Artificial Intelligence and Finance

Hervé BOCO

Introduction

Over the course of the two last decades the phenomenal improvements of the computer capacity as well as the higher speed in communication have opened a new era for AI (Artificial Intelligence).

In particular, AI is rapidly transforming the global financial services industry. As a group of related technologies that include machine learning (ML) and deep learning (DL), AI has the potential to disrupt and refine the existing financial services industry.

Introduction

AI in finance combines the study of the data framework without considering a priori and the use of econometric tools.

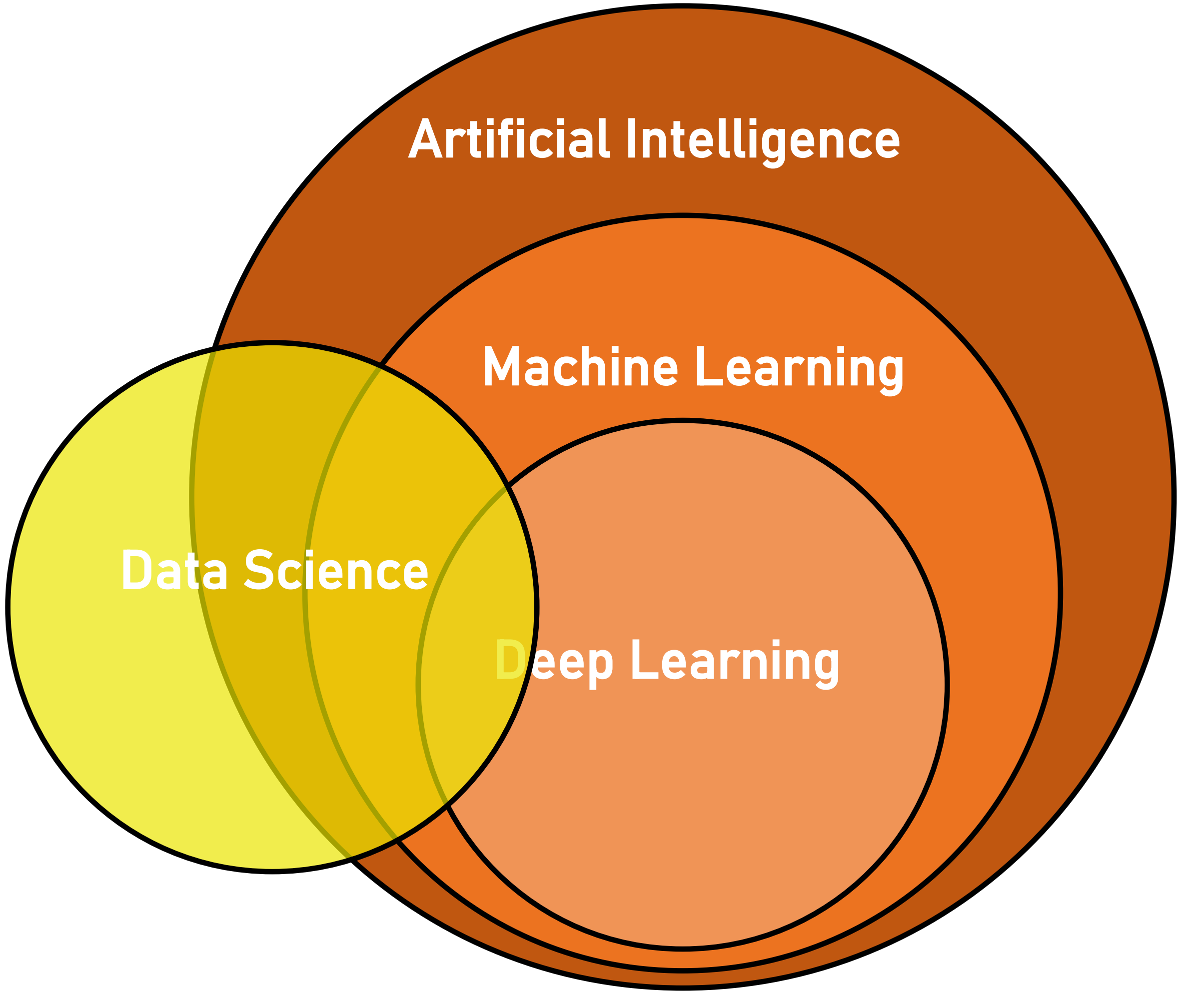
Financial services have pointed out several techniques of AI to understand the underpinning data features.

AI has revolutionised all financial services sectors from costumer relationship to anti-money laundering without forgetting asset trading.

Generally speaking, AI allows financial firms to predict costumer’s behaviour or the future market trends.

AI

Artificial intelligence is the field of study by which a computer (and its systems) develop the ability to successfully accomplish complex tasks that usually require human intelligence. These tasks include, but are not limited to, visual perception, speech recognition, decision making, and translation between languages. AI is usually defined as the science of making computers do things that require intelligence when done by humans.



