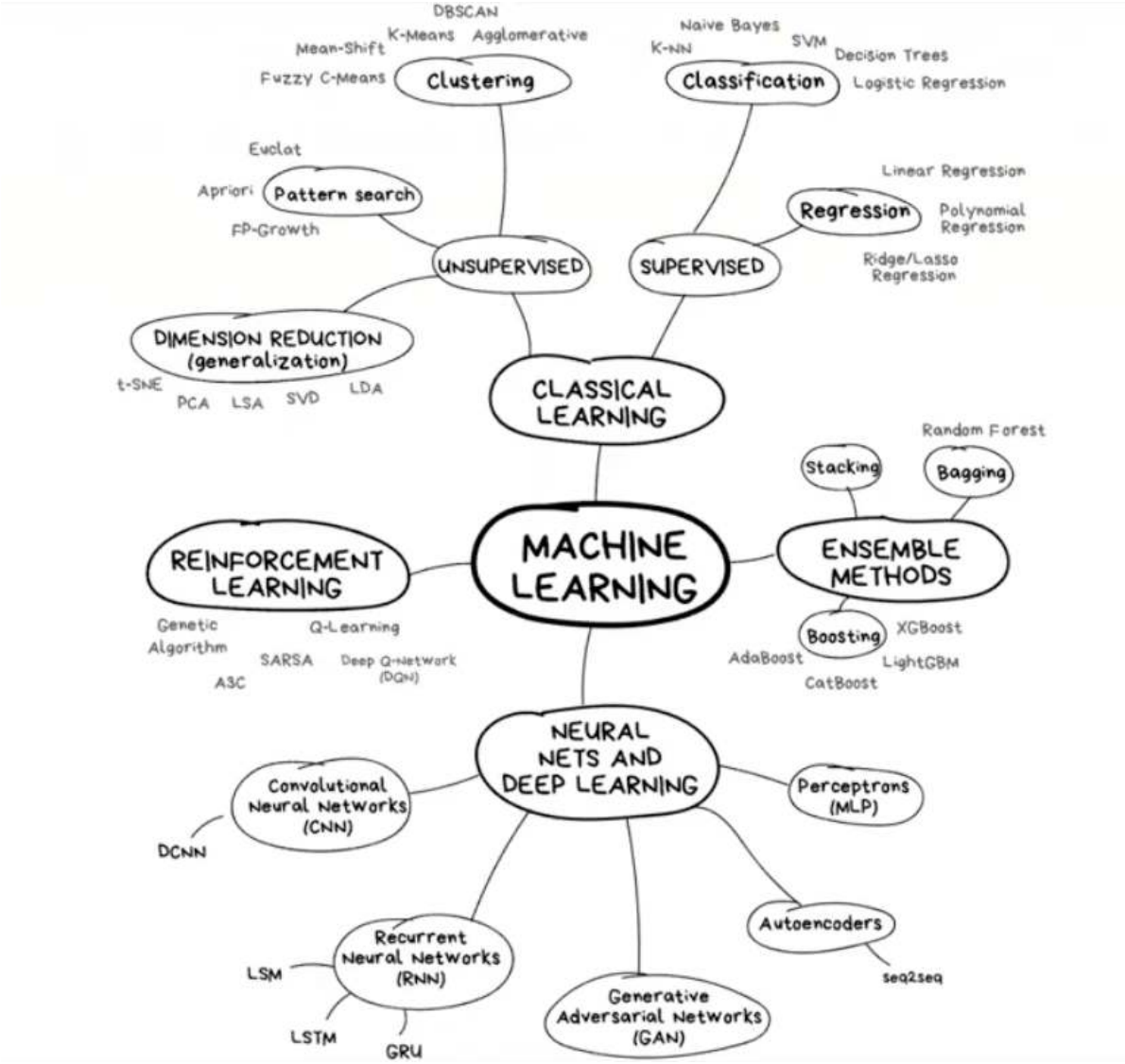
Association is when we want to discover rules that describe our data.



Apriori algorithm

This algorithms is used for mining frequent itemsets and devising association rules. It is created to operate on a database containing a lot of transactions, for instance, items brought by customers in a store.

→ This is the algorithms behind: “you may also like”


MACHINE LEARNING
CLASSICAL LEARNING
UNSUPERVISED
SUPERVISED
Clustering
DBSCAN
K-Means
Agglomerative
Mean-Shift
Fuzzy C-Means
Classification
Naive Bayes
K-NN
SVM
Decision Trees
Logistic Regression
Pattern search
Euclat
Apriori
FP-Growth
Regression
Linear Regression
Polynomial Regression
Ridge/Lasso Regression
DIMENSION REDUCTION (generalization)
t-SNE
PCA
LSA
SVD
LDA
REINFORCEMENT LEARNING
Genetic Algorithm
ASC
Q-Learning
SARSA
Deep Q-Network (DQN)
ENSEMBLE METHODS
Stacking
Bagging
Random Forest
Boosting
XGBoost
AdaBoost
LightGBM
CatBoost
NEURAL NETS AND DEEP LEARNING
Convolutional Neural Networks (CNN)
DCNN
Perceptrons (MLP)
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Autoencoders
seq2seq
Generative Adversarial Networks (GAN)

Term Paper

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/d50837a2-9060-4179-92b4-089699f2317a/Final_Project_209.pdf

Resources to Read

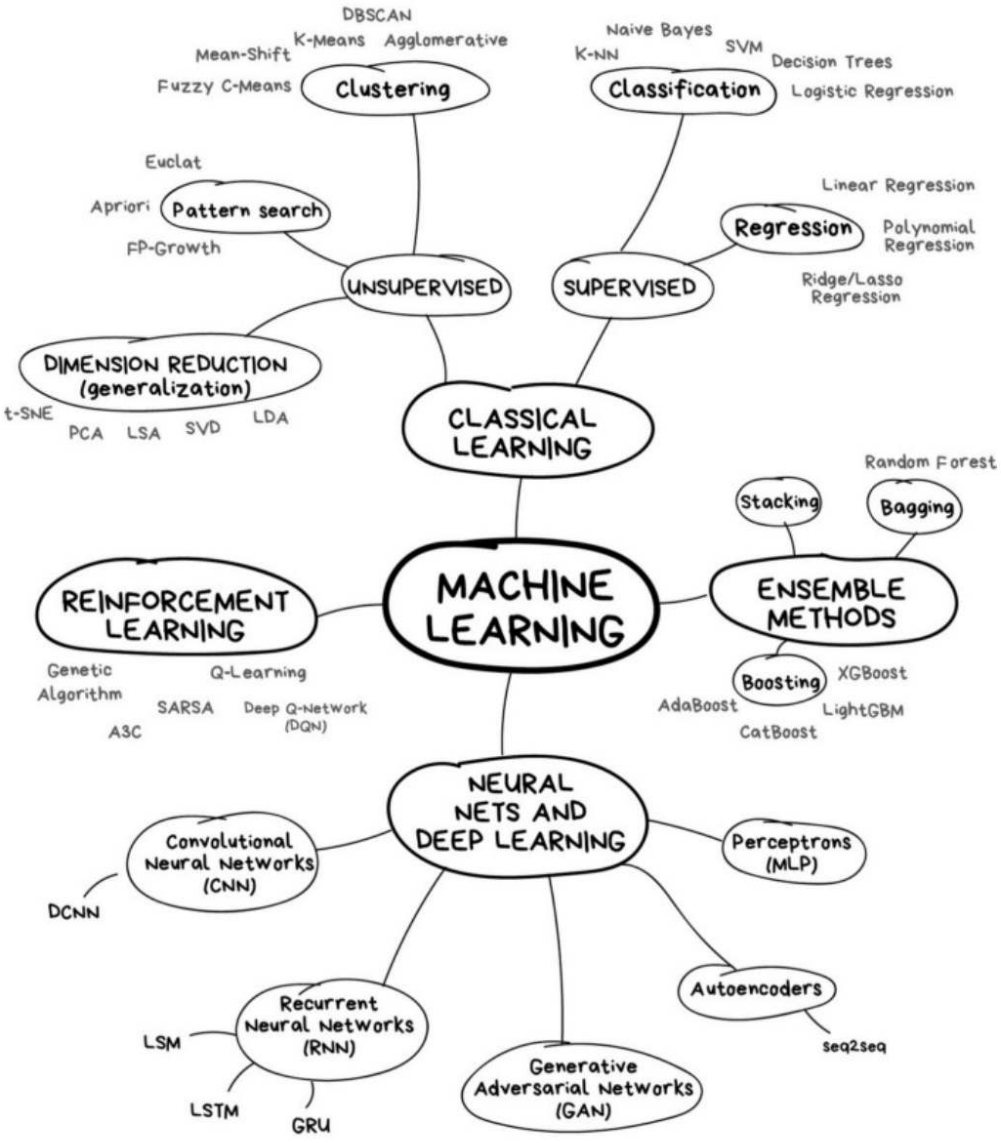
https://s3-us-west-2.amazonaws.com/secure.notion-static.com/a82e606d-4a0e-47d5-8350-4cd5c7934a11/Ng-MLY01-MachineLearning_Yearning.pdf

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/b33f07d4-fb8c-4230-a09c-0ec881cfb72b/DIKU_004_-_Supervised_Machine_Learning.pdf

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/36257cf8-ac30-40d1-ba0f-f377e70410d7/Zero_to_AI_Intro_and_History_of_AI.pdf

Previous Exam

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/f322b683-69ac-4a06-a639-58c51c900f03/Exam2021-DAVE3625_(1).xlsx



Lecture 1

What is artificial Intelligence?

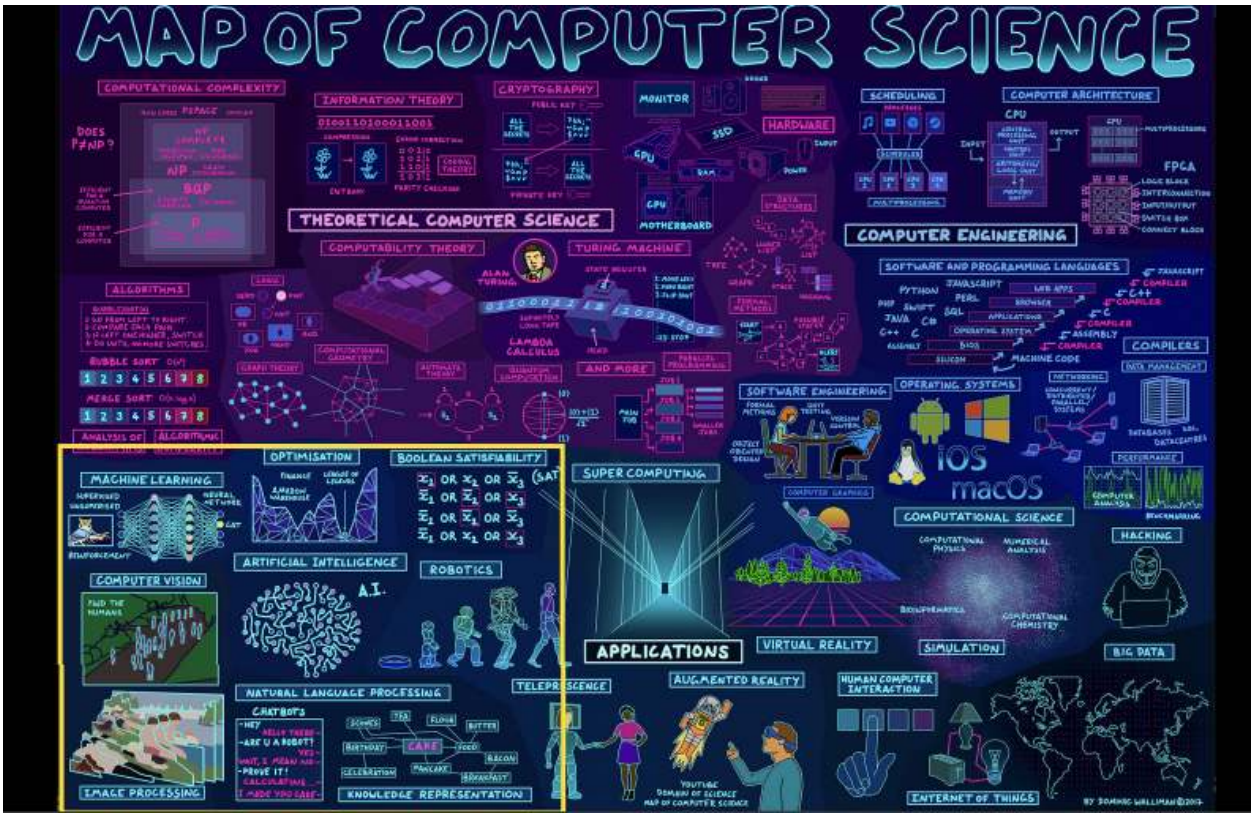
AI is when we artificially introduce intelligence in machines, thereby giving us:

- A machine which mimics humans like intelligence
- A machine which has decision making capabilities
- A machine which learns on its own

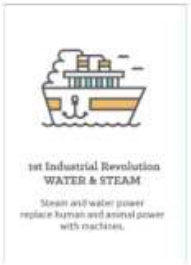
“Software that solves a problem without explicit human instruction”

Artificial Intelligence versus Conventional programming

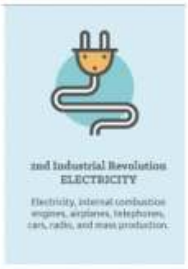
Conventional programming	Artificial Intelligence
Programmers look at the problem (desired output) and build an algorithm / application to solve this problem	AI programmers show the problem (desired output) to the AI algorithms and expect the algorithm to find a solution
A programmer has complete control over their application	An AI programmer can never claim to have full control over over their AI applications (. Explainable AI is hard achieve)
The software must follow a logical series of steps to reach a conclusion (hard coded instructions by the programmer)	AI applications use the technique of search and pattern matching
Its easy to explain a conventional algorithm	Its very hard to explain how an AI algorithm reached its desired output
The most important element here is the algorithm	The most important elements here are data and algorithms



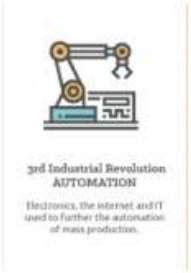
The 4th and biggest industrial revolution



1784



1870

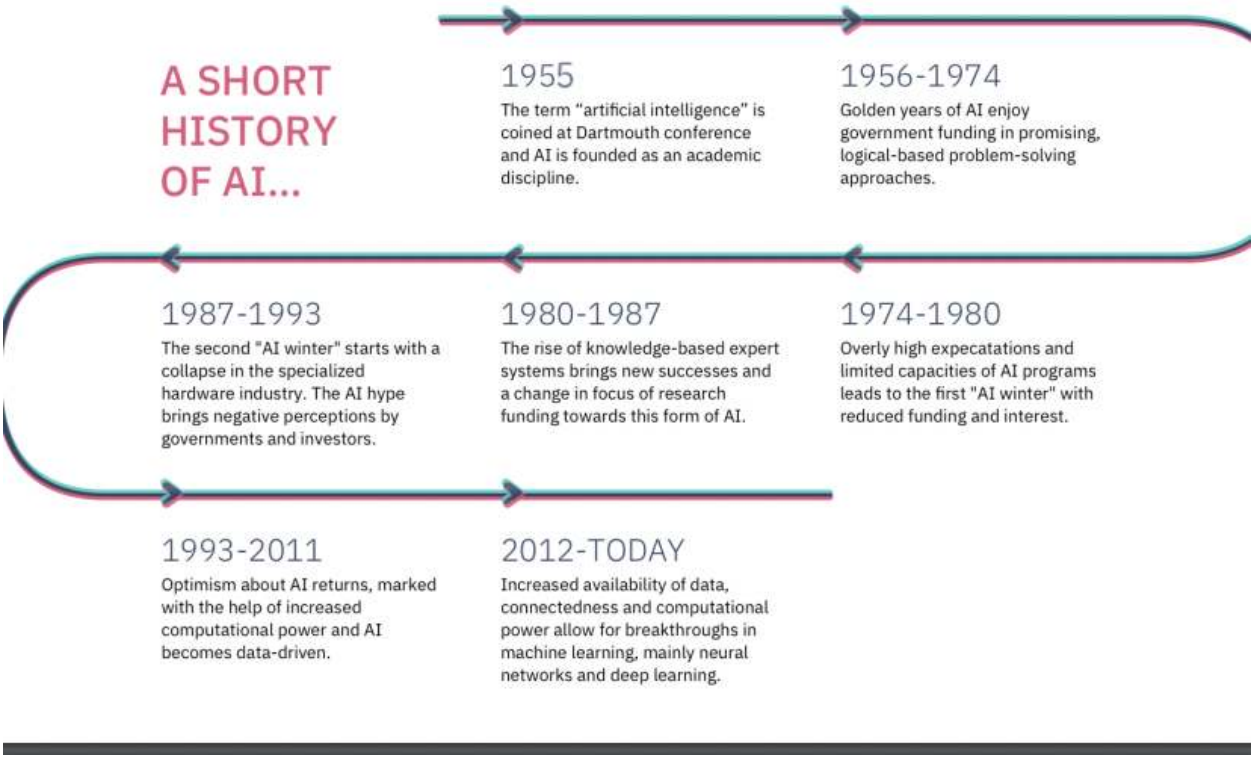


1969

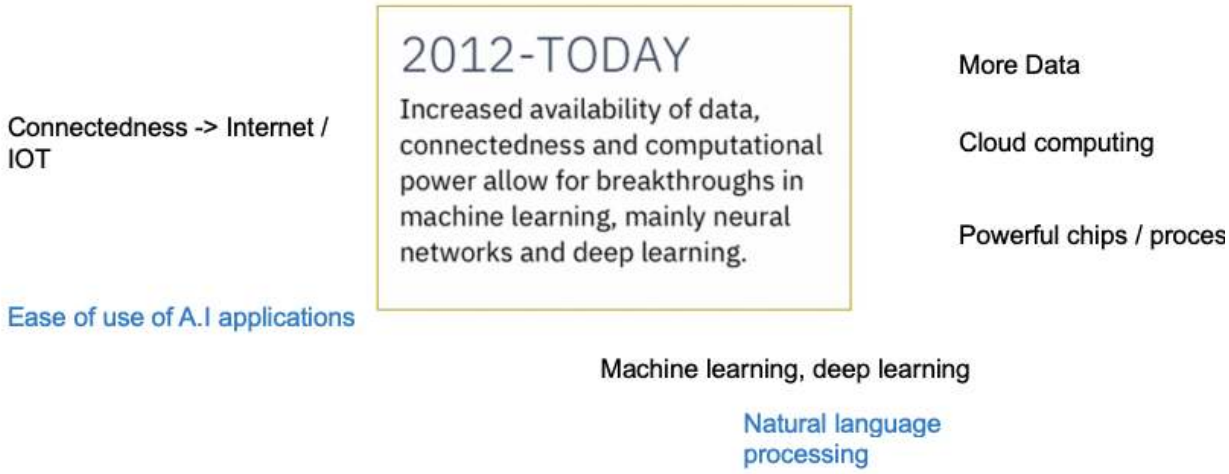
Now or future ?

History of AI - important points

- Efforts to create intelligent machines started as early as 1642
  - First mechanical calculating machine - Blaise Pascal
- Turing test was introduced in 1950
- In 1955 we coined them “Artificial Intelligence”



Why?



Can AI exist on its own?

Q: can you name a company which is purely an AI company?