Definitions

Machine Learning

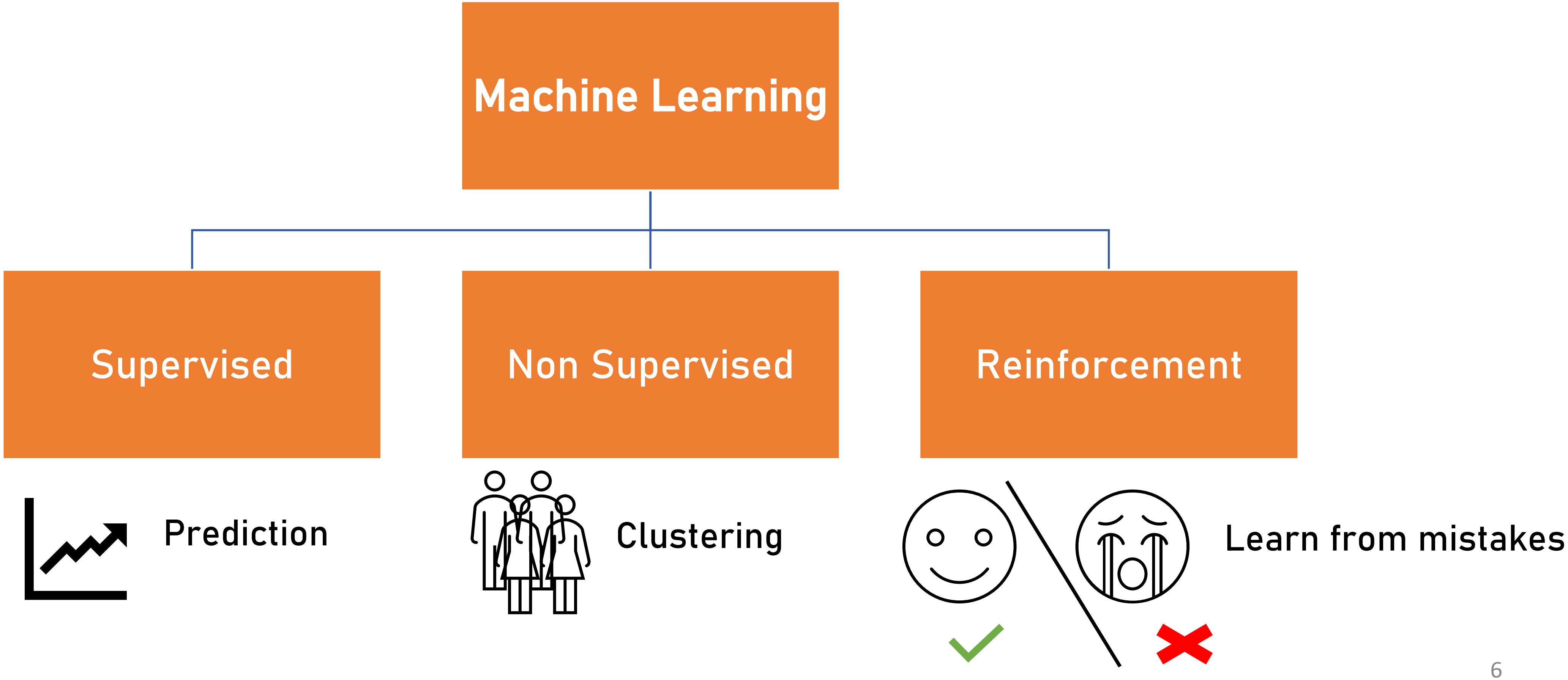
Machine learning is an application of artificial intelligence that provides the AI system with the ability to automatically learn from the environment and apply those lessons to make better decisions. There are a variety of algorithms that machine learning uses to iteratively learn, describe and improve data, spot patterns, and then perform actions on these patterns.

Deep Learning

Deep learning is a subset of machine learning that involves the study of algorithms related to artificial neural networks that contain many blocks (or layers) stacked on each other. The design of deep learning models is inspired by the biological neural network of the human brain. It strives to analyze data with a logical structure similar to how a human draws conclusions.

Machine Learning

There are three types of Machine Learning



Supervised

The main goal in **supervised learning** is to train a model from labeled data that allows us to make predictions about unseen or future data. Here, the term supervised refers to a set of samples where the desired output signals (labels) are already known. There are two types of supervised learning algorithms: classification and regression.

Supervised

Classification

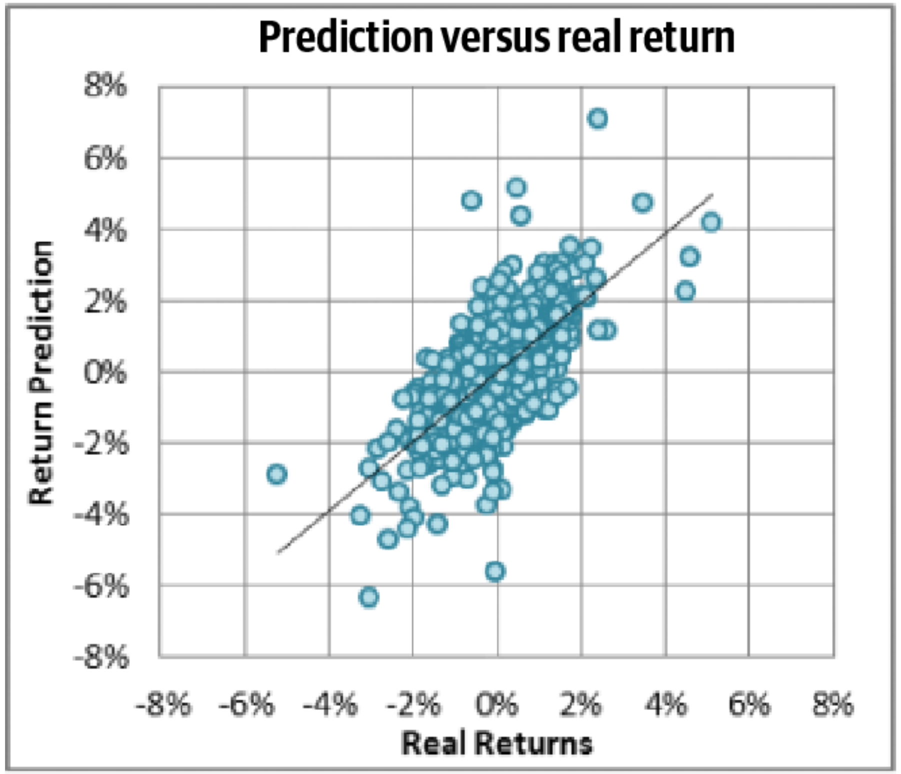
Classification is a subcategory of supervised learning in which the goal is to predict the categorical class labels of new instances based on past observations.