AI helps other software services to do their job better e.g:

- Voice recognition and intelligent search used in AI assistants
- Natural Language Processing used in Translation services

Examples of AI in our daily lives

Self driving cars

Image / video analysis

OSLOMET

Voice and chat assistants

Examples of A.I in our daily lives

Content reccomendation

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

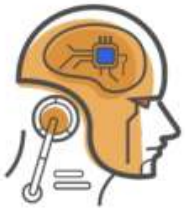
Types of Artificial Intelligence

Narrow AI

Dedicated to assist with or take over specific tasks.

General AI

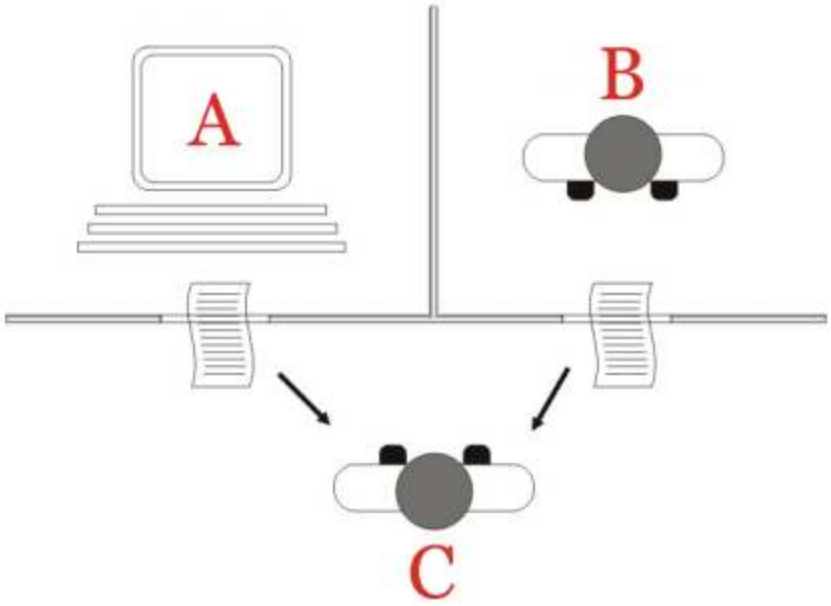
Takes knowledge from one domain, transfers to other domain.

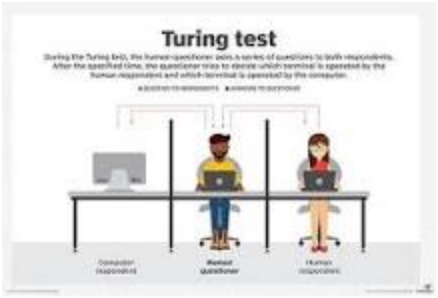


Super AI

Machines that are an order of magnitude smarter than humans.

Turing Test – General AI

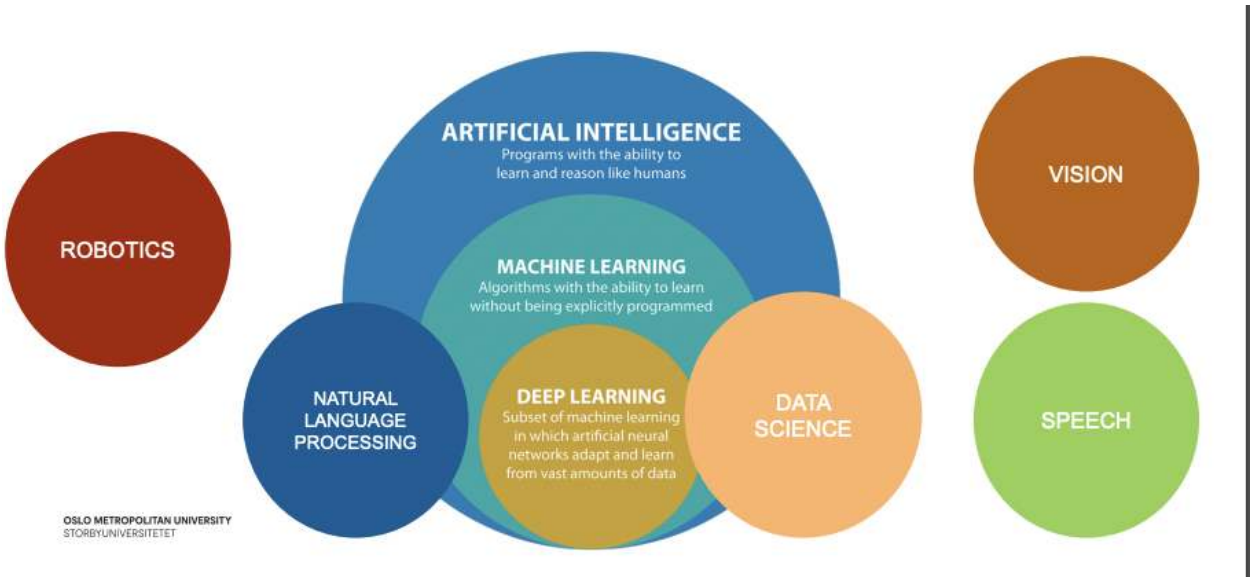




The Turing Test is **a method of inquiry in artificial intelligence (AI) for determining whether or not a computer is capable of thinking like a human being**. The test is named after Alan Turing, the founder of the Turing Test and an English computer scientist, cryptanalyst, mathematician and theoretical biologist.

→ Google Duplex passed the turing test in 2018

Branches of AI



Skillset needed to work with AI

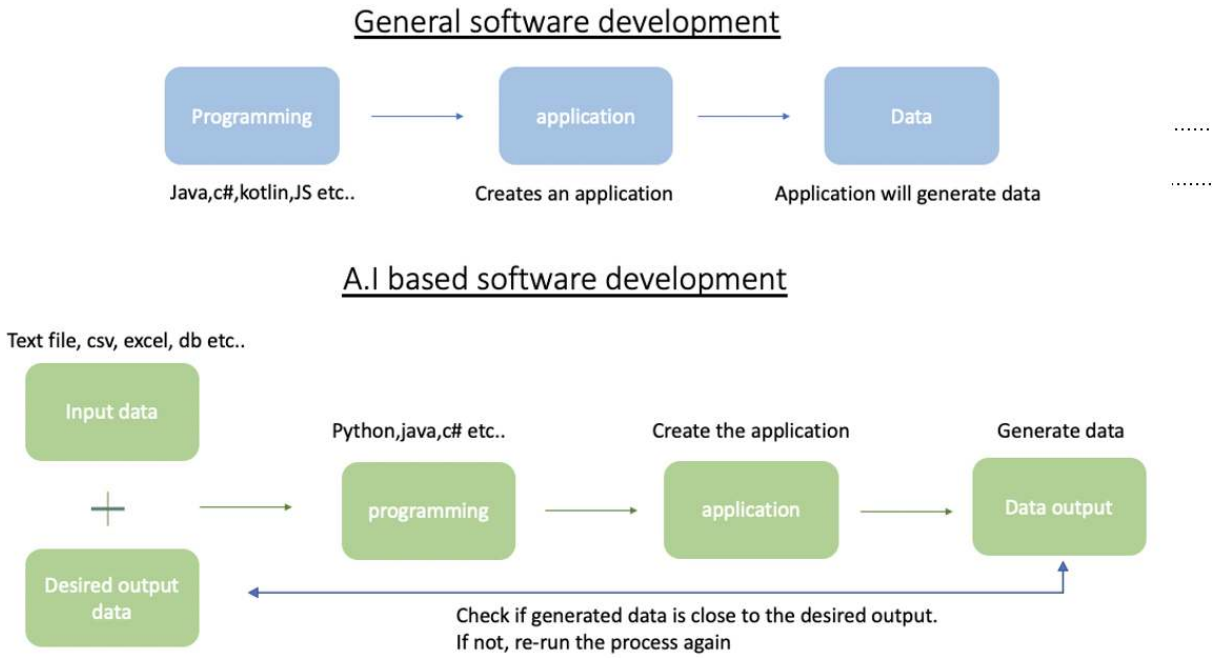
- Math
  - The theoretical background necessary to conduct and apply AI research
- Statistics
  - Empirical skills needed to fit and measure the impact of AI models
- Machine learning
  - Skills needed to build self-learning models like deep learning and other supervised models that power most AI applications today
- Statistical programming

    - Programming skills needed to implement AI models such as in Python and related packages, like sci-kit learn and pandas
- Software Engineering
    - Programming skills needed to design and scale AI powered applications

Lecture 2

Steps to design an AI system

1. Identify the problem
2. Prepare the data
3. Choose the algorithms
4. Train the algorithms with the data
5. Run on a selected platform



AI is about algorithms, tools and data

Algorithms and tools

DATA

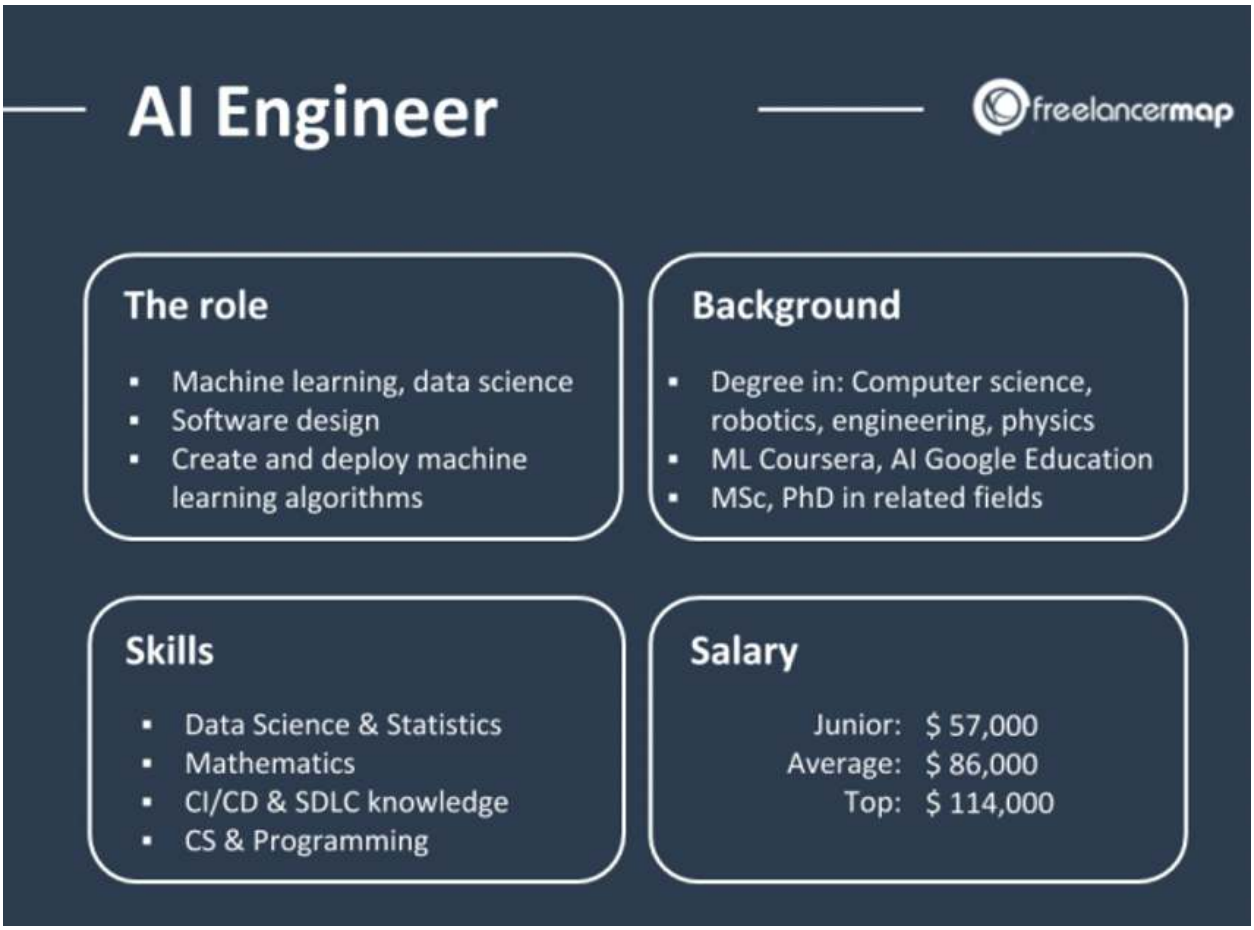
Open data sets. e.g Kaggle

Private data, e.g. any organizations data

Daily life of an AI programmer

- 80% time spend on data (cleaning, preparing, labeling, analyzing etc)
- 5% on deployment (cloud / on premise)
- 15% on AI development

This means there is a huge demand for data engineers. Organizations are not looking for AI engineers. They are looking for data engineers / software engineers who can also work with AI.



Data pitfalls (problems which can occur with data)f

- Assuming the data is clean

    - Spelling mistakes
- Outliers
    - Excluding outliers : Those 21 people using your product more than a thousand times a day? Either they´re your biggest fans, or bots crawling your site for content. Whichever they are, ignoring them would be a mistake
    - Including outliers : While those 21 people using your product a thousand times a are interesting from a qualitative perspective, because they can you things you didn´t expect, they can be problematic when building models. You probably want to exclude them when building data products; otherwise the “you may also like “ feature on your site will have the same items everywhere
    - In statistics, an outlier is a data point that differs significantly from other observations.

- Ignoring seasonality
    - Easter vacations, summer holidays, Black Friday : At Ruter we can assume that the passenger capacity will be the same on special seasonal days
- Context is critical
    - Ignoring size when reporting growth : When you´ve just started, technically, your dad signing up just doubled your user base
- Poor data insights : We need eyes into our data. Popular way to get these insights is create dashboards or reports
- Not connecting with external data: It helps to combine your data with external data, like social media data, weather data etc.
- Lacking business understanding : Its critical to know the business before working with the data. Now knowing the reason or the intention why the data was produced in the first place can mess up its desired output.

How to work with data? 👾

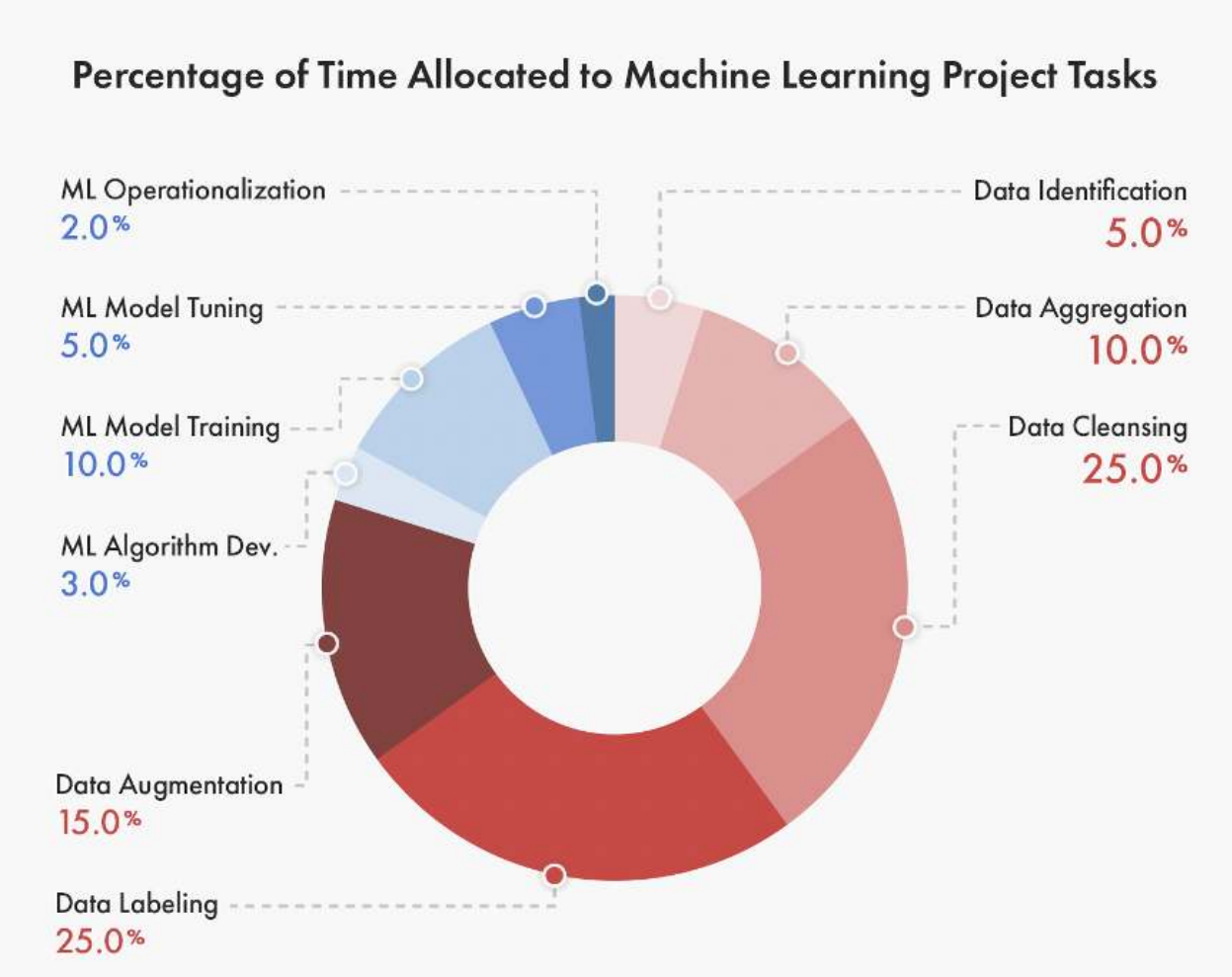
- Data labeling / annotation

- Data Annonmization
- Synthetic data
- Data preperation
  - Data cleansing + feature engineering
- Data wrangling
- Data mining
- Data warehousing
  - ETL programming
- Data Engineering
  - Infrastructure, big data, cloud etc..