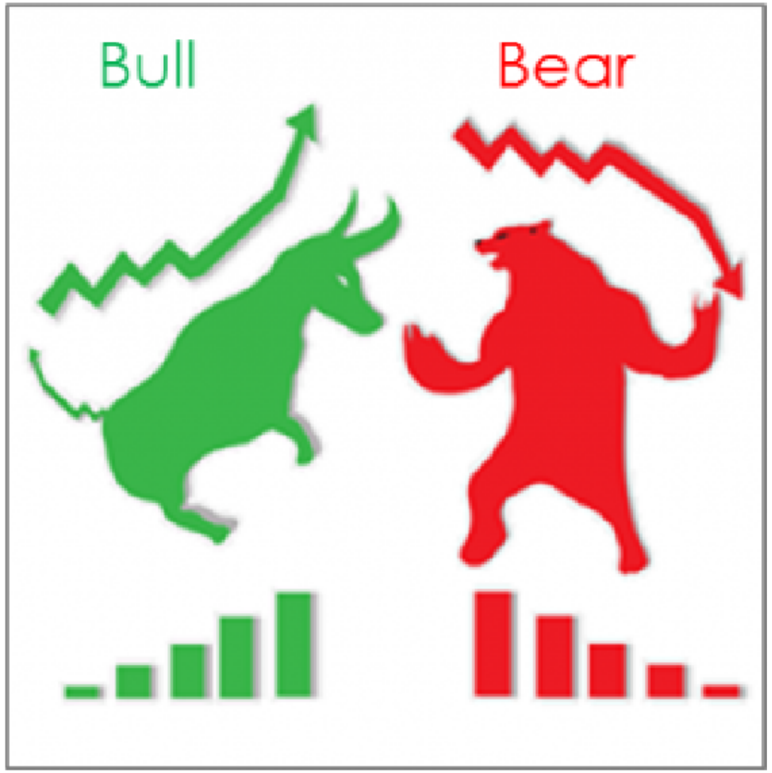
Regression

Regression is another subcategory of supervised learning used in the prediction of continuous outcomes. In regression, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome or target), and we try to find a relationship between those variables that allows us to predict an outcome.

Regression vs classification

Regression **Versus** **Classification**





Unsupervised

Unsupervised learning is a type of machine learning used to draw inferences from datasets consisting of input data without labeled responses. There are two types of unsupervised learning: dimensionality reduction and clustering.

Dimensionality reduction

Dimensionality reduction is the process of reducing the number of features, or variables, in a dataset while preserving information and overall model performance. It is a common and powerful way to deal with datasets that have a large number of dimensions.

Unsupervised

Clustering

Clustering is a subcategory of unsupervised learning techniques that allows us to discover hidden structures in data. The goal of clustering is to find a natural grouping in data so that items in the same cluster are more similar to each other than to those from different clusters.

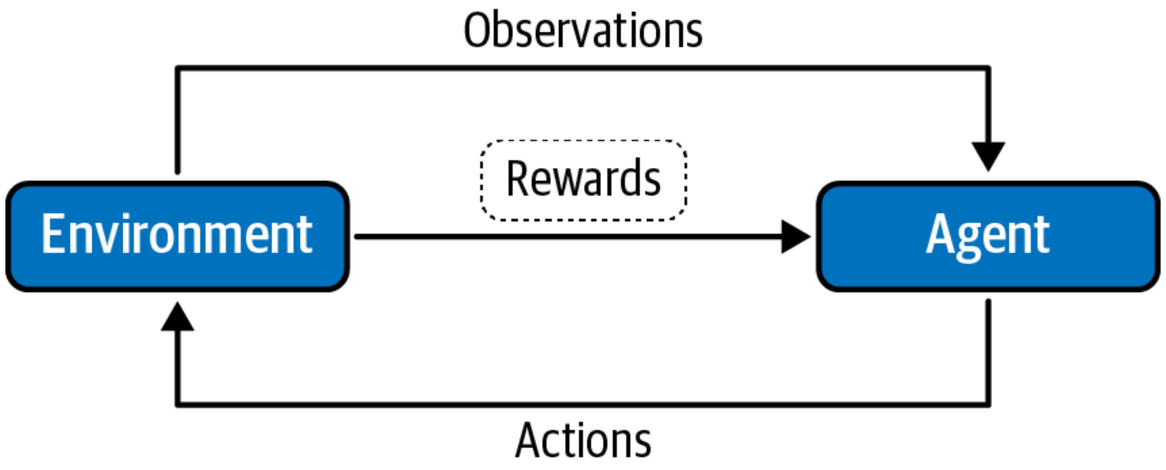
Reinforcement Learning

The steps of the reinforcement learning are as follows:

1. First, the agent interacts with the environment by performing an action.

2. Then the agent receives a reward based on the action it performed.

3. Based on the reward, the agent receives an observation and understands whether the action was good or bad. If the action was good—that is, if the agent received a positive reward—then the agent will prefer performing that action. If the reward was less favorable, the agent will try performing another action to receive a positive reward. It is basically a trial-and-error learning process.



Data Science

Data science is an interdisciplinary field similar to data mining that uses scientific methods, processes, and systems to extract knowledge or insights from data in various forms, either structured or unstructured. Data science is different from ML and AI because its goal is to gain insight into and understanding of the data by using different scientific tools and techniques.

Linear Regression

Y=aX+b

Implementation in Python

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X, Y)
```

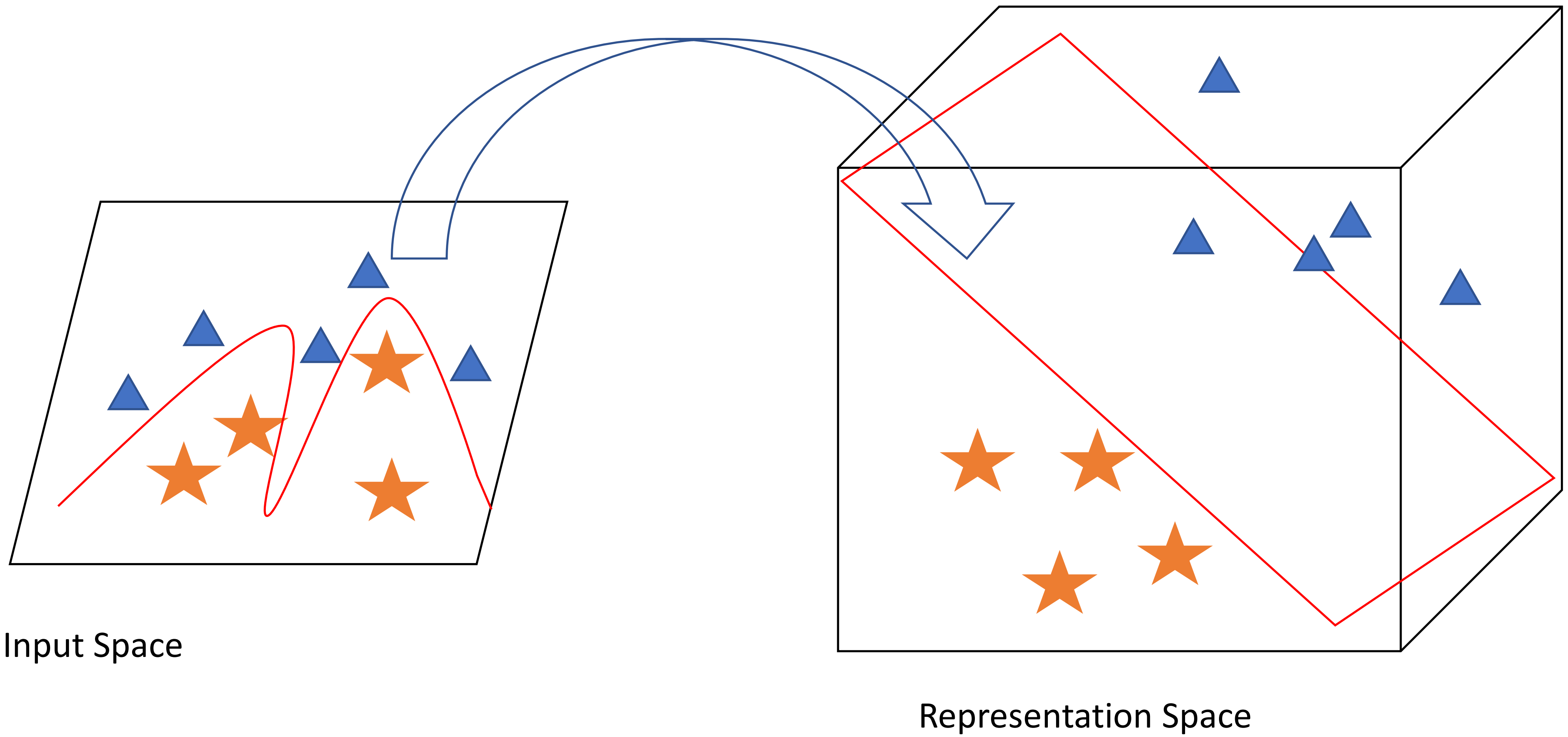
Logistic Regression

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p (*)$$

Implementation in Python

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X, Y)
```

SVM



SVM

Implementation in Python

Regression

```
from sklearn.svm import SVR
model = SVR()
model.fit(X, Y)
```

Classification

```
from sklearn.svm import SVC
model = SVC()
model.fit(X, Y)
```

LSTM

```
from keras.models import Sequential
from keras.layers import Dense
from keras.optimizers import SGD
from keras.layers import LSTM
def create_LSTMmodel(learn_rate = 0.01, momentum=0):
# create model
model = Sequential()
model.add(LSTM(50, input_shape=(X_train_LSTM.shape[1],\
X_train_LSTM.shape[2])))
#More number of cells can be added if needed
model.add(Dense(1))
optimizer = SGD(lr=learn_rate, momentum=momentum)
model.compile(loss='mse', optimizer='adam')
return model
LSTMModel = create_LSTMmodel(learn_rate = 0.01, momentum=0)
LSTMModel_fit = LSTMModel.fit(X_train_LSTM, Y_train_LSTM, validation_data=\
(X_test_LSTM, Y_test_LSTM),epochs=330, batch_size=72, verbose=0, shuffle=False)
```

Content

I AI History

1 Definitions

2 AI pioneering scientists

3 Information Theory

4 Global Growth of AI industry

1 Definitions

The term "artificial intelligence" was coined in ***1956*** by John McCarthy.