1) Data labeling / annotation

OSLOMET

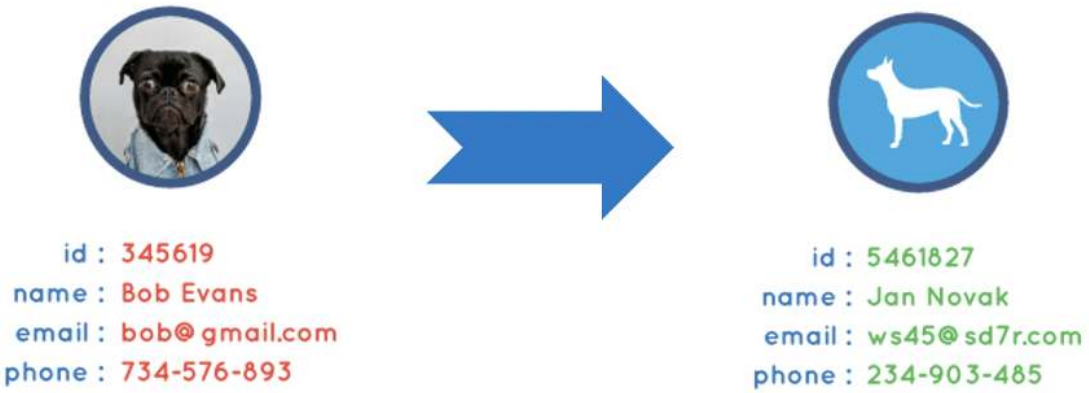
PIE CHART OF ALLOCATED TIME TO MACHINE LEARNING TASKS



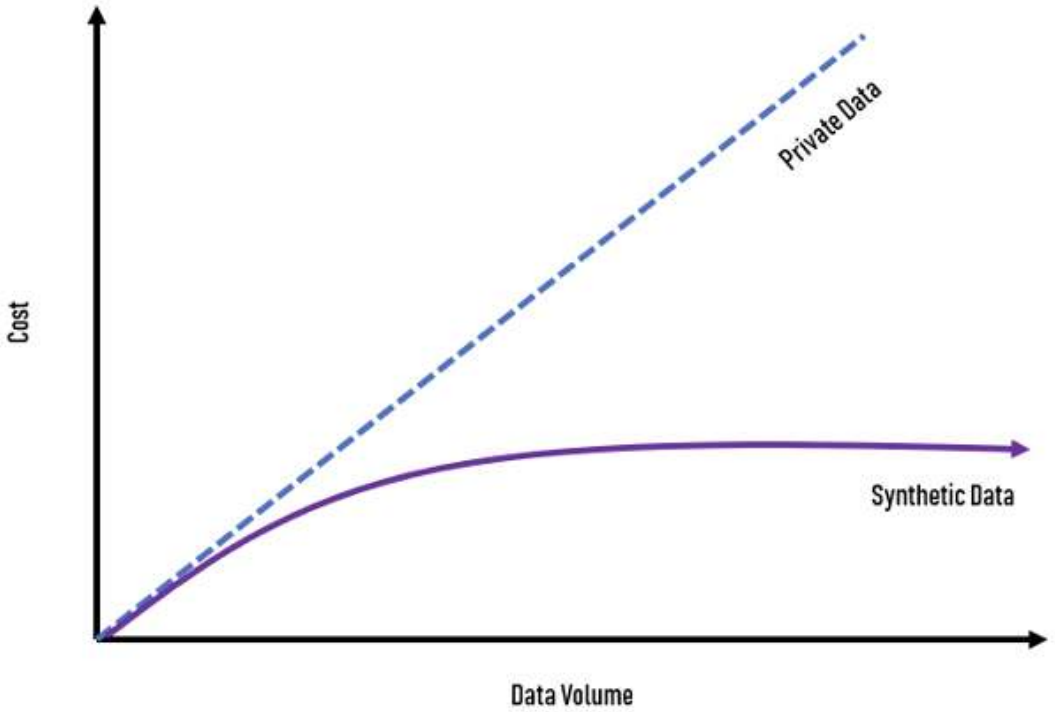
Lidar

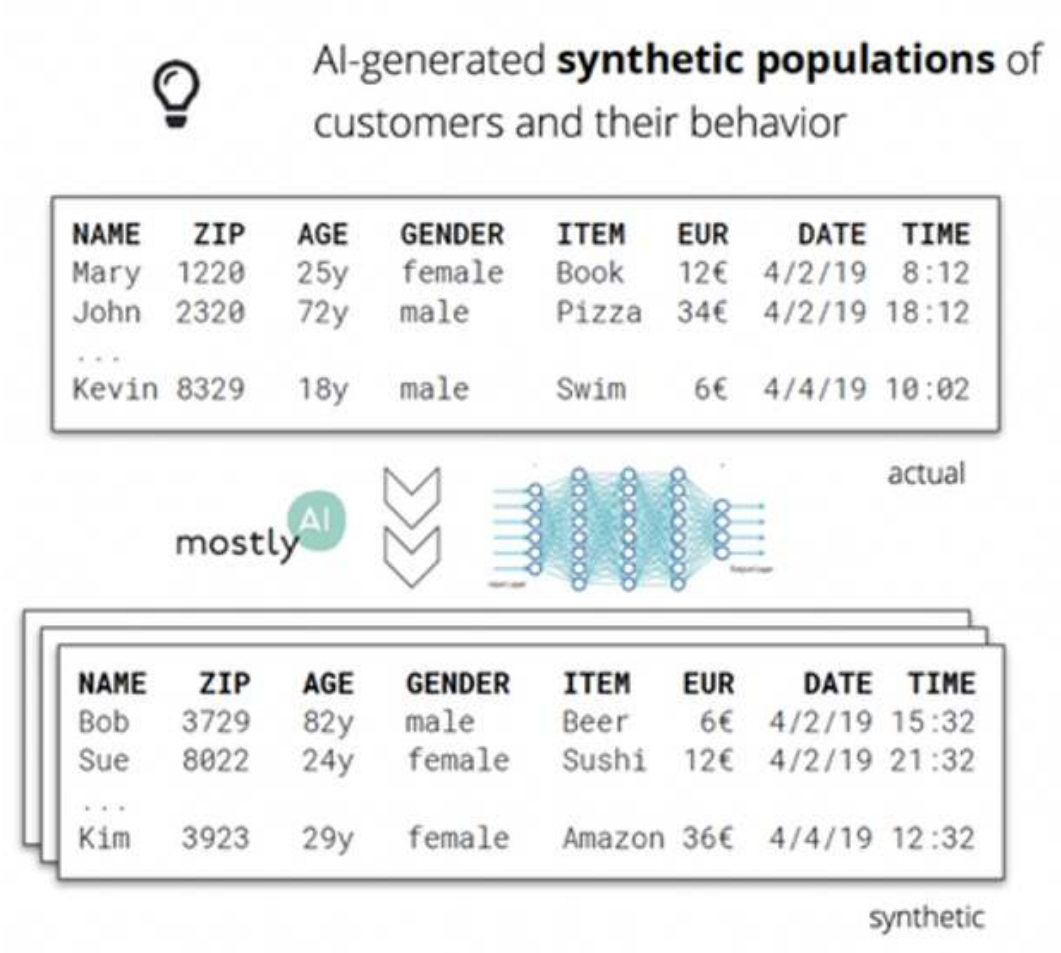
Lidar is a method for measuring distances (ranging) by illuminating the target with laser light and measuring the reflection with a sensor. Differences in laser return times and wavelengths can then be used to make digital 3D representations of the target. It has terrestrial, airborne and mobile applications

2) Data annonimization



3) Synthetic Data





4) Data preperation (cleansing + feature engineering)

Feature engineering also known as feature creation is the process of constructing new features from existing data to train a machine learning model.

In machine learning and pattern recognition, a feature is an individual measureable property or charachteristic of a phenomenon being observed.

Features engineering is the process of using domain knowledge to extract features from raw data via data mining techniques. These features can be used to improve the performance of machine learning algorithms. Feature engineering can be considered as applied machine learning itself.

For example

Character recognition: features may include histograms counting the number of black pixels along horizontal and vertical directions, number of internal holes, stroke detection and many others

Speech recognition: features for recognizing phonemes can include noise ratios, length of sounds, relative power, filter matches and many others

Spam detection: feature may include the presence or absence of certain email headers, the email structure, the language, the frequency of specific terms, the grammatical correctness of the text

Computer vision: there are a large number of possible features, such as edges and objects

Feature Extraction:

There usually have some meaningful features inside existing features, you need to extract them manually.

Some examples:

- Location
  - address, city, state and zip code (categorical or numeric)
- Time
  - year, month, day, hour, minute, time ranges (numeric)

- Morning, noon, afternoon, evening (categorical)

- Numbers
  - Turn age number into ranges (ordinal or categorical)

Consider a dataset with 2 patterns

Pattern 1: ^+*B*++*T*+*C*+

Pattern 2: +*R*++*T*+*C*

Extraction of features

Patterns	Features extracted
^+B++T+C+	^+B++, ++T+, +C+
+R++T+C	+R++, ++T+, +C

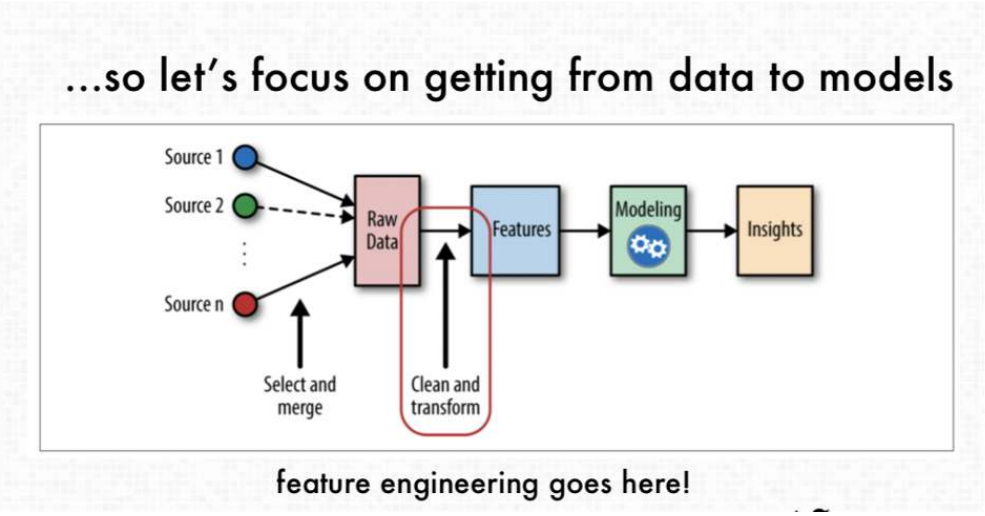
Representing patterns as a vector

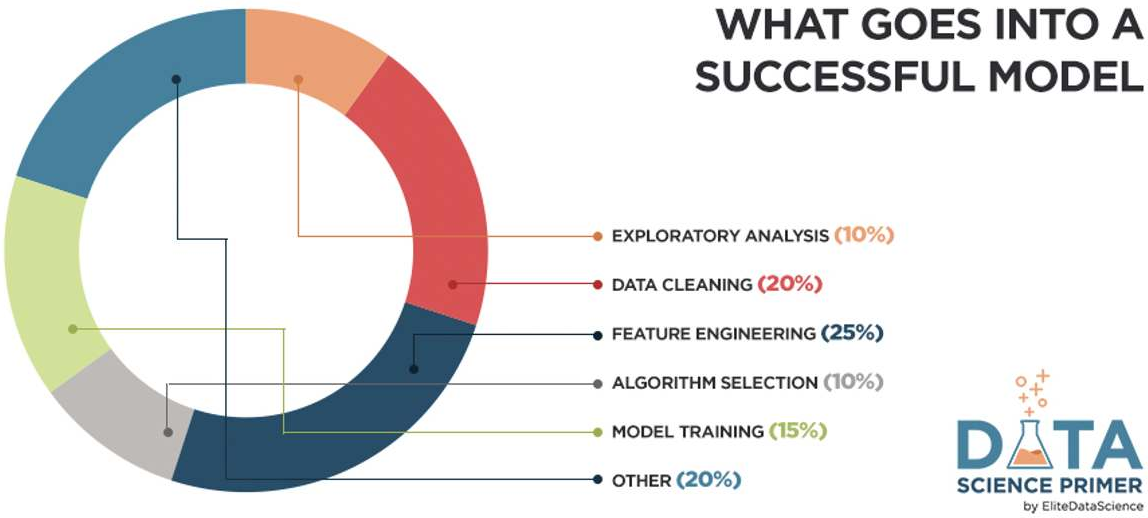
Patterns	^+B++	++T+	+C+	+R++	+C
^+B++T+C+	1	1	1	0	0
+R++T+C	0	1	0	1	1

* Numbers in the cell indicate presence of that particular feature

Data preperation(cleansing + feature engineering)

OSLOMET





Feature Engineering →

Transformations

Transformations
- Create new features out of one or more of the existing data.

client_id	joined	income	credit_score
46109	2002-04-16	172677	527
49545	2007-11-14	104564	770
41480	2013-03-11	122607	585
46180	2001-11-06	43851	562
25707	2006-10-06	211422	621

client_id	joined	income	credit_score	join_month
46109	2002-04-16	172677	527	4
49545	2007-11-14	104564	770	11
41480	2013-03-11	122607	585	3
46180	2001-11-06	43851	562	11
25707	2006-10-06	211422	621	10

Feature engineering requires extracting the relevant information from the data and getting it into a single table which can then be used to train a machine learning model. We can group the operations of feature creation into two categories: **transformations** and **aggregation**

Feature Engineering → Aggregations

Feature engineering requires extracting the relevant information from the data and getting it into a single table which can then be used to train a machine learning model. We can group the operations of feature creation into two categories: **transformations** and **aggregation**

Aggregations

- Performed across data and usually calculates statistics

client_id	joined	income	credit_score
46109	2002-04-16	172677	527
49545	2007-11-14	104564	770
41480	2013-03-11	122607	585
46180	2001-11-06	43851	562
25707	2006-10-06	211422	621

client_id	joined	income	credit_score	join_month	log_income	mean_loan_amount	max_loan_amount	min_loan_amount
46109	2002-04-16	172677	527	4	12.059178	8951.600000	14049	559
49545	2007-11-14	104564	770	11	11.557555	10289.300000	14971	3851
41480	2013-03-11	122607	585	3	11.716739	7894.850000	14399	811
46180	2001-11-06	43851	562	11	10.688553	7700.850000	14081	1607
25707	2006-10-06	211422	621	10	12.261611	7963.950000	13913	1212
39505	2011-10-14	153873	610	10	11.943883	7424.050000	14575	904
32726	2006-05-01	235705	730	5	12.370336	6633.263158	14802	851
35089	2010-03-01	131176	771	3	11.784295	6939.200000	13194	773
35214	2003-08-08	95849	696	8	11.470529	7173.555556	14767	667
48177	2008-06-09	190632	769	6	12.158100	7424.368421	14740	659

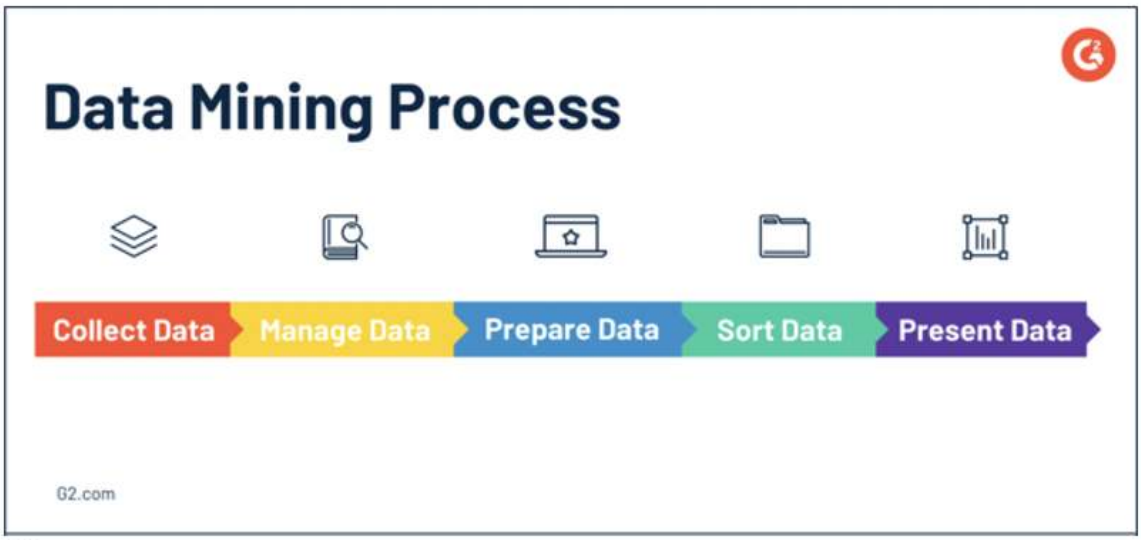
5) Data Wrangling

The process of transforming and mapping data from one “raw” data form into another format with the intent of making it more appropriate and valuable for a variety of downstreams purposes such as analytics

Tools: python, excel, tablaeu, google data prep

6) Data mining

The process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics and database systems.

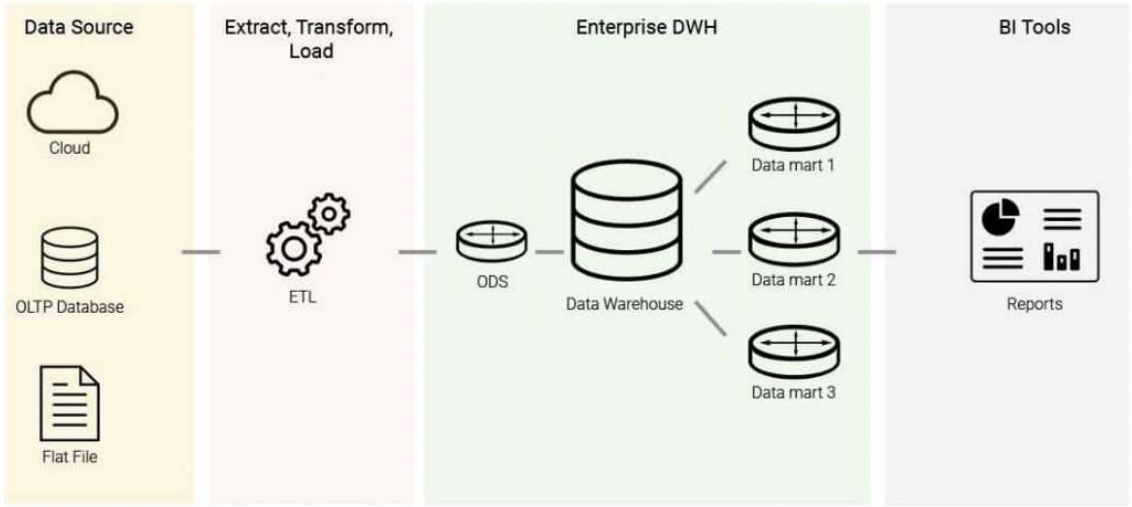


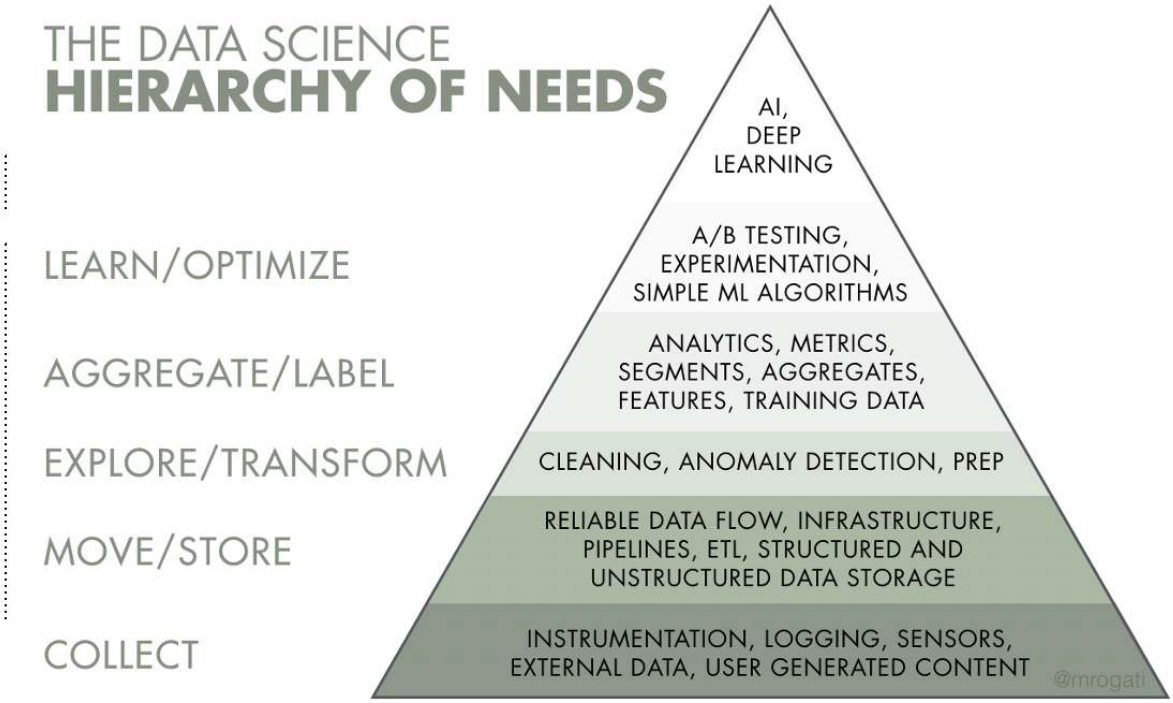
NIVERSITY

For example, through data mining, a business may be able to see which customers are buying specific products at certain times of the year. This information can then be used to segment those customer. Customer segmentation is important for targeting sales and marketing campaigns - which may lead to higher profits, but also point toward potential trend or two.

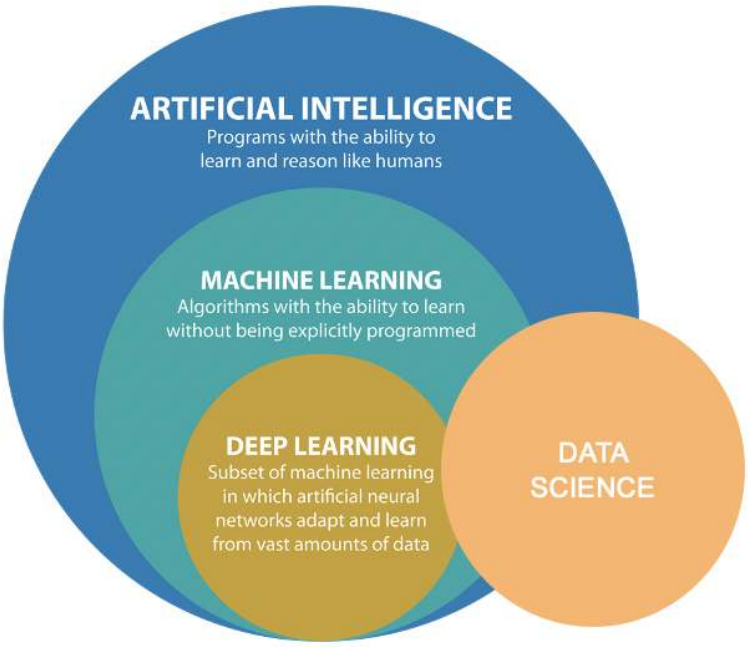
In addition to automated decision-making, data mining is also an important tool because it can accurately predict and forecast trend for your business based on historical information and current condition. It also has the capability to allow for more efficient use and allocation of resources so that businesses can plan and make automated decisions to maximize cost reduction.

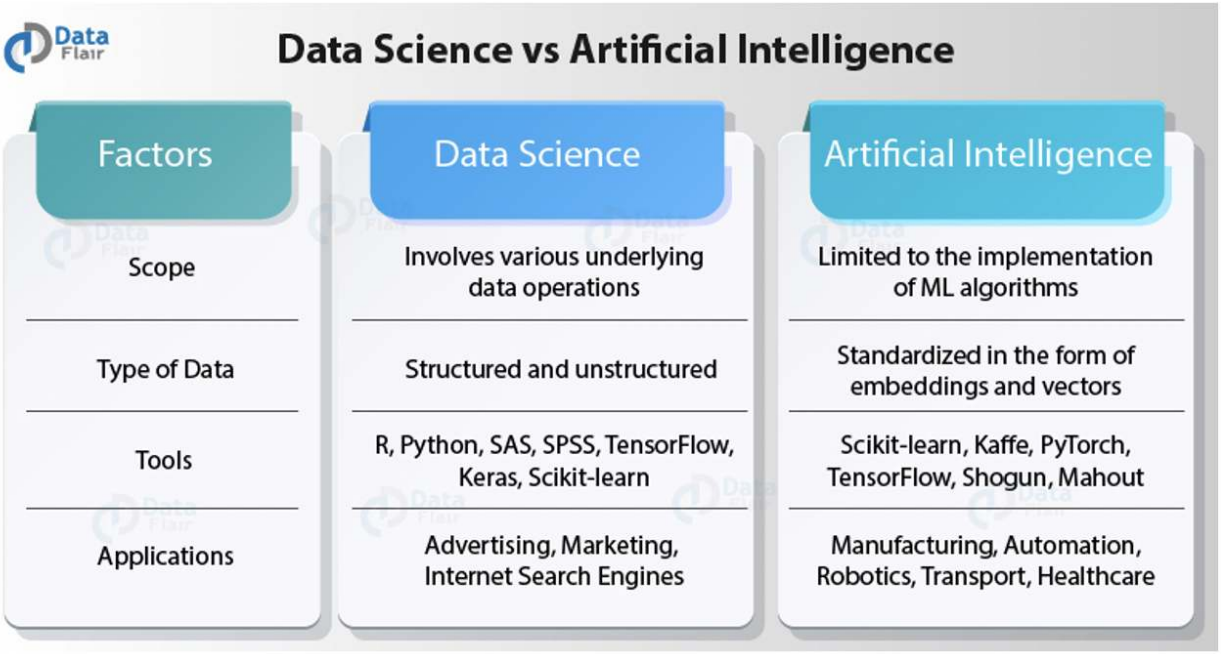
7) Data warehousing


Data Source
Extract, Transform, Load
Enterprise DWH
BI Tools
Cloud
OLTP Database
Flat File
ETL
ODS
Data Warehouse
Data mart 1
Data mart 2
Data mart 3
Reports


THE DATA SCIENCE
HIERARCHY OF NEEDS
AI, DEEP LEARNING
LEARN/OPTIMIZE
A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS
AGGREGATE/LABEL
ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA
EXPLORE/TRANSFORM
CLEANING, ANOMALY DETECTION, PREP
MOVE/STORE
RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE
COLLECT
INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT
@mrogati

This has given rise to a field related to AI,





Lecture 3

Statistics is a traditional field, broadly defines as a branch of mathematics dealing with data collection, organization, analysis, interpretation and presentation.

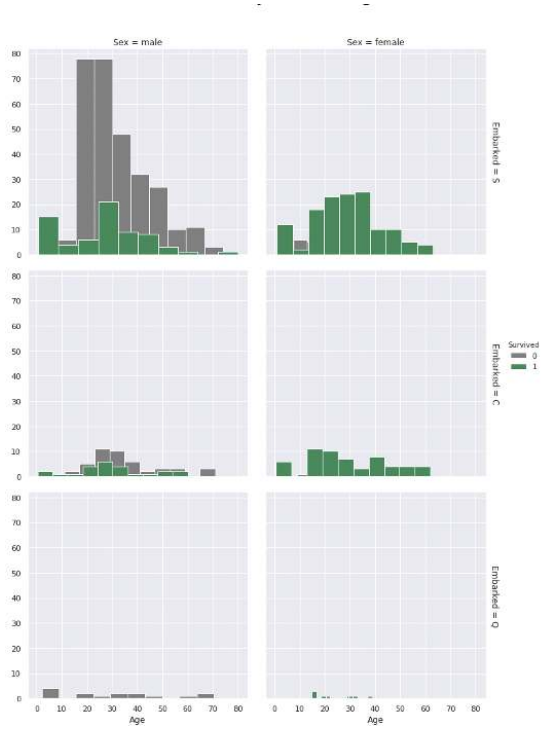
Titanic dataset

OSLOMET

- Most passengers seem to be boarded on Southampton(S).
- More than 60% of the passengers died boarded on Southampton.
- More than 60% of the passengers lived boarded on Cherbourg(C).
- Pretty much every male that boarded on Queenstown(Q) did not survive.
- There were very few females boarded on Queenstown, however, most of them survived.

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

DAVE3625- INTRO TO A.I. BY UMAIR M.I

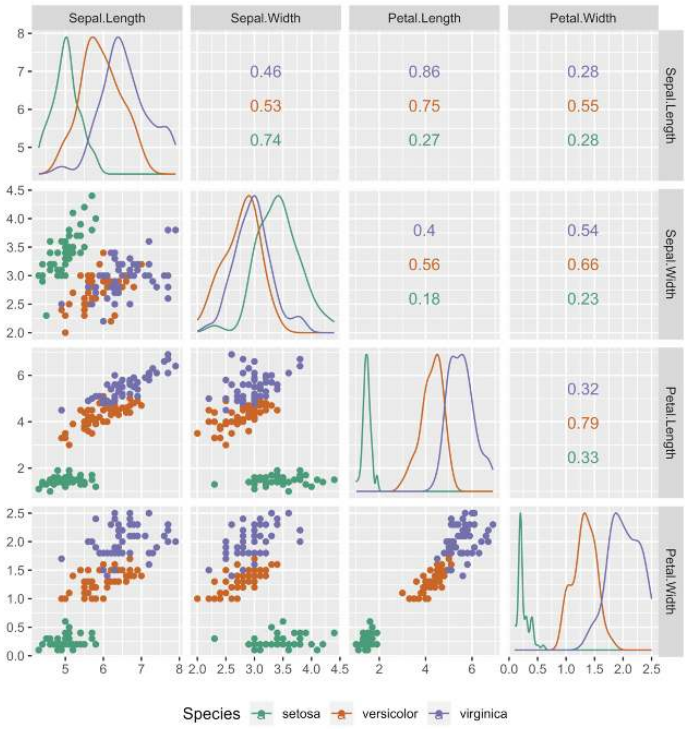


This is a compelling facet grid illustrating four features relationship at once. They are Embarked, Age, Survived & Sex.

- The color illustrates passengers survival status (green represents survived, gray represents not survived).
- The column represents Sex (being male, right stands for female)
- The row represents Embarked (from top to bottom: S, C , Q)

OSLOMET

Scatter plots



OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Scatter plots are used in descriptive statistics to show the observed relationships between different variables, here using the iris flower datasets.

What is machine learning?

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

Machine Learning algorithms can learn its own from past experiences, just like humans do. When exposed to new data, these algorithms learn, change and grow by themselves without you needing to change the code every single time.

Uses of machine learning

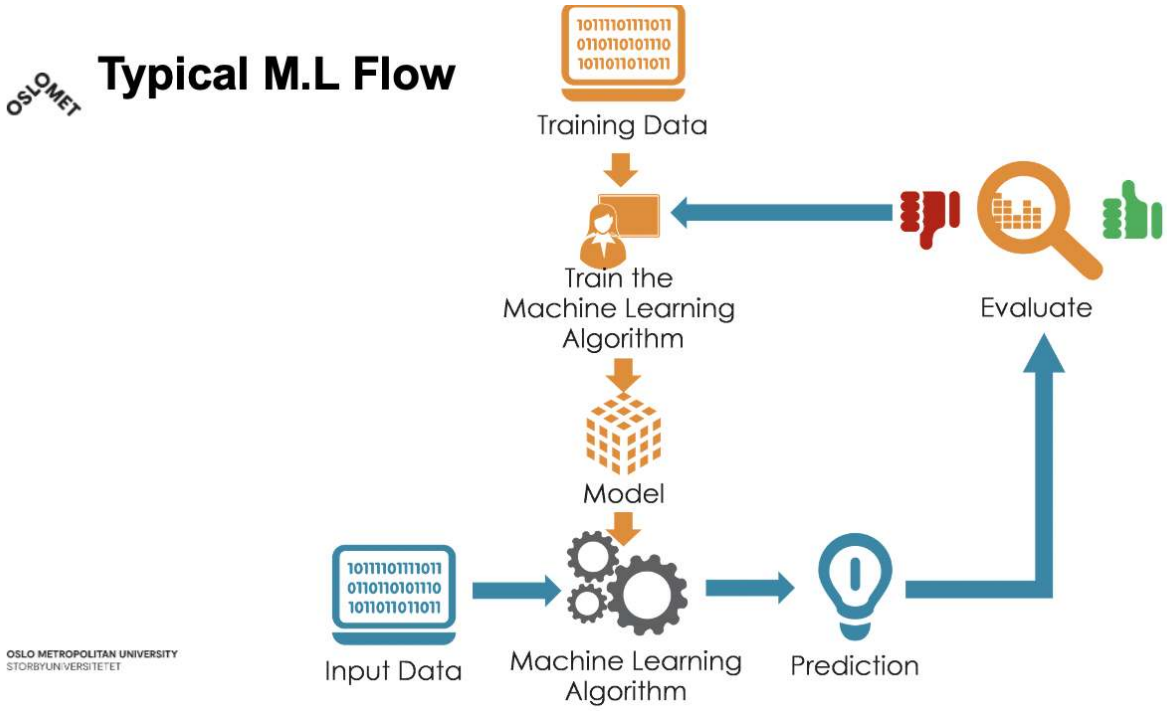
Example: Search for cats on google. How does this search work?

- Google first gets a large quantity of examples of photos labeled “cat”.
- Then the machine learning algorithm looks for patterns of pixels and patterns of colors that will help it predict if the image is of “cat”
- At first, Google´s computers make a randoms guess of what patters are good in order to identify an image of a cat
- If it makes a mistakes, the a set of adjustments are made (by humans) in order for the algorithm to get it right
- In the end the algorithm will learn such patterns and improve its output

How would you build such a machine learning algorithm?

1. Collect data
2. Label the data
3. Train the machine learning model with data
    a. the model will look for patterns in the images in order to identify cats and lions
4. Once the model is trained, we give it different input images to test and see if the model gives you the right answer

Typical Machine Learning Flow

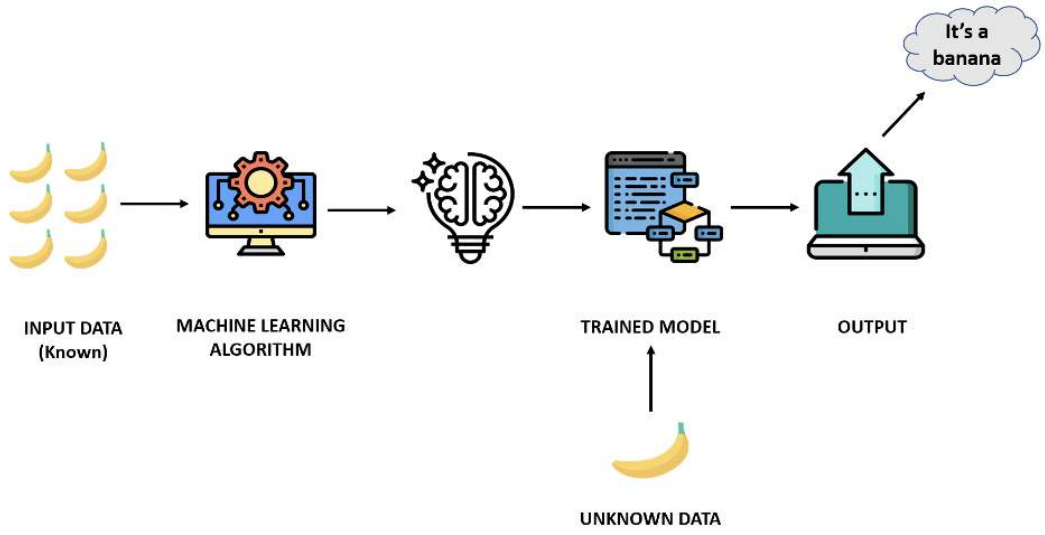


Types of Machine Learning algorithms

- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Recommender systems

Supervised learning

In a supervised learning model, the algorithm learns on a labeled dataset, providing an answer key that the algorithm can use to evaluate its accuracy on training data.

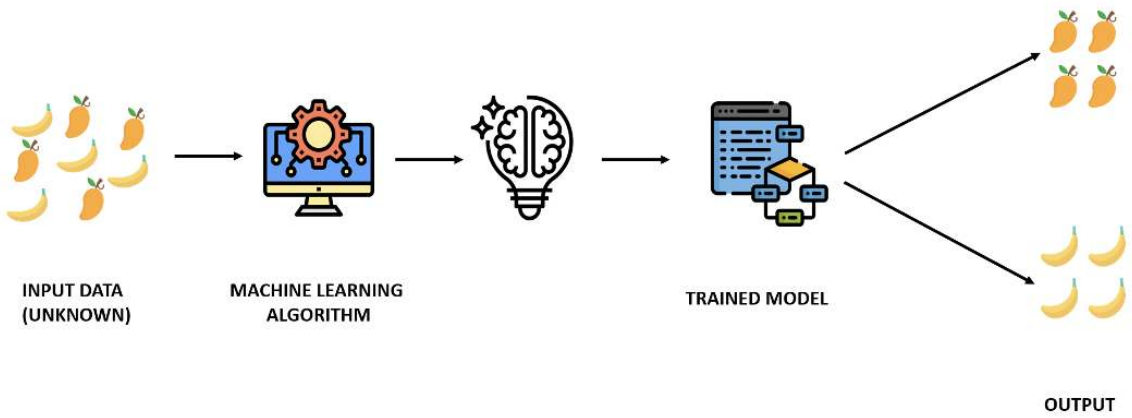


OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

INTRODUCTION TO A.I - UMAIR M.I

Unsupervised Learning

An unsupervised model, provides unlabeled data that the algorithm tries to make sense of by extracting features and pattern on its own.



OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

INTRODUCTION TO A.I - UMAIR M.I

1. The training data is unknown or unlabelled
2. The model tries to find patterns and relationships in the dataset by creating clusters in it
3. Model cannot add labels to the data

Reinforcement Learning

Reinforcement learning is a type of learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions experiences

Whenever the model predicts or produces a result, it is penalized if the prediction is wrong or rewarded if the prediction is correct

Challenges with reinforcement learning

The main challenge in reinforcement learning lays in preparing the simulation environment, which is highly dependent in the task to be performed.

Scaling and tweaking the neural network controlling the agent is another challenge. There is no way to communicate with the network other than through the system of rewards and penalties.

When the model has to go superman in Chess, Go or Atari games, preparing the simulation environment is relatively simple. When it comes to building a model capable of driving an autonomous car, build a realistic simulator is crucial before letting the car ride on the street. The model as to figure out how to brake or avoid collision in a safe environment, where sacrificing even a thousand cars comes at a minimal cost. Transferring the model out of the training environment and into to the real world is where things get tricky.

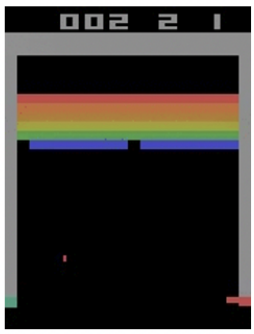
There are agents that will optimize the prize without performing the task it was designed for. An interesting example can be found in the OpenAI video (below), where the agent learned to gain rewards, but not to complete the race.