Definition :

1. The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making and translation between languages.
2. The theory and development of computer systems able to perform tasks that have traditionally required human intelligence.

2 AI Pioneering scientists

Turing (1950).

He detailed an operational test (the Turing Test) for intelligent behaviour. In his seminal work, Turing provided the major components for future AI work with language, reasoning, knowledge, learning and understanding.

Through the Turing Test, Turing laid the ground work for ML, genetic algorithms and reinforcement learning. The attempt to replicate the logical flow of human decision making through processing symbols became known as the "symbol processing hypothesis"

2 AI Pioneering scientists

McCulloch and Pitts (1943)

They introduced the first a computational model for neural networks.

Hebb

He presented a learning hypothesis based on the mechanism of neural plasticity that became known as Hebbian learning.

Farley and Clark(1954)

They first used computational machines, then called "calculators", to simulate a Hebbian network.

2 AI Pioneering scientists

Rosenblatt (1958)

He created the perceptron which will be the cornerstone of deep technologies 40 later.

Kelley and Bryson (1960)

They established the AI algorithms, More precisely the basics of continuous backpropagation.

3 Information Theory

In 1948 Shannon introduced a new paradigm into all the scientific fields: Information Theory.

The purpose of all science is to process, store and transmit information.

Biology: DNA (Deoxyribonucleic Acid) contains genetic information and determines the phenotype of an organism.

3 Information Theory

Economy: In 1962 Machlup described the information society. Several labels stand for analysing this concept: information economy, post-industrial society, postmodern society, network society, the information revolution, informational capitalism, network capitalism.

One can distinguish five knowledge sectors: education, research and development, mass media, information technologies, information services.

Artificial intelligence is at the core of innovation and progress in these knowledge sectors.

3 Information Theory

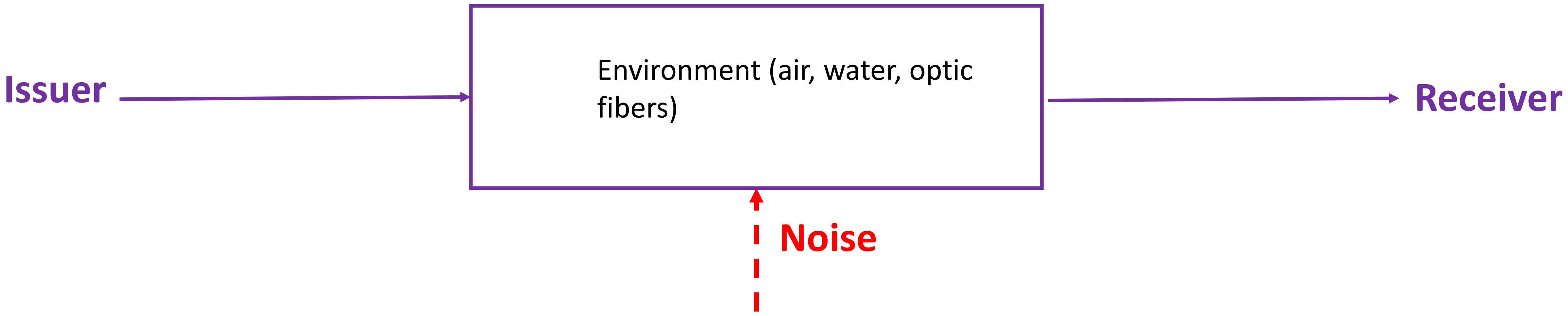
Finance: Informational advantages provide the main sources of market participants' profits. In other hand, the price formation depends crucially on information.

Markets incorporate information into prices.

3 Information Theory

Definition: Information theory studies the transmission, processing, extraction, and utilization of information.

The information transmission is a stochastic process since it randomly depends on the issuer, the receiver and the environment in which the information evolves.



3 Information Theory

One may look for a quantification for information.

Example: We are looking for a blue book in a library.

There are n blue books out of a total of N books in the library.

Question: What is the value of this information?

The smaller the value of n by comparison with N the higher is this information value. Therefore we introduce the concept of quantity of information.

3 Information Theory

Example:

If one book out of two has a blue color then the uncertainty is halved

$$\frac{N}{n} = 2$$

If one book out of ten has a blue color then the uncertainty is reduced by nine-tenths

$$\frac{N}{n} = 10$$

3 Information Theory

Definition:

The quantity of information is defined as:

$$I = -\log_2\left(\frac{n}{N}\right)$$

We choose the logarithm in base 2 in order to I=1 represents a halving of uncertainty

Remark: The quantity of information is always positive

3 Information Theory

The quantity of information is measured in different units:

- Bits
- Shannon
- Nat
- Hartley

A bit is the occurrence, for example, of a 0 or a 1 in a dichotomous question that provides precise information.