NIVERSITY

OSLOMET

Breakout game - Atari

Initial performance

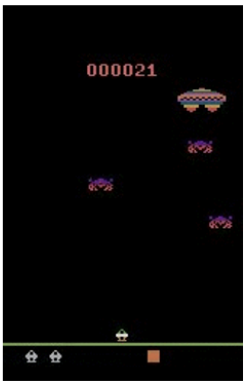


After 15 min of training

After 30 min of training

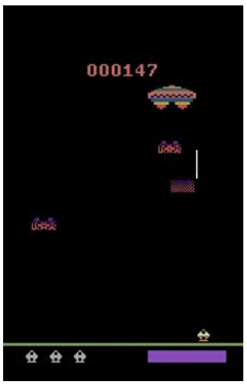
Assault game - Atari

Initial performance



After 15 min of training

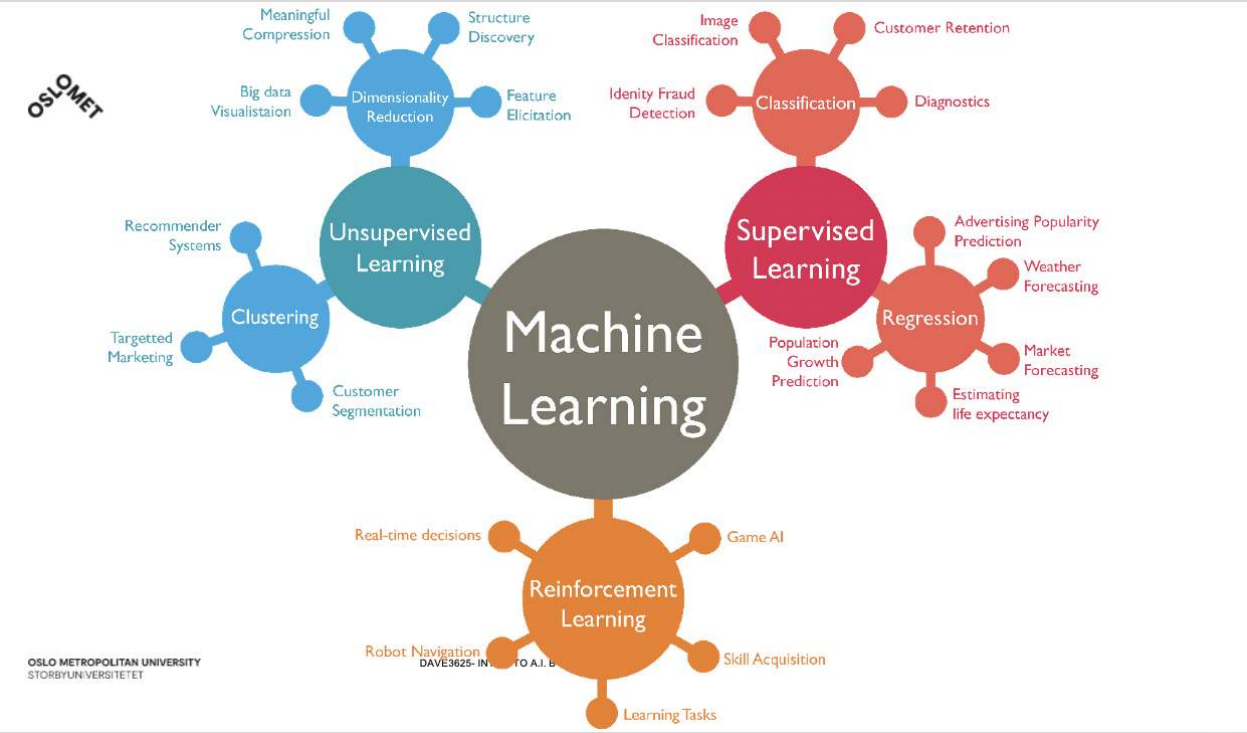
After 30 min of training



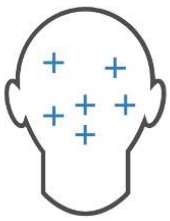
Recommender systems

Recommender system are an important class of machine learning algorithms that offer relevant suggestions to users.

These system predict the most likely product that the users are most likely to purchase and are of interest to them.



Where should we use machine learning?



Hand-written rules and equations are too complex—as in face recognition and speech recognition.

The rules of a task are constantly changing—as in fraud detection from transaction records.

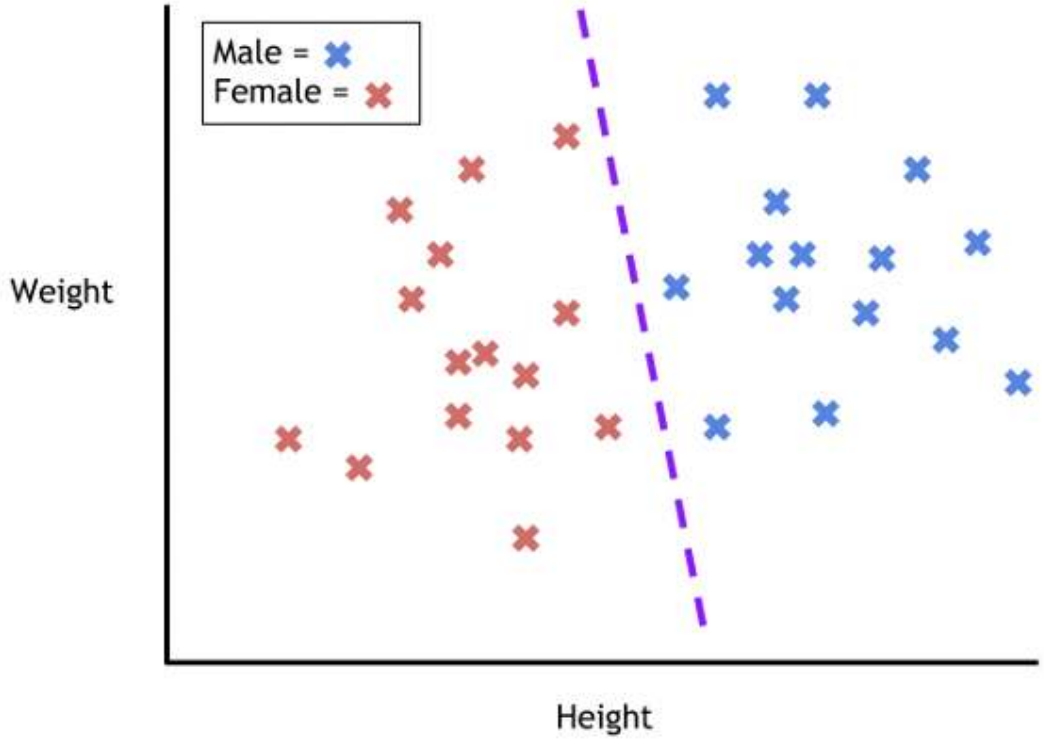
The nature of the data keeps changing, and the program needs to adapt—as in automated trading, energy demand forecasting, and predicting shopping trends.

Lecture 4

Types of Machine Learning Algorithms

- Supervised Learning
- Unsupervised Learning
- Reinforcement learning
- Recommender systems

Supervised Machine Learning



INTRODUCTION TO A.I - UMAIR M.I

Classification algorithm is to take an input value and assign it a class, or category that it fits into based on the training data provided. The most common example of classification is determining if an email is spam or not. With two classes to choose from (spam or not spam), this problem is called binary classification problem.

The algorithm will be given training data with emails are both spam and not spam. The model will find the features within the data that correlate to either class and create the mapping function mentioned earlier: Y = f(x). Then, when provided with an unseen email, the model will use this function to determine wether or not the email is spam → Divides the data in classes / categories

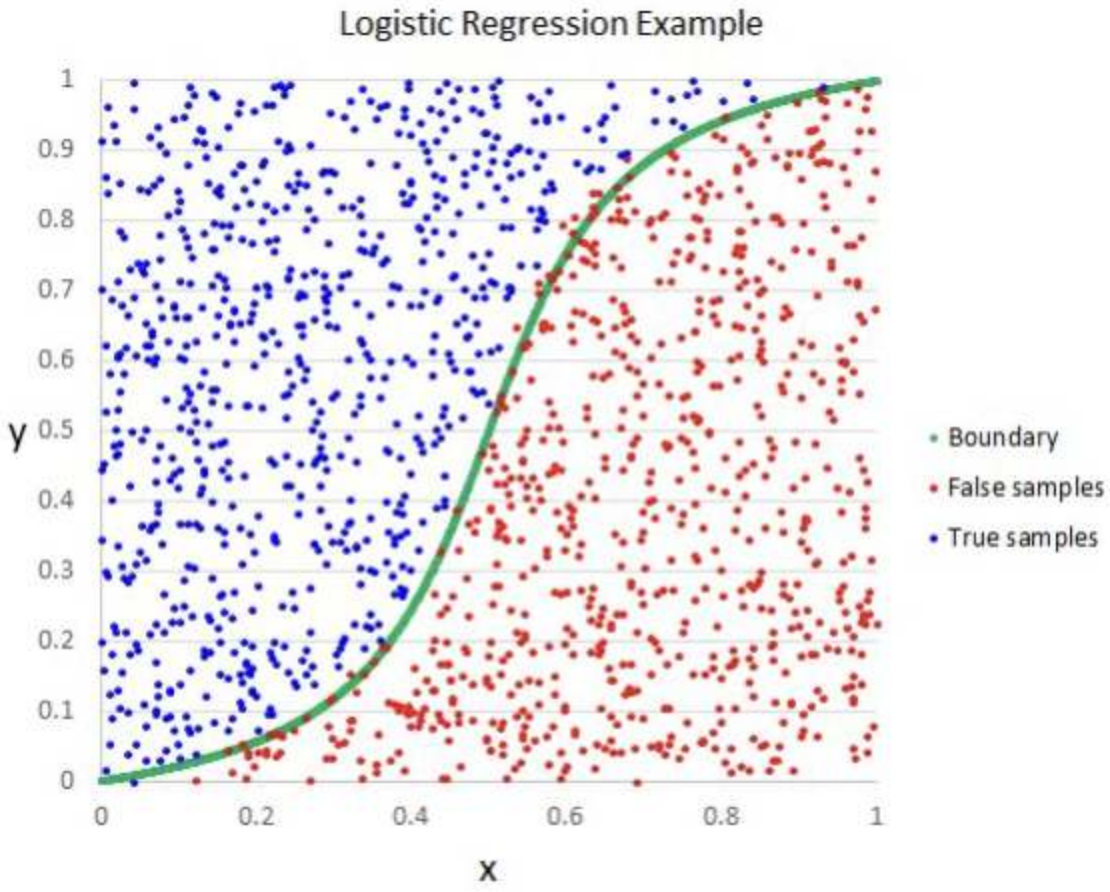
Use cases:

- Spam detection (classification: spam or normal)
- Analysis of the customer data to predict wether they will buy computer accessories (Classification: Yes or no)
- Classifying fruits from features like colour, taste, size, weight (classification: Apple, Orange, Cherry, Banana)
- Gender classification from hair length (classification: Male or female)
- Stock market price prediction (classification: high or low)

Classification versus Clustering algorithms

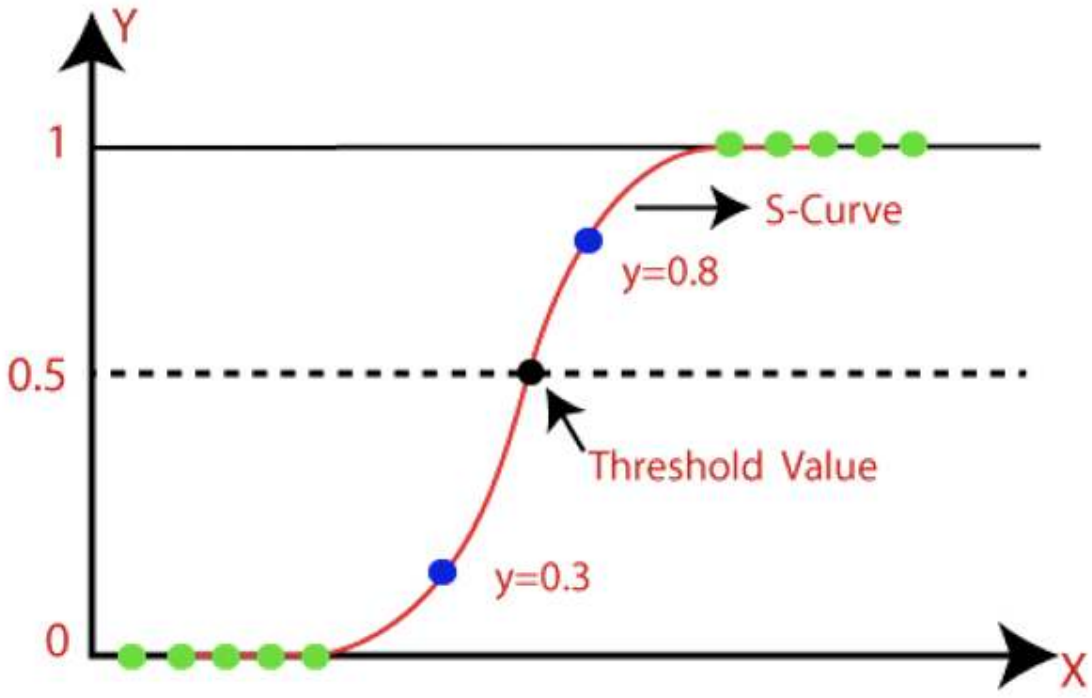
Clustering groups similar kind of things together, however classification predicts similar kind of things together based on the historic trained data.

Classification → Logistic regression is used when you have a classification problem : yes / no, pass / fail, win / lose, Alive / dead, healthy / sick etc….



Sigmoid function

Allows to put a threshold value, e.g, 0.5 (use case: spam detection)



Uses: Logistics regression is used to predict the occurance of some event, e.g;

- Predicts wether rain will occur or not
- Spam detection, Diabetes prediction, cancer detection etc.

Types of Logistics regression

- **Binary logistics regression (e.g pass / fail)**
- **Multiclass logistic regression (e.g cats, dogs, sheep)**
- **Ordinal (low, medium, high)**

Example run of an AI algorithm

1. Develop the algorithm

2. Evaluate the algorithm

Develop (using binary logistic regression)

Step 1: Visualization

Say we´re given data on student exam results and our goal is to predict wether a student will pass or fail based on number of hours slept and hours spent studying. We have two features (hours slept, hours studied) and two classes: Passed (1) and failed (0)

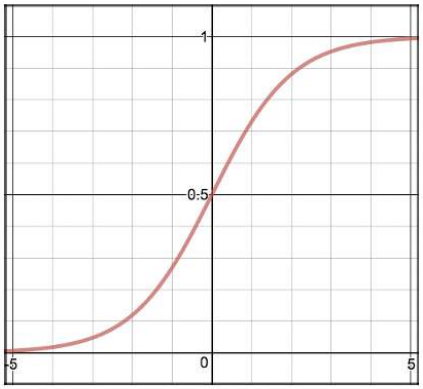
Step 2: Sigmoid Activation

$$S(z) = \frac{1}{1 + e^{-z}}$$

Code

```
def sigmoid(z):
  return 1.0 / (1 + np.exp(-z))
```

$S(z)$ = output between 0 and 1 (probability estimate)
z = input to the function (your algorithm's prediction e.g. mx + b)
e = base of natural log



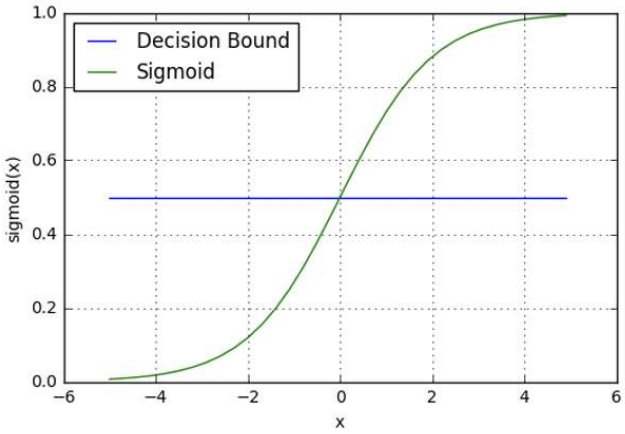
In order to map predicted to probabilities, we use the sigmoid function. The function maps any real value into another value between 0 and 1.

→ In machine learning, we use to map predictions to probabilities

Step 3: Decision boundary

- We need a threshold value

$$p \geq 0.5, class = 1$$
$$p < 0.5, class = 0$$



Our current prediction functions returns a probability score between 0 and 1. In order to map this to a discrete class (true / false, cat / dog), we select a threshold value or tipping point above which we will classify values into class 1 and below which we classify values into class 2.

For example, if our threshold was .5 and our prediction function returned .7, we would classify this observation as positive. If our prediction was .2 we would classify the observation as negative. For logistic regression with multiple classes we could select the class with the highest predicted probability.

Step 4: Making predictions

Using our knowledge of sigmoid functions and decision boundaries, we can now write a prediction function. *A prediction function in logistic regression returns the probability of our observation being "True", or "Yes. We call this class and its notation is P(class=1)*

$$P(class = 1) = \frac{1}{1 + e^{-z}}$$

If the model returns .4 it believes there is only 40% change of passing. If our decision boundary was .5, We would categorize this observation as “Fail” → Code it

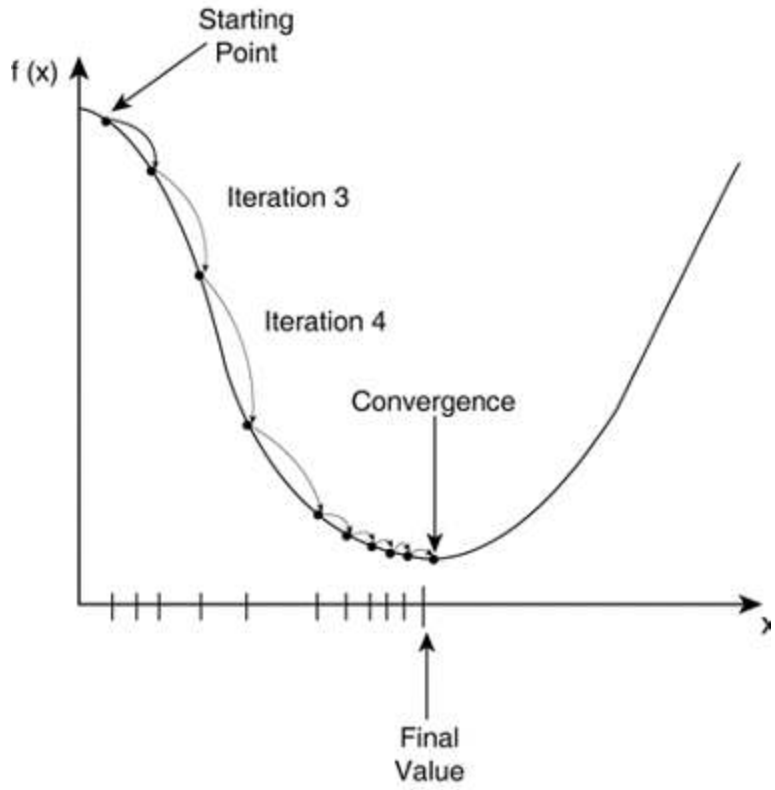
Second stage: Evaluate the algorithm

What is a cost function? The cost functions helps the learner to correct / change behaviour minimize mistakes. In machine learning, cost function are used to estimate how badly models are performing.