We can generalize the previous definition by introducing probability measures. Indeed, if we consider the probability p that a specific information occurs then $I = -\log_2(p)$

3 Information Theory

Messages that convey information that is certain to happen contain no real information I=-log(1). Infrequently occurring messages contain more information than more frequently occurring messages

$$\boldsymbol{p_1} < \boldsymbol{p_2} \rightarrow -\log(p_1) > -\log(p_2)$$

Remarks: objective and subjective probabilities: information provides a distribution of probabilities but probabilities induce a quantity of information but which precedes one or the other?

3 Information Theory

Example:

	Large Companies	Small firms	Total
Strong growth	25	25	**50**
Slight growth	25	75	**100**
Flat growth	25	175	**200**
Recession	25	25	**50**
Total	**100**	**300**	**400**

3 Information Theory

Exercise 1

Calculate the quantity of information :

"The company you are looking for is in strong growth".

Exercise 2

Calculate the quantity of information :

“You are looking for a firm with flat growth".

3 Information Theory

Exercice 3

Calculate the quantity of information :

“You are looking for a large firm with a strong growth".

3 Information Theory

Answer 1

$$I(Strong\ growth) = -\log_2\left(\frac{50}{400}\right) = \log_2(2^3) = 3bits$$

Comment: Uncertainty has therefore been divided by 8. However, in binary data mode, the identification of a number between 1 and 8 requires 3 bits.

$$[000; 001; 010; 011; 100; 101; 110; 111]$$

3 Information theory

Answer 2

$$I(Flat\ growth) = -\log_2\left(\frac{200}{400}\right) = \log_2(2) = 1bit$$

Comment: This information is of lesser importance because there are many more stagnant companies than companies in very strong growth. Here the uncertainty is halved.

3 Information Theory

Answer 3

$$I(Strong\ growth) = 3bits$$

$$I(Large\ firm) = 2bits$$

$$I(Large\ firm + Strong\ growth) = -\log_2\left(\frac{25}{400}\right) = 4bits$$

Comment: We have lost 1 bit of information: when we learn that we are looking for a firm with a strong growth, the probability that it is a large firm increases: one out of every two strongly growing companies is a large firm, whereas only one firm out of four is large. As a result, the remaining uncertainty is only halved by adding the information on the size of the company after it is known that the company is growing strongly. And a division of uncertainty by 2 corresponds well to 1 bit of information.

3 Information Theory

Definition: The entropy *H* is a measure of the informational potential of an information set. $H = \sum_i I_i$

Example

	Female	Male
Number	3	1
Probability	3/4	1/4

$$H = \frac{3}{4}\left(-\log_2 \frac{3}{4}\right) + \frac{1}{4}\left(-\log_2 \frac{1}{4}\right) = 0.8$$

4 Global Growth of AI industry

In the last 60 years the AI field has experienced its share of successes and failures.

Currently, governments around the world are competing to create superior AI facilities and research with a view to AI being a lever for greater economic power and influence.

Between 2012 and 2016 the US invested ***$18.2 billion*** into AI compared with **$2.6 billion** in China and ***$850 million*** in the UK. The Japanese Government Pension Investment Fund (the world's biggest manager of retirement savings) is considering AI to ultimately replace human fund managers. In February 2018, BlackRock announced it would establish an AI lab. With ***$6.3 trillion*** assets under management, the firm already employs text analysis and analyses corporate website traffic and smartphone geolocation data and is now looking at ML to deploy in asset management.

4 Global Growth of AI industry

However, the recent trend has been one of rapid growth. According to a Wushen Institute Report (2017), 5,154 AI startups have been established globally during the past five years, representing a 175% increase relative to the previous 12 years.

There are two explanations for this impressive growth. First, exponential advances in computing power have led to declining processing and data storage costs and secondly, the immense data availability has increased, creating more possibilities in the AI field.

4 Global Growth of AI industry

Historically, the US has dominated the AI industry. Between 2000 and 2016 there were 3,033 AI startups in the US, accounting for ***37.41%*** of the worldwide total (Buchanan and Cao, 2018).

However, the proportion has been decreasing and in 2016 dropped to under ***30%*** for the first time.

During the same period, the US received ***$20.7 billion*** in funding, accounting for ***71.78%*** of the world's total funding (Wushen Institute Report, 2017).

4 Global Growth of AI industry

In 2017 China surpassed the US for the first time in terms of AI startup funding.

In 2012 China accounted for ***48%*** of global AI startup funding and in 2017 the total global AI funding was ***$15.2 billion***. AI equity deals increased ***141%*** relative to the previous year and since 2016 more than 1,100 new AI companies have raised their first round of equity financing. However, the US is losing its global AI equity deal share, decreasing from ***77%*** to ***50%*** of equity deal share during the last five years.

In terms of AI growth, China leads the Asian market. During the past five years China accounted for ***68.67%*** of Asian AI startups and corresponding AI funding was ***60.22%*** of the Asian total. Many Chinese cities and provinces dominate other Asian countries. In terms of the number of AI companies, there are ***454*** in Beijing, ***319*** in Guangdong and 224 in Shanghai compared with ***57*** in Singapore and ***283*** in India (Wushen Institute Report, 2017).

4 Global Growth of AI industry

However, Chinese and American big tech firms have differed in terms of their AI focus.