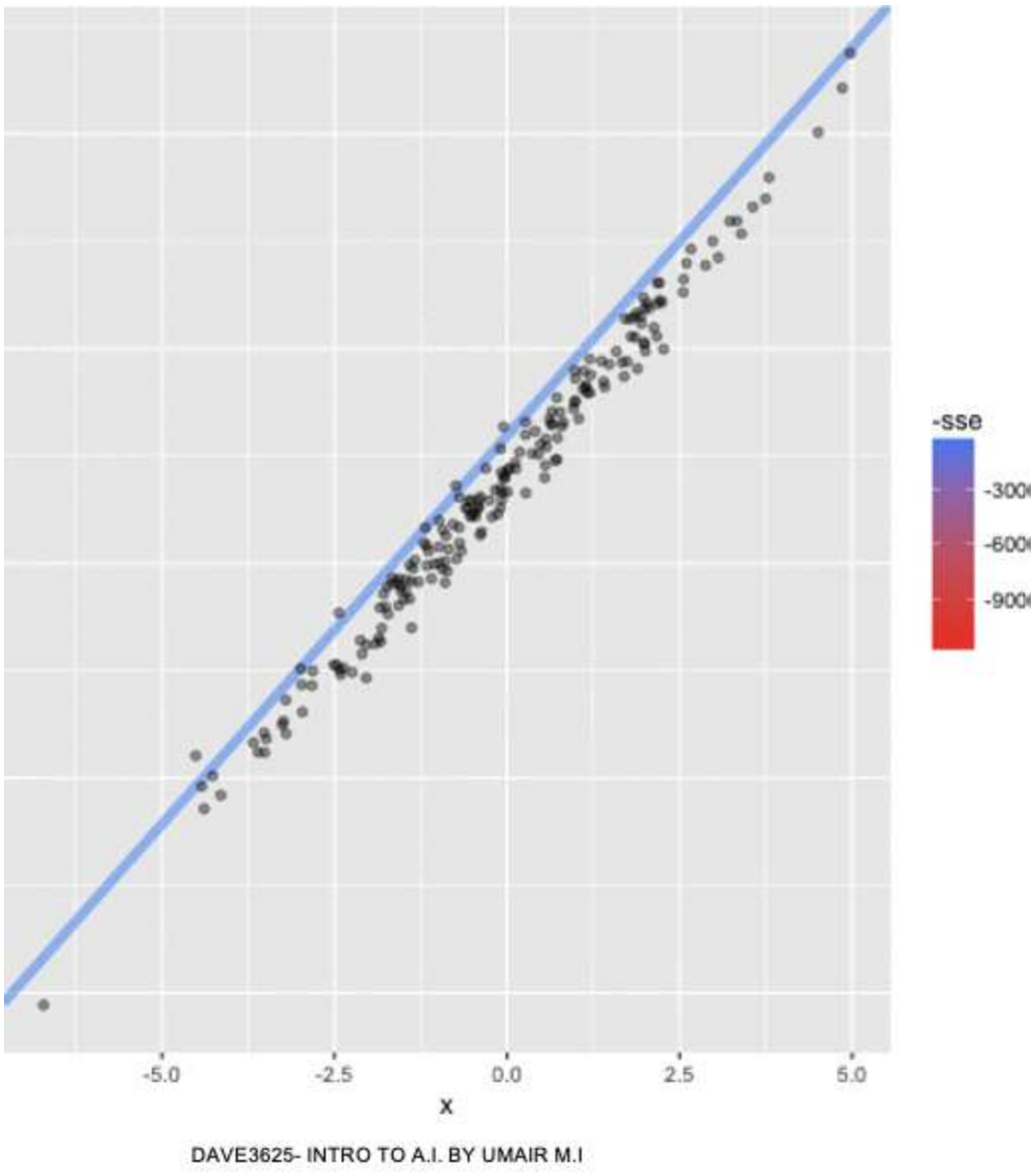
→ Put simply, a cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and Y.

Cost function: Gradient descent

Gradient Descent is an optimization algorithm used to tweak parameters iteratively to minimize the cost function.



‘Gradient descent enables a model to learn the gradient or direction that the model should in order to reduce errors (difference between actual Y and predicted Y)



Develop: Map probabilities to classes

The final step is assign class labels (0 or 1) to our predicted proabilities.

Decision boundary

```
def decision_boundary(prob):
  return 1 if prob >= .5 else 0
```

Convert probabilities to classes

```
def classify(predictions):
  '''
  input  - N element array of predictions between 0 and 1
  output - N element array of 0s (False) and 1s (True)
  '''
  decision_boundary = np.vectorize(decision_boundary)
  return decision_boundary(predictions).flatten()
```

Example output

```
Probabilities = [ 0.967, 0.448, 0.015, 0.780, 0.978, 0.004]
Classifications = [1, 0, 0, 1, 1, 0]
```

Final stage: Finish it up

- Train the model
- Evaluate the model
  - Minimize the cost with repeated iterations
- Measures the accuracy of your outputs
- Measure the probability score

Classification → Naive Bayes Classification Algorithm

A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other features.

E.g a fruit may be considered to be an apple if it is red, round and about three inches in diameter. Its an easy to build algorithm and a very powerful one.

Lecture 5

Types of supervised learning

Classification

- Support vector machines
- Decision Trees
- K-Nearest Neighbour
- Random Forest
- Logistic Regression
- Naive Bayes Classifier

Regression

- Linear Regression
- Polynomial Regression

Naive Bayes Classification Algorithm (classification)

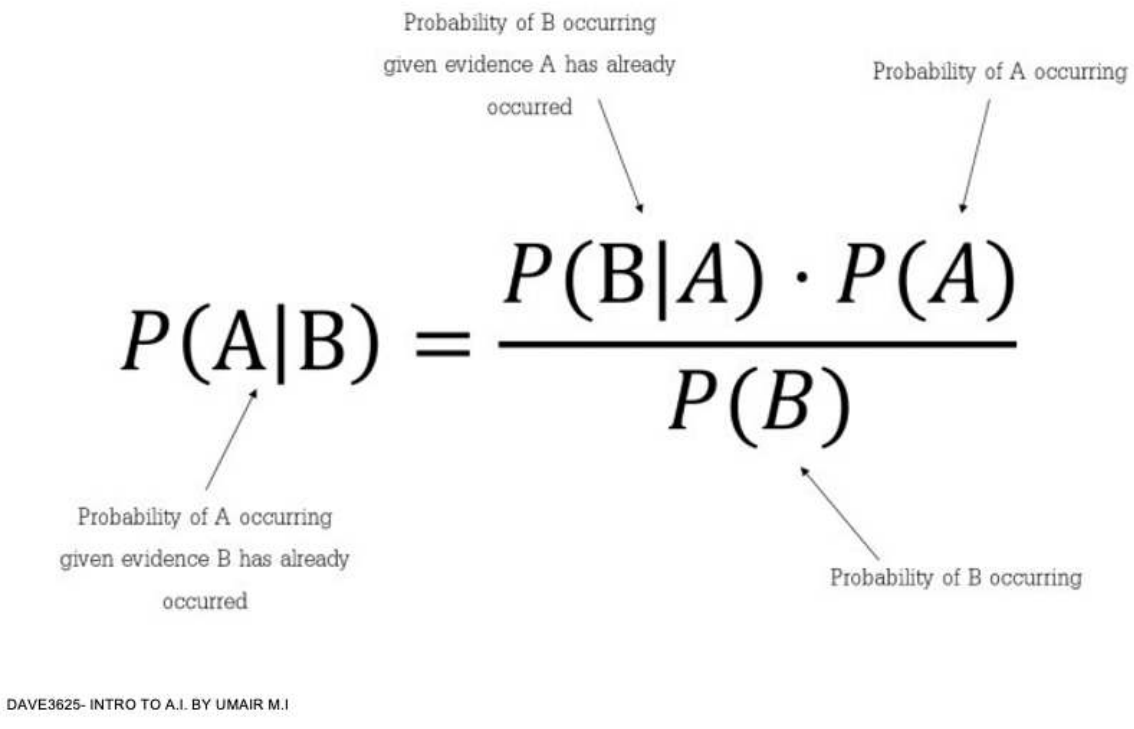
A Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes does not take into account the relationship between features.

E.g:

- A dog will be considered a dog because it has 4 legs and a tail
  - But a Naive Bayes will consider everything a dog which has 4 legs or tail (e.g squirrel) not because its an animal. Being an animal is the relationship → not legs / tail
- A fruit may be considered to be an apple if it is red, round and about three inches in diameter.

Bayes Theorem

Bayes´s theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event.



Using Bayes Theorem, we can find the probability of A happening, given that B has occured. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors / features are independent. That is presence of one particular feature does not affect the other → Hence it is called naive.

Fruit	Long	Sweet	Yellow	Total
Banana	400	350	450	500
Orange	0	150	300	300
Other	100	150	50	200
Total	**500**	**650**	**800**	**1000**

We can see that:

- 50% of the fruits are bananas
- 30% are oranges
- 20% are other fruits

We can also say:

- From 500 bananas, 500 (0.8) are Long, 350 (0.7) are Sweet, and 450 (0.9) are Yellow.
- Out of 300 oranges 0 are Long, 150 (0.5) are Sweet and 300 (1) are Yellow
- From the remaining 200 fruits, 100 (0.5) are long, 150 (0.75) are sweet and 50 (0.25) are yellow

Banana:

$$P(Banana|Long, Sweet, Yellow)$$
$$= \frac{P(Long|Banana) \cdot P(Sweet|Banana) \cdot P(Yellow|Banana) \cdot P(Banana)}{P(Long) \cdot P(Sweet) \cdot P(Yellow)}$$
$$= \frac{0.8 \times 0.7 \times 0.9 \times 0.5}{P(evidence)}$$
$$= \frac{0.252}{P(evidence)}$$

Orange:

$$P(Orange|Long, Sweet, Yellow) = 0$$

Other Fruit:

$$P(Other|Long, Sweet, Yellow)$$
$$= \frac{P(Long|Other) \cdot P(Sweet|Other) \cdot P(Yellow|Other) \cdot P(Other)}{P(Long) \cdot P(Sweet) \cdot P(Yellow)}$$
$$= \frac{0.5 \times 0.75 \times 0.25 \times 0.2}{P(evidence)}$$
$$= \frac{0.01875}{P(evidence)}$$

E.g Let’s say given the features of a piece of fruit and we need to predict the class. If we ´re told that the additional fruit is Long, Sweet and Yellow, we can classify it using the following formula and subbing in the values for each outcome, wether it’s a Banana, an Orange or Other fruit. The one with the highest probability (score) being the winner. In this case, based on the higher score 0.01875 < 0.252 we can assume this Long, Sweet and Yellow fruit is, in fact a, a Banana.

Naive Bayes use in Natural Language Processing

Task: Determine if the text “I feel terrible” is either mood category or not mood.

TEXT	CATEGORY
I am so angry.	mood
I feel like superstar.	mood
It is going to rain.	not mood
I want to cry.	mood
They will come together.	not mood

Naive Bayes will count 2 occurances for mood category and 0 for not mood category. Hence classifying the text as mood.

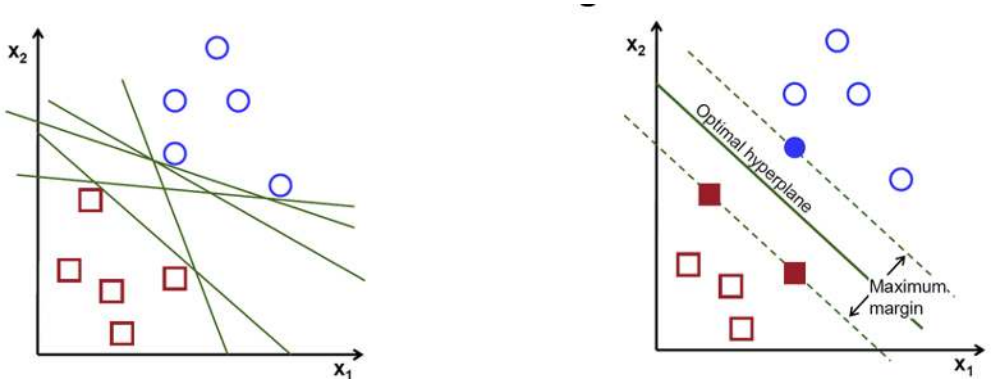
I	mood
feel	mood
so	
terrible	

Support Vector Machines (Classification)

Main use is to classify unseen data. Support vector machines classifies the data by finding a clear separation between the data points

→ **It looks for a hyperplane**

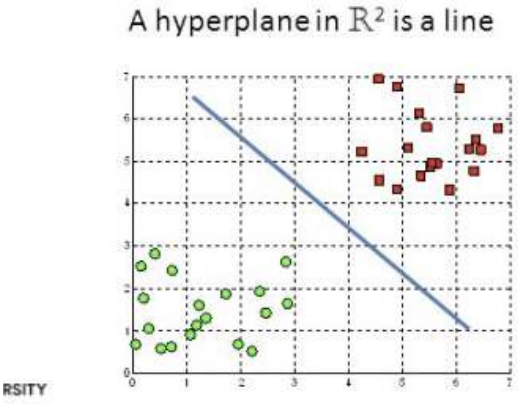
Can also be used for both classification and regression.

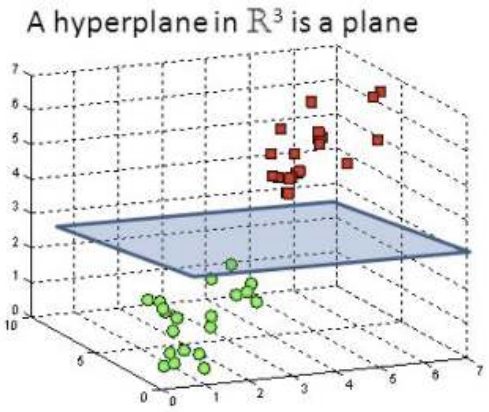


To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e, the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that the future data points can be classified with more confidence.

Hyperplanes

Hyperplanes are boundaries that help classify the data points. We can use a 2 dimension hyperplane (if number of features is 2) or 3 dimensional hyperplane (if number of features is 3) → We can also have n dimensional hyperplane (for more features)





Another reason we use SVM is because they can find complex relationships between your data without you needing to do a lot of transformations on your own. It’s a great option when you are working with smaller datasets that have tens to hundreds of thousands of features.

Uses of support vector machines

- Used to detect cancerous cells
- Used to predict driving routes
- Face detection
- Image classification
- Handwriting detection

Pros	Const
Effective on datasets with multiple features	They do not provide probability estimates. Those are calculated using an expensive five fold cross validation
Effective in cases where number of features is greater than the data points	Works best on small sample sets because of its high training time
Memory efficient	

Other classification algorithms

- Decision trees
- K-nearest Neighbour
- Random Forest

Regression

Regression models are used to predict a continuous value, e.g:

- Predicting prices of a house given a set of features (size, price, location etc.)
- Predicting sales revenue of a company based on previous sales figures

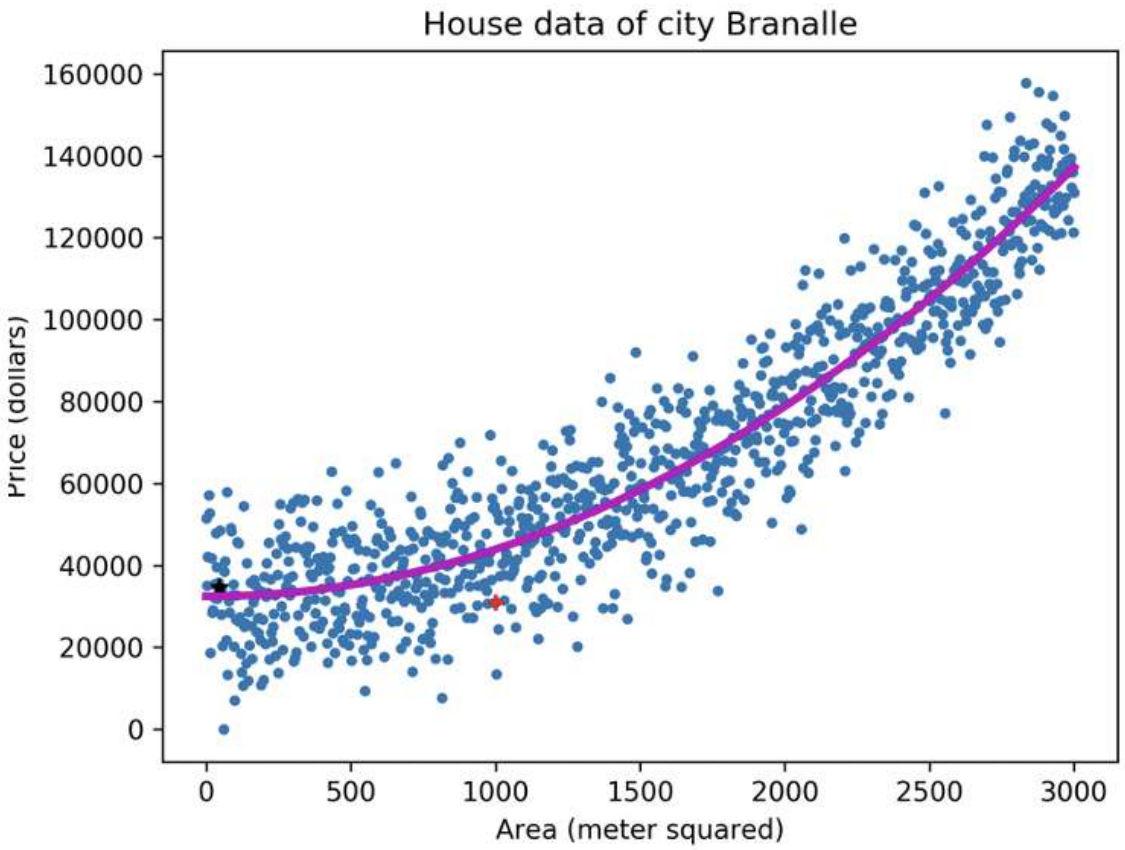
> The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age etc, and classification algorithms are used to predict / classify the discrete values such as Male or Female, True or False, Spam or Not Spam etc.

→ Regression algorithms = continuous values

→ Classification algorithms = discrete values

- Churn: Classification
- Stock market price prediction: Classification
- Spam email: classification
- Prediction of price of an oil: Regression
- Salary prediction: Regression
- Age prediction: Regression
- Gender prediction: Classification

Example of Regression



Regression is used for prediction, forecasting, time series modeling and determining the causal effect between variables.

Regression Algorithms	Classification Algorithms
In regression, the output variable must be of continuous nature or real value	In Classification, the output variable must be a discrete value
The task of the regression algorithm is to map the input value (x) with the continuous output variable (y)	The task of the classification algorithm is to map the input value(x) with the discrete output variable (y)
Regression Algorithms are used with continuous data	Classification algorithms are used with discrete data
In regression, we try to find the best fit line, which can predict the output more accurately	In classification, we try to find the decision boundary, which can divide the dataset into different classes

Regression algorithm can be used to solve the regression problems such as Weather prediction, House Price Prediction etc.	Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech identification of cancer cells, etc.
The regression algorithms can be further divided into Linear and non-linear regression	The classification algorithms can be divided into Binary Classifier and Multi-class Classifier

Types of regression

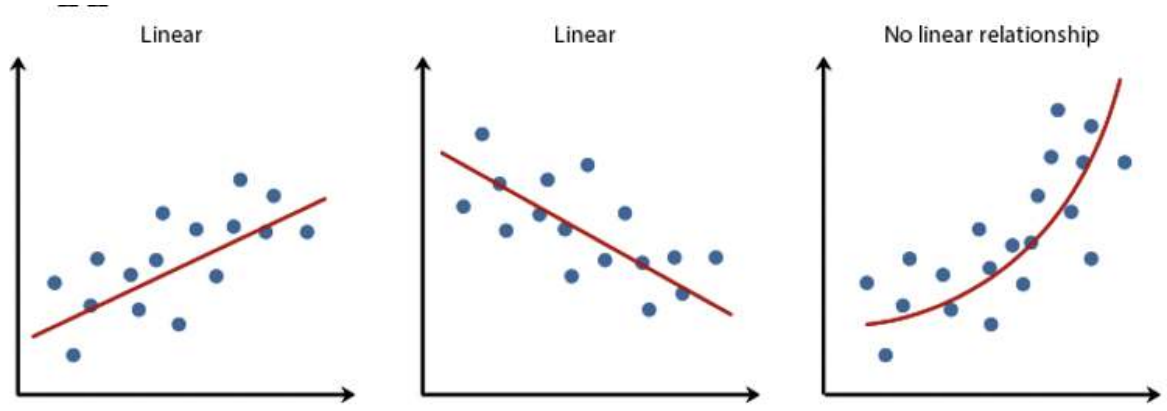
- Linear regression
- Logistic regression
- Polynomial regression
- Support vector regression
- Decision Tree regression
- Random forest regression
- Ridge regression
- Lasso regression

1) Linear regression

It’s a linear model (an equation) which describes a relationship between two quantities that show constant rate of change. e.g the older i get, the wiser i will be (hopefully).