Microsoft, Google and IBM focus on ML, speech recognition and speech synthesis whereas Tencent, Alibaba and Baidu focus on image recognition and AI searching.

Chinese company Cambricon is developing chips for DL. Ant Financial uses facial recognition for payments at Alibaba owned retail stores.

In 2016, Ant Financial, Foxconn and the city of Hangzhou partnered for the “City Brain” project using AI data from social feeds and surveillance cameras. Additionally, 55 cities participate in the “sharp eyes” project whose surveillance data may end up powering the nation’s Social Credit System, a measure to gauge “trustworthiness”.

II Financial Data

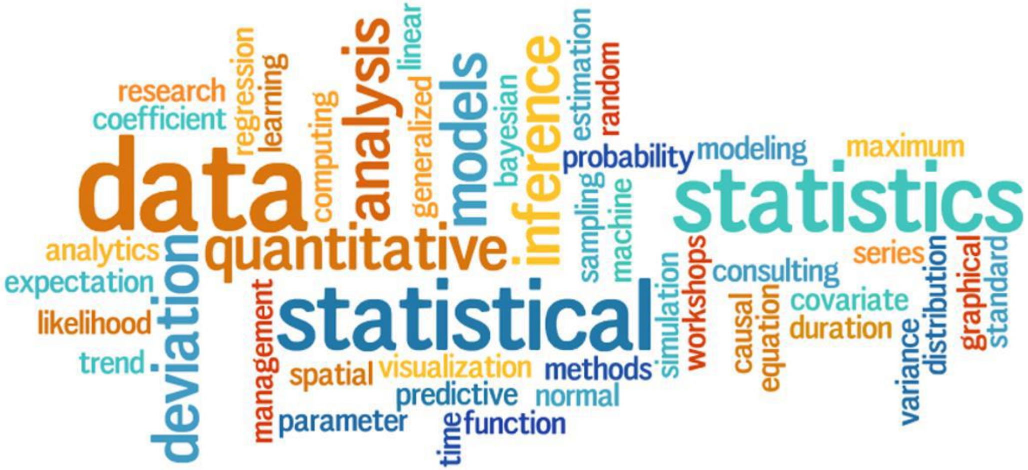
1 Big Data

2 Market Stock news

3 Kyle Model

1 Big Data

FACEBOOK'S HADOOP CLUSTER CONTAINS OVER 100 PB OF DATA
NETFLIX STREAMS OVER 1 BILLION HOURS OF VIDEO MONTHLY
AMAZON OWNS THE 3 LARGEST LINUX DBs WITH DATA STORES UP TO 24.7 TB
DATA CENTER TO END USER TRAFFIC WILL BE 561 EB IN 2013
PayPal
PAYPAL PROCESSES OVER $315M IN PAYMENTS PER DAY
DATA
400,000,000 TWEETS PER DAY ON Twitter
350 MILLION IMAGES UPLOADED DAILY TO FACEBOOK
EBAY PROCESSES OVER 1 BILLION TRANSACTIONS DAILY
AMAZON CLOUD HANDLES 500,000 TRANSACTIONS PER SECOND
72 HOURS OF VIDEO ARE UPLOADED TO YOUTUBE EVERY MINUTE



research
coefficient
regression
learning
computing
analysis
linear
generalized
models
bayesian
inference
estimation
random
data
probability
modeling
maximum
statistics
sampling
machine
analytics
quantitative
expectation
deviation
series
consulting
likelihood
statistical
simulation
workshops
causal
equation
covariate
duration
distribution
graphical
standard
trend
management
spatial
visualization
methods
variance
predictive
normal
parameter
time
function

1 Big Data

Data

- Is the cornerstone of an organization's performance (**decision making**)
- is the company's most important **intangible asset**
- is the **means to success** in a globalised and increasingly competitive economic context
- are tipping companies into **a knowledge-based economy**
- enable **innovation** to be steered

Data: Black Gold for firms

The Internet and social networks are huge providers of "free" data.

The exploitation of digital traces is still in its beginnings.

Illustration: chat bots

1 Big Data

Definitions

Big Data: Big data concerns data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time.

Big data requires a set of techniques and technologies with new forms of integration to reveal insights from data-sets that are diverse, complex, and of a massive scale.

CRM: analyse and manage the interactions with past, current and potential customers. It uses data analysis about customers' history with a company to improve business relationships with customers, specifically focusing on customer retention and ultimately driving sales growth.

Datawarehouse : providing an architectural model for data flow from operational systems to decision support environments. Storage, upload, redundancy, data cleaning, data extracting

1 Big Data

Data Sourcing:

It is the location where data that is being used come from. In a database management system, the primary data source is the database, which can be located in a disk or a remote server. The data source for a computer program can be a file, a data sheet, a spreadsheet, an XML file or even hard-coded data within the program.

Datamining: is defined as a process used to extract usable data. It implies analysing data patterns in large batches of data using one or more software. This helps businesses be closer to their objective and make better decisions For segmenting the data and evaluating the probability of future events, data mining uses sophisticated mathematical algorithms.

Business Analytics: can be defined as systems that combine Data gathering, Data storage, Knowledge management. The purpose of BI is optimization and prediction by using statistic tools

A new challenge

Big Data encompasses a family of tools that satisfy a rule known as the three "V's".