→ Attending all lectures in Intro to AI course will result in passing the exam.

There is always an input variable (x) and an output variable (y) → Y can be calculated from a linear combination of input variables (x)



Linear regression is a predictive statistical approach for modeling relationship between a dependent variable with a given set of independent variables.

- Simple Linear regression → When there is a single input variable (x)
- Multiple linear regression → Multiple input variables

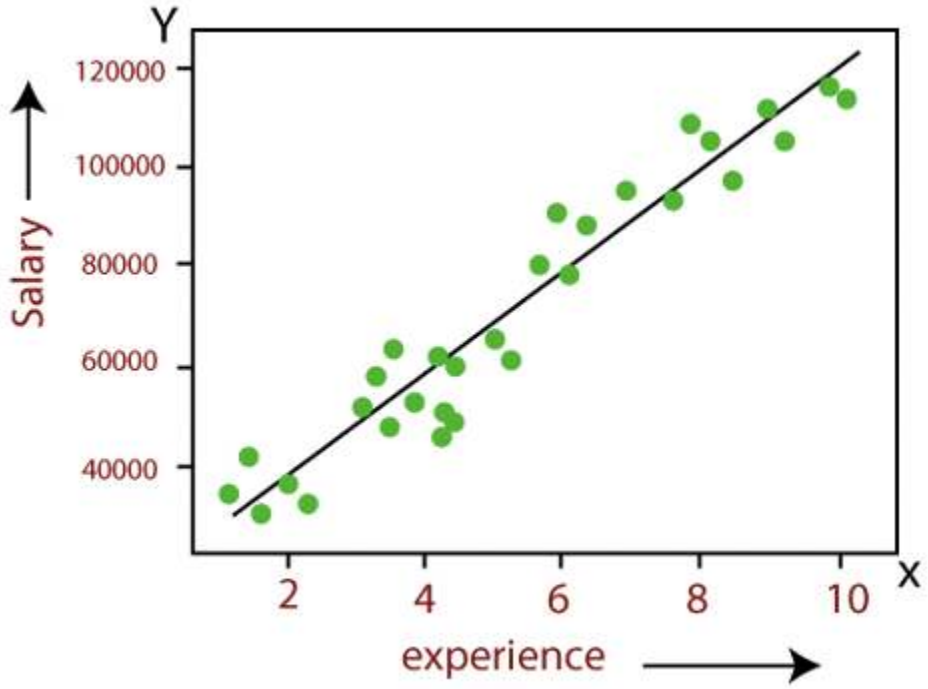
e.g

Y = salary

X = Employees age

A,B: Coefficients of the equation

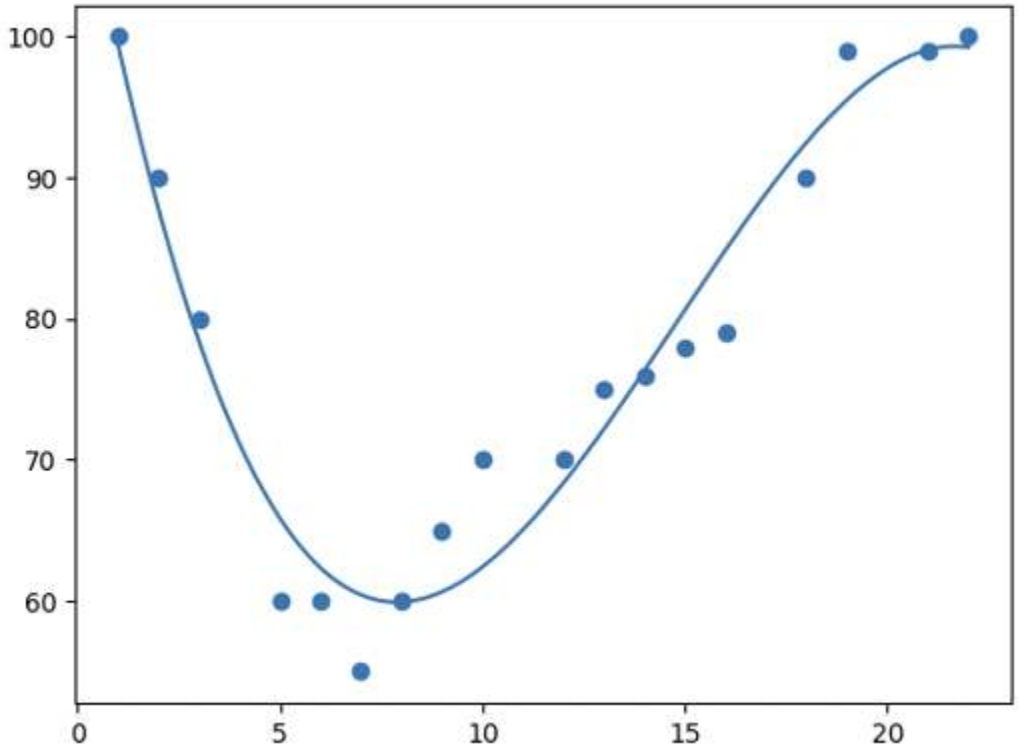
Coefficients are estimates of the unknown. These are calculated for a regression model and help us predict value of Y for each value of X.



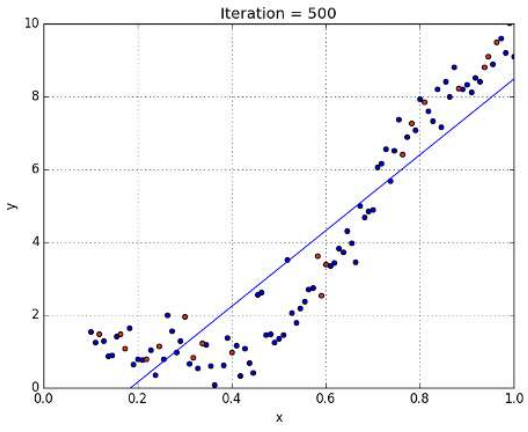
So in order to predict Y (salary) given X (age), we need to know the values of a and b (the model´s coefficients). While training and building a regression model, it is these coefficient which are learned and fitted to training data. The aim of the training is find the best fit line such that cost function is minimized. The cost function helps in measuring error. During the training process, we try to minimize the error between actual and predicted values and thus minimizing the cost function.

2) Polynomial regression

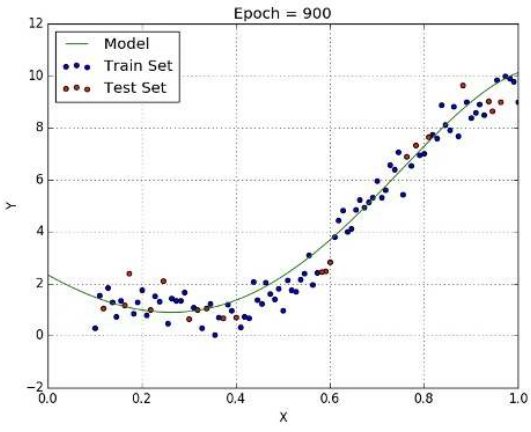
It is also a linear model, however, it’s never a straight line.



Linear regression



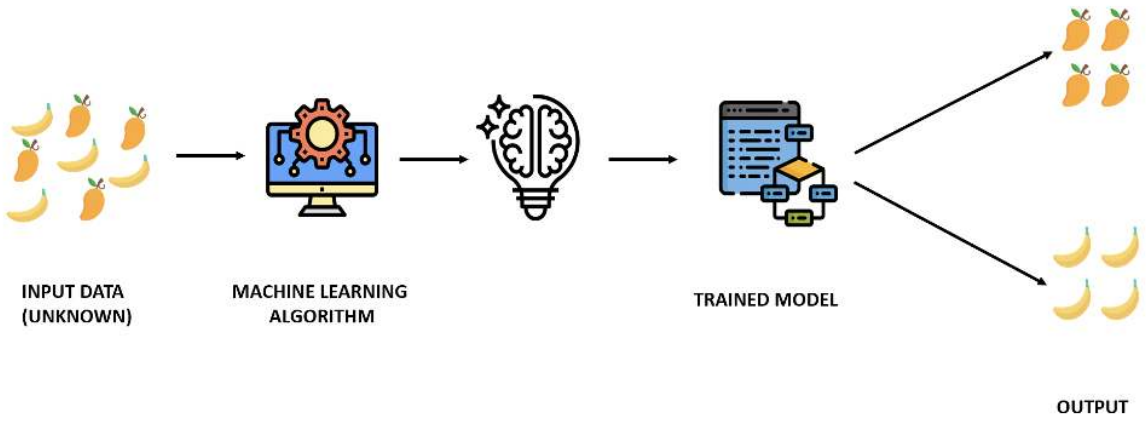
Polynomial regression



Lecture 6

Unsupervised Machine Learning

Looks for undetected patterns in a data set (with no labels and minimum human supervision)



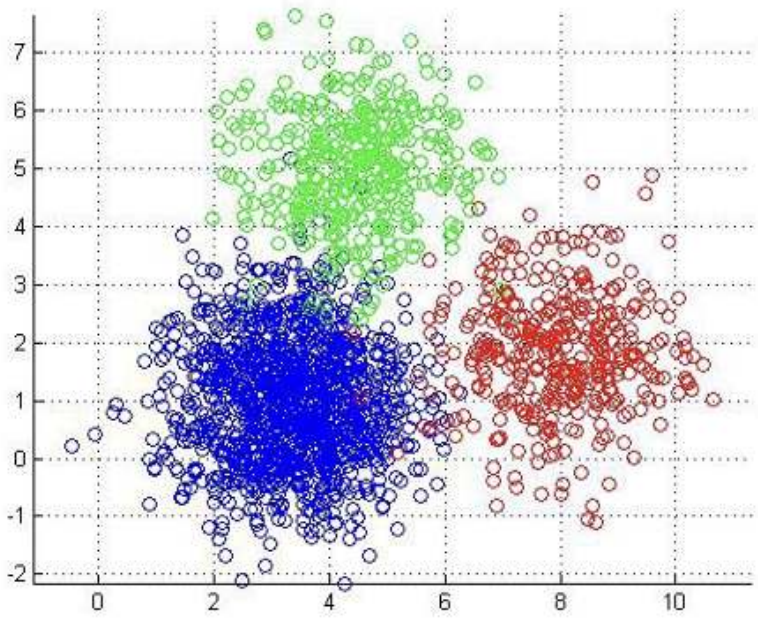
An unsupervised models provides unlabeled data the algorithm tries to make sense of by extracting features and patterns on its own.

1. The training data is unknown or unlabelled
2. The model tries to find patterns and relationships in the dataset by creating clusters in it
3. Model cannot add labels to the data

Types of Unsupervised learning

- Clustering
    - Where we want to discover the groupings in data

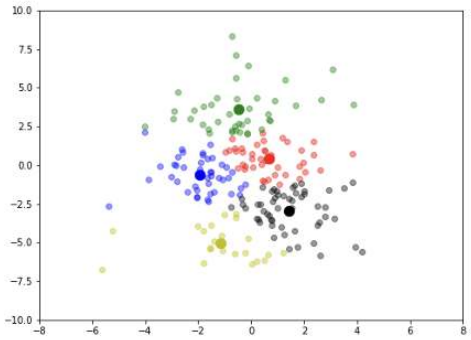
sample Cluster/group

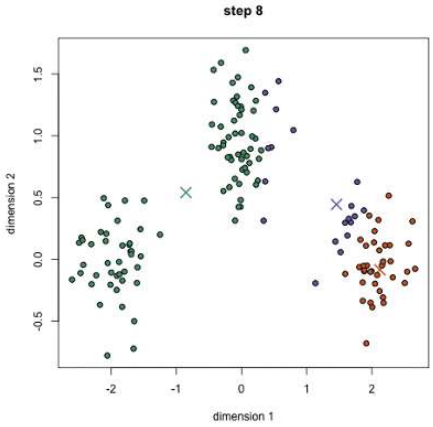


- **K-Means Clustering**
  - K-Means is a distance based algorithm where we calculate distances between data points to assign a point to the cluster.

You will define a target number K, which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. In other words, the K-means algorithm identifies K number of centroids, and then allocates every data point to the nearest cluster, while keeping the

centroids as small as possible. **The “means” in the K-means refers to averaging of the data; that is, finding the centroid**





It will stop when the centroids have been stabilized - there is no change in their values since clustering has been successful

→ A centroid is the imaginary or real location representing the center of the cluster.

→ The defined number of iterations have been reached.

Advantage of K-means

- Very simple to run (Choose K and run it a number of times)
- Most projects do not need quality sensitive clusters

Uses of K-means Clustering

- Document classification
- Customer segmentation
- Fraud detection (insurance and bank)
- Ride share data analysis (uber etc)
- Detection of anomalies
- Sorting sensor measurements

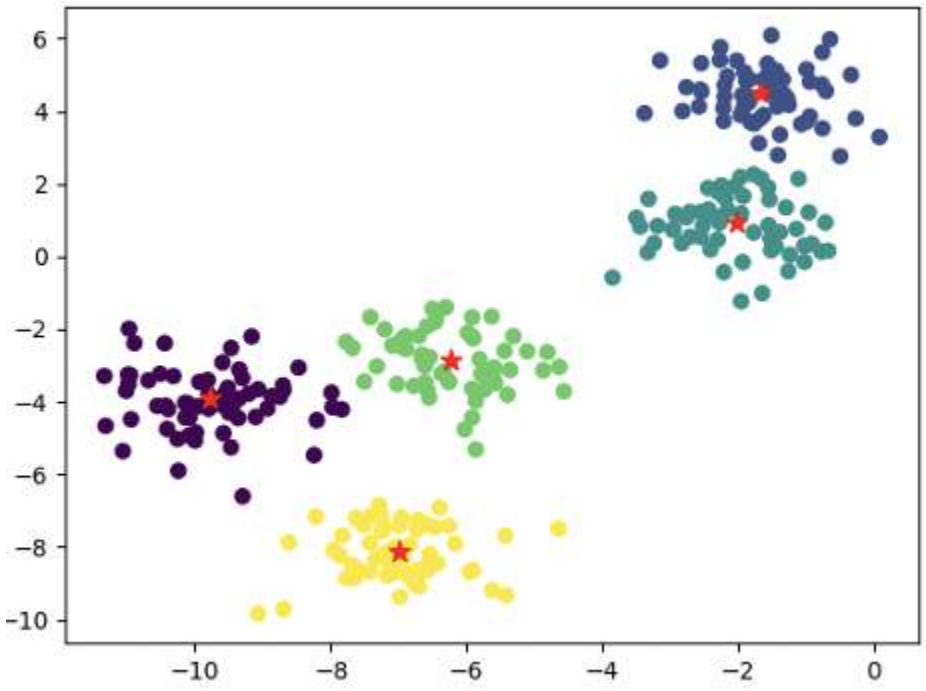
Mean Shift Clustering Algorithm

It locates the heavy density clusters in a data. Mean shift clustering is a sliding-window-based algorithm that attempts to find dense areas of data points. It is a centroid-based algorithm meaning that the goal is to locate the center pointer of each group / class, which works by updating candidates for center points to be the mean of the points within the sliding - window.

→ Window Sliding Technique is **a computational technique which aims to reduce the use of nested loop and replace it with a single loop, thereby reducing the time complexity**.

Uses of Mean Shift Clustering Algorithm:

- Computer vision
- Image processing



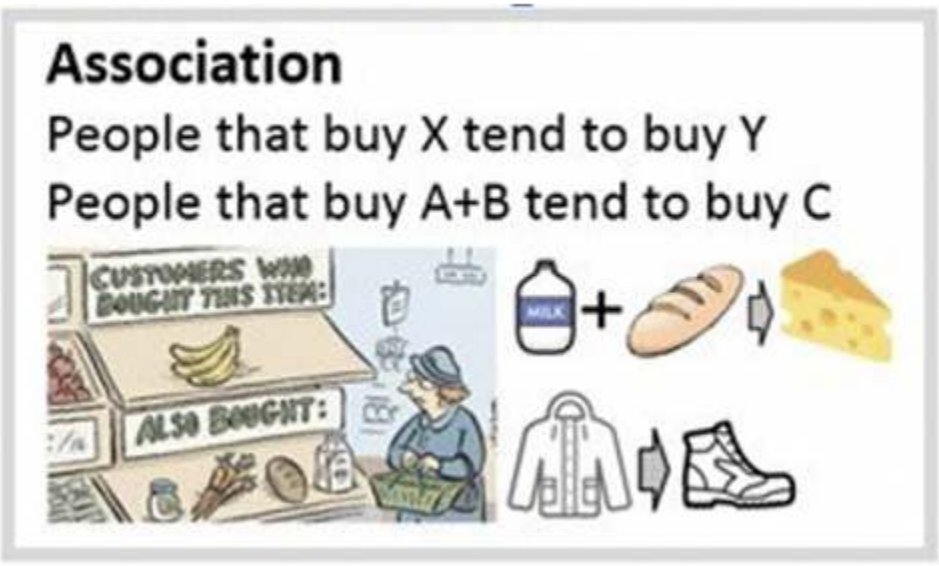
DBSCAN Algorithm

Stands for Density-Based Spatial Clustering of Applications with Noise. This is also a density based algorithm. It separates regions by areas of low-density so that it can detect outliers between the high-density clusters.

Uses-two parameters:

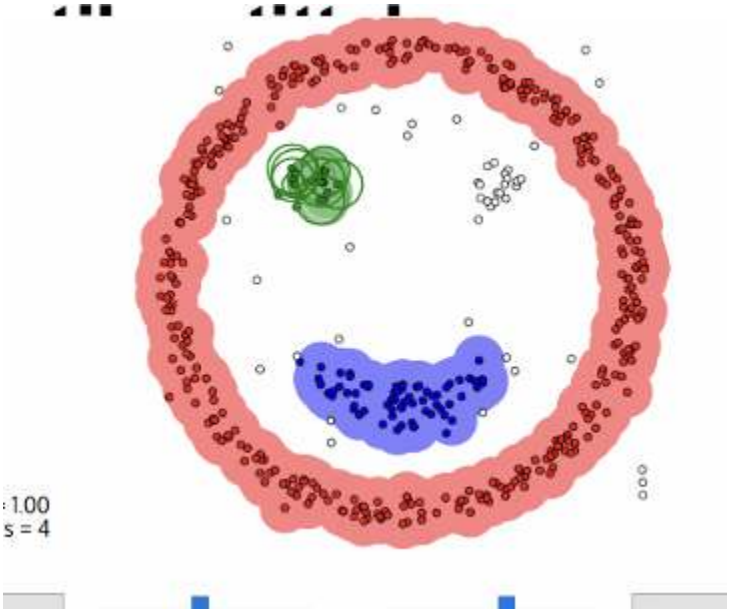
- minPts: the minimum number of data points that need to be clustered together for an area to be considered high-density
- Eps: The distance used to determine if a data point is in the same area as other data points → the minimum number of points to form a dense region, e.g, if we set the minPoints parameter as 5, then we need at least 5 points to form dense region.