Volume

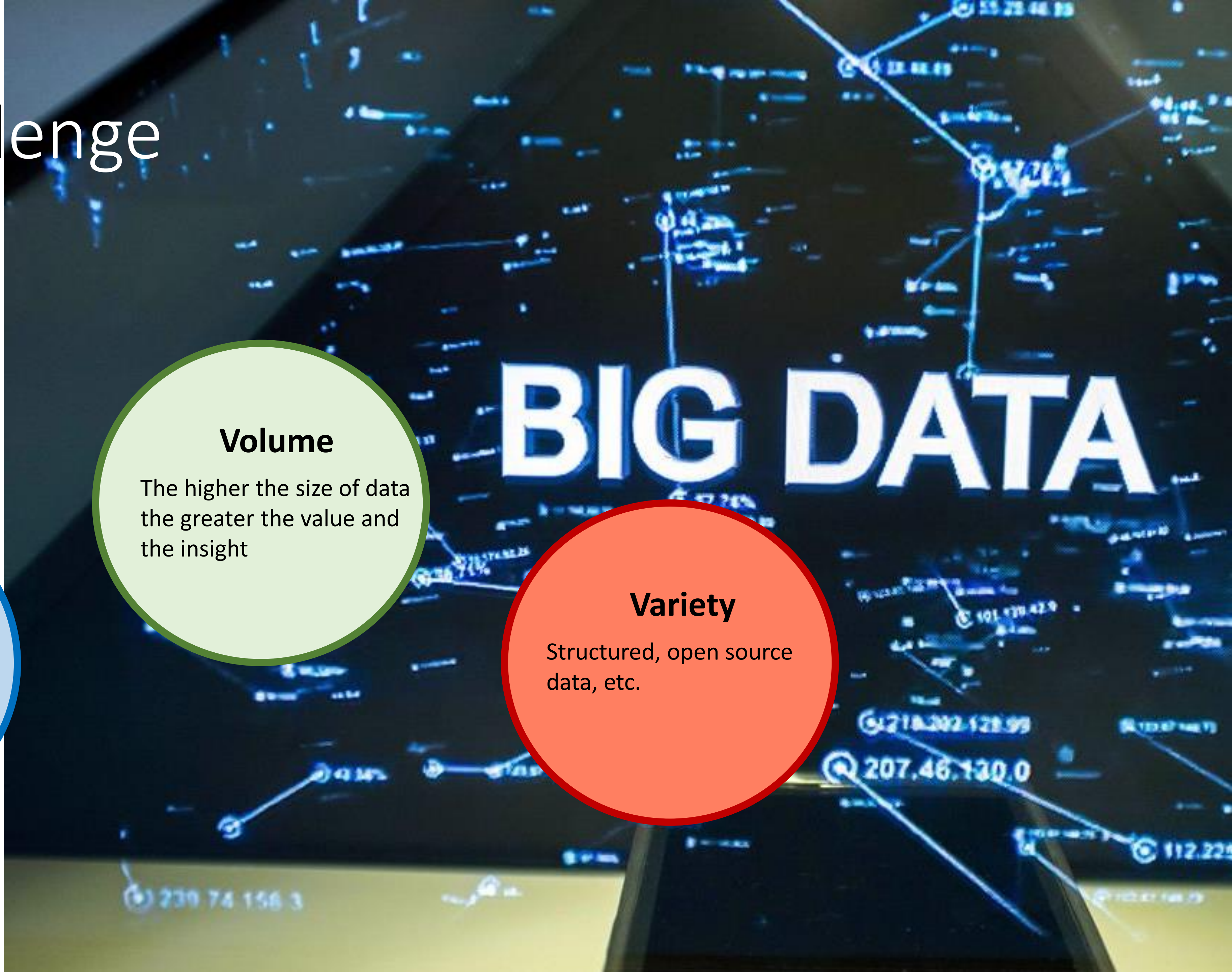
The higher the size of data the greater the value and the insight

Velocity

The speed at which the data is generated and processed

Variety

Structured, open source data, etc.



1 Big Data

- Commercial companies have always collected data to get to know their customers, suppliers, partners and employees better. These are **THE STRUCTURED DATA.**

- Today, the Internet gives access to much more data, which is more difficult to use but very rich. These are THE UNSTRUCTURED DATA (ex: a video)
- The data has several **TYPES** or formats: numerical data, text (or string of characters), categorical data, etc.

2 Market Stock Data

Market participants receive different signals about the fundamental value of the risky assets:

- Earning annoucements
- Dividends payments
- Board changes
- Orders
- Newspapers
- And so on

2 Market Stock Data

Information comes from different sources: Internet, newspaper, phone, firm statement, market figures, etc.

The speed at which information is processed and handled is very crucial. Big Data is linked with HFT.

2 Stock Market Data

- Media
- Website
- Forums
- Etc.

Business Platform:
- Electricity suppliers
- Payments
- Financial Management

Government agencies:
- Social security
- Civil affairs
- Public security
- Industry and Commerce
- Taxation
- Court

Financial institutions:
- Bank
- Insurance
- Markets

Financial Big Data

Text **Speech** **Video** **Image**

2 Market Stock Data

The last two decades have seen the explosion of computerized trading. High Frequency Trading (HFT) is only one aspect of computerized or algorithmic trading. A definition of HFT is quite complex and can be given by describing its properties such as proprietary trading, very short holding periods, submission of a large number of orders that are rapidly cancelled, flat position at the end of the trading day, low margin per trade and the use of co-location services. HFT offers different challenges such as how to measure