- Association
    - When we want to discover rules that describe our data



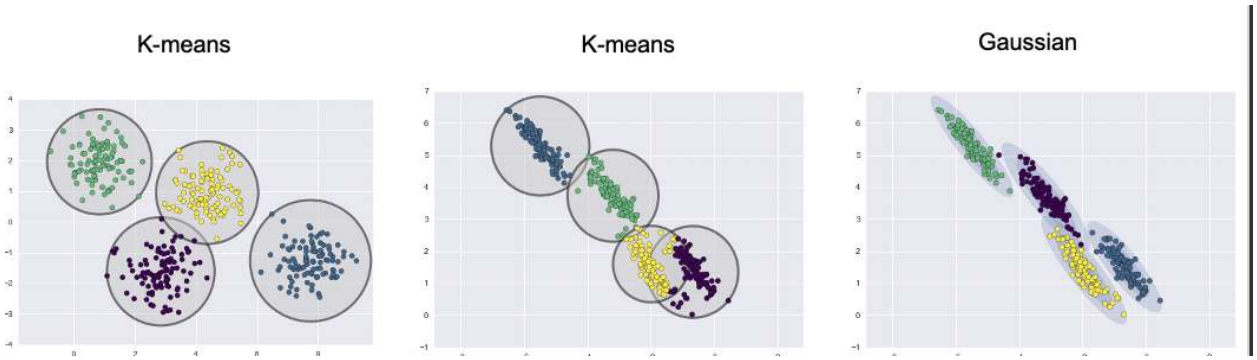
- Store example in USA where they increased sales of beer by putting beers and diapers together. Since they found that men bought beers and diapers on Thursdays
- Frequent items which commonly appear together are: {wine, diapers, soy milk}. Association rule such as diapers → wine. This means that if someone buys diapers, there is a good chance they will buy wine.

Transaction number	Items
0	soy milk, lettuce
1	lettuce, diapers, wine, chard
2	soy milk, diapers, wine, orange juice
3	lettuce, soy milk, diapers, wine
4	lettuce, soy milk, diapers, orange juice



Gaussian Mixture Model

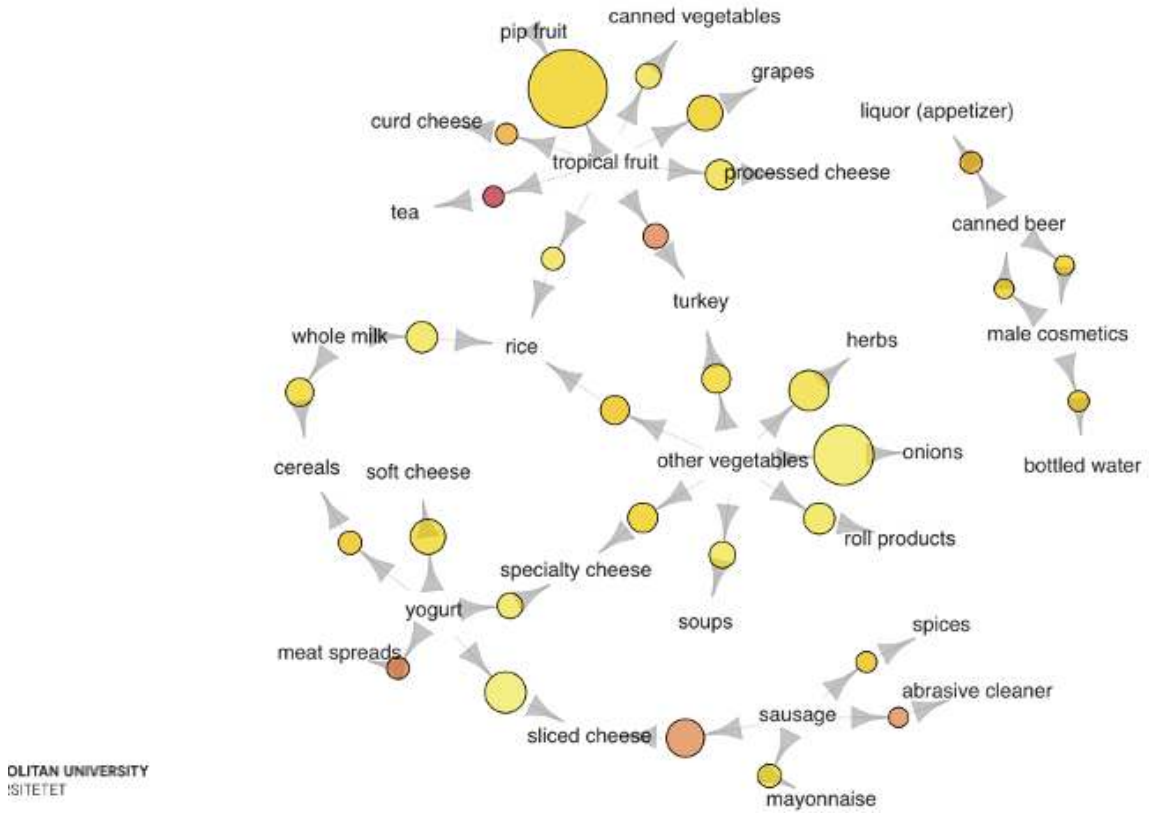
The gaussian mixture model is very similar to K-means however, K-means follows a circular format. Gaussian can take on any format.



Few more clustering algorithms

- BIRCH algorithm
- Affinity Propagation clustering algorithm
- OPTICS algorithm
- Agglomerative Hierarchy clustering algorithm

Association algorithms



We use a dataset on grocery transactions from the arules R library. It contains actual transactions at a grocery outlet over 30 days. The network graph below shows associations between selected items. Larger circles imply higher support, while red circles imply higher lift.

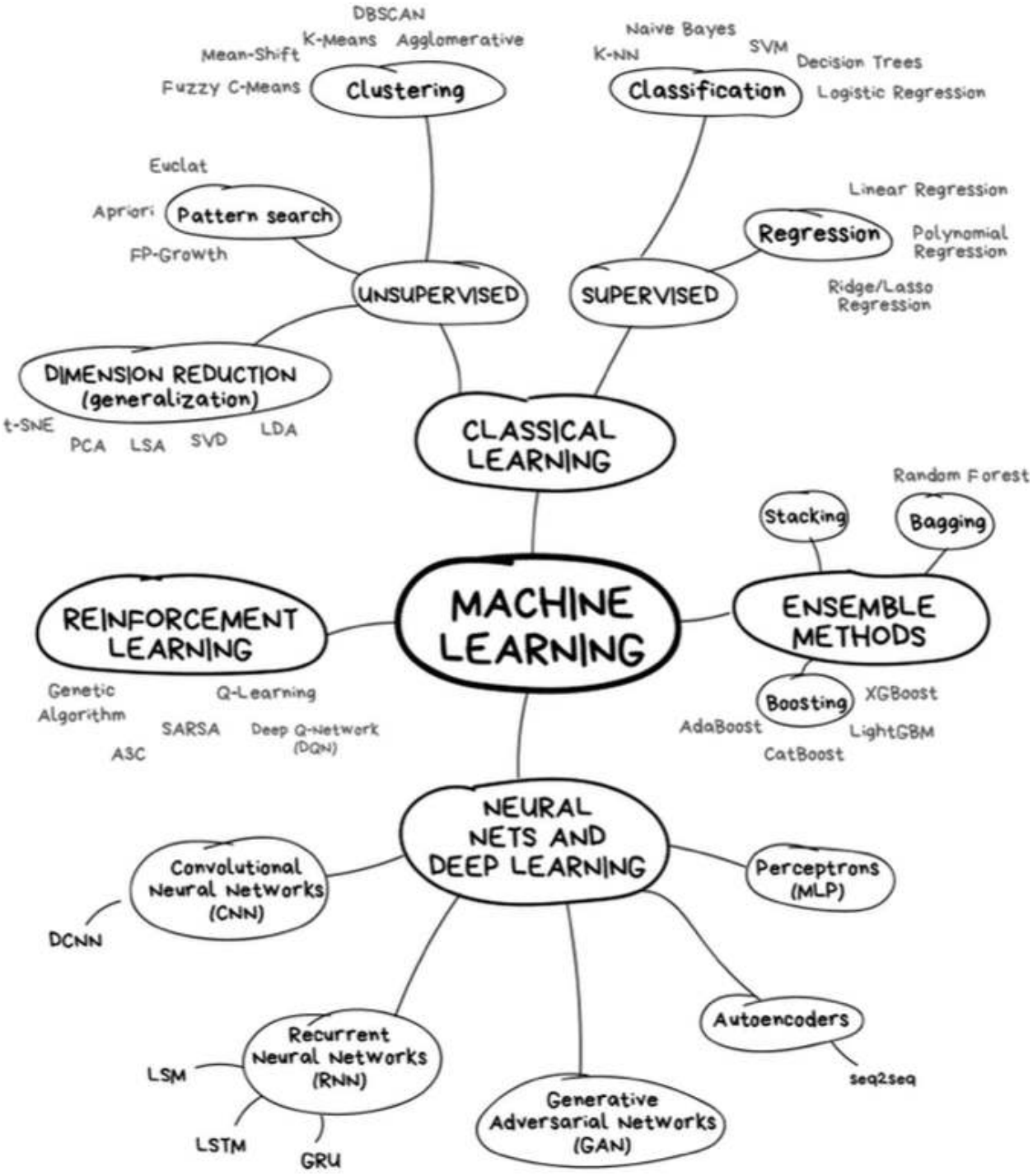
The most popular transaction was of pip and tropical fruits. Another popular transaction was of onions and other vegetables. If someone buys meat spreads, he is likely to have bought yoghurt as well, relatively many people buy sausage along with sliced cheese. If someone buys tea, he is likely to have bought fruit as well, possibly inspiring the production of fruit - flavoured tea.

Apriori algorithm

The algorithm is used for mining frequent item sets and devising association rules. It is created to operate on a database containing a lot of transactions for instance, items

brought by customers in a store.

→ This is the algorithm behind: “you may also like”



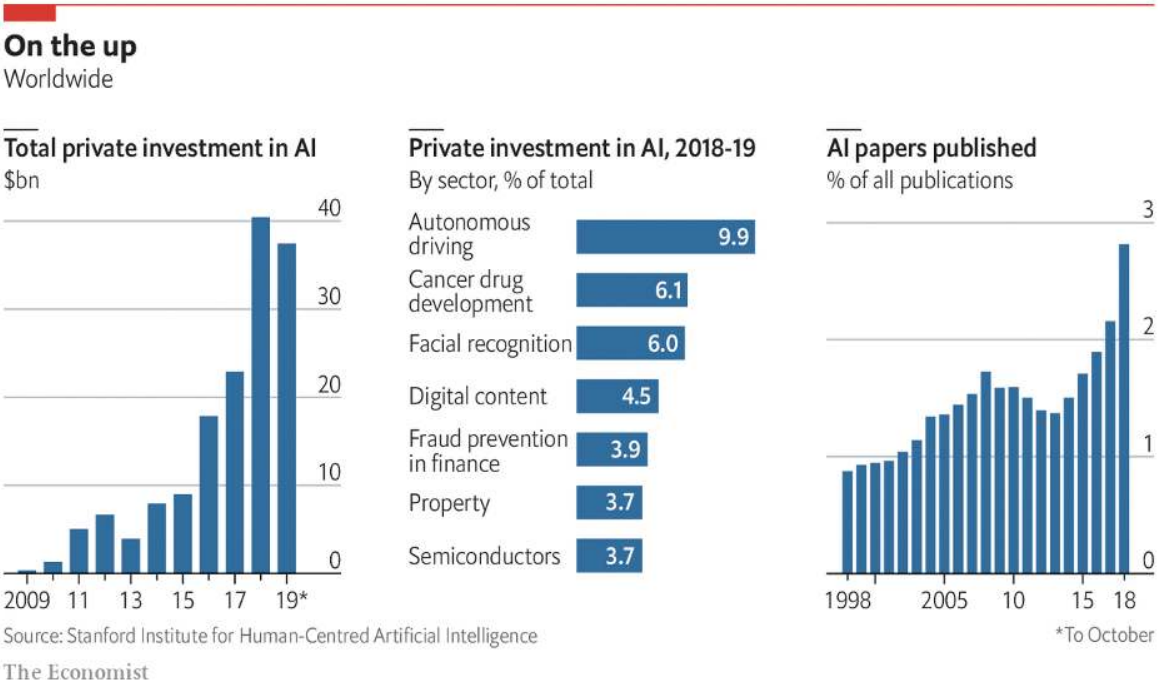
Lecture 7

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/267ec069-41f8-4285-ab7a-09eac416bef2/Limitations-with-AI-DAV_E3625.pptx

Limitations with Artificial Intelligence

Excitement around Artificial Intelligence

- PwC and McKinsey have predicted that AI will add $ 16 trillion to the economy by 2030.
- Google´s boss has described development in AI as more profound that fire or electricity.
- AI is an advanced algorithm: All things which were impossible to do by human programmers can be done by AI.
- Autonomous robots (we will be able to make General AI). AI is the new electricity: its a force which will power everything
- Self driving car companies predict that robo taxis will revolutionalize transport
- In 2016 Geoffrey Hinton, remarked that we should stop training radiologist and instead unleash the power of AI in health.
- An AI firm called BlueDot claims it spotted signs of a novel virus in reports from Chinese hospitals as early as December 2019 (ref economist)
- Almost all companies today either have an AI strategy or are thinking of adopting one



Reality of AI

AI has evolved but it is still far away from the promise of a fully autonomous system. Self driving cars are more capable but still are not safe enough to put on the streets. Use of AI in health is taking longer than expected. There is still a worldwide shortage of human radiologists. Most companies find AI hard to implement.

→ A survey of European AI startups by mmc, a venture-capital fund, found that 40% did not seem to be using any AI at all (ref; Economist)

The hype of AI has far exceeded the science of AI.

1) The state of driverless cars

1. Tesla claims to have autnomous driving (2018) but drivers still need to keep their hands on the wheels
2. General Motors planned to launch self-driving taxis in San Fransisco by 2019
3. In 2019 a self driving car by Uber became the first to kill a pedestrian
4. Waymo in America and WeRide in China are geographically limited and rely on human safety drivers
5. Many driverless car startups going bankrupt because the technology is hard to master

Side Note: Deep learning approaches are fundamentally statistical, linking inputs to output in ways specified by their training data. That leaves them unable to cope with what engineers call “edge cases” - unusual circumstances that are not common in those training data. Driving is full of such oddities.

Some are dramatic: An escaped horse in the road, say, or a light aircraft. One study, for instance, found that computer vision systems were thrown when snow partly obscured lane markings .

Another found that a handful of sticker could cause a car misidentify a “stop” sign as one showing a speed limit of 45mph.

Training an A.I algorithm to achieve the last 10% is much harder than the first 90%

In driverless cars it would be ten thousand times harder.

- Can you trust a driverless car if the success rate of the algorithm is 95%
- Car companies are not talking about driver-assistance tools such as automatic lane-keeping or parking systems, rather than full-blown autonomous cars.

2) Data is harder to come by

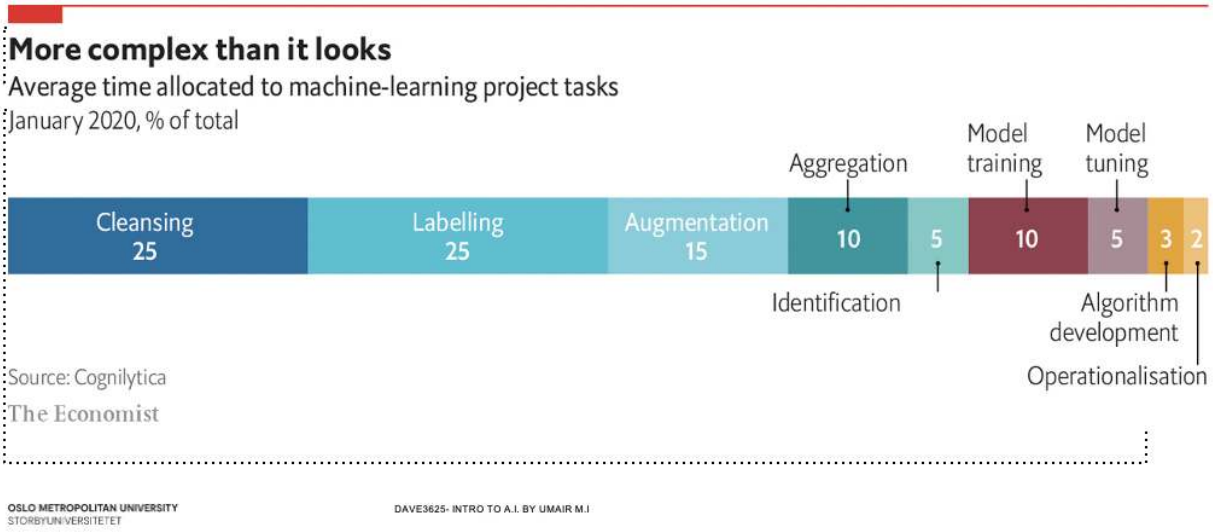
Data issues

1. Amazon Go stores was a unique idea. So unique that it was hard for them to find data to train their algorithms.
    a. There was no such training set featuring people in browsing in shops
    b. They fixed this by creating a virtual shopping environment
    c. The system mush handle crowded stores in which people disappear from view behind other customer. It must recognise individual customers as well as friends or family groups.
    d. Facial recognition fixes: such deficiencies are, at least in theory, straightforward to fix (IBM offered a more representative dataset for anyone to use) → E.g Amazon requirement tool which became sexist towards women.

2. An AI health system designed to work for one hospital might not have the same results for another hospital
    a. Jobs: Everything from hobbies to previous jobs to area codes in telephone number could contain hints that an applicant is likely to be female or young from an ethnic minority

Data Bias:

1. Most facial recognition system today use a higher proportion of white faces as training data (Study by IBM in 2019)
2. Companies today (cannot discriminate on sex, age or race while recruiting people but A.I algorithms can outsmart this process by using variables to reconstruct forbidden information.



Data anonymization

Self driving cars use a lot of virtual reality environments to train their cars since there are not so many self-driving cars. Data **anonymization** is hard and still does not work 100%. Facial recognition system are struggling to identify faces in the covid times where everyone covers their faces.

Data Markets

- Data preparation market was worth more than $1.5 billion in 2019 and could grow to $3.5 billion by 2024 (ref economist).
- The data labeling business could be $4.1 billion by 2024 (ref. economist). D
- Data is the new oil term is not relevant anymore → processed data is the key

3) Hard for businesses to adopt AI

Use of AI in corporations

1. Most AI examples we hear come from big tech giants (e.g Facebook, Amazon, Google, TikTok etc)
    a. Their brilliant algorithms justify their over the top valuations
2. But adoption of AI stops with tech giants
3. Non-tech companies find it harder to see the benefits of using AI (ref Boston consultancy)

Google uses machine learning to refine search results, and target advertisements; Amazon and Netflix use it to recommend products and television shows to watch; Twitter and TikTok to suggest new user to follow.

Feature engineering requires extracting the relevant information from the data and getting it into a single table which can then be used to train a machine learning model. We can group the operations of feature creation into two categories: **transformations** and **aggregation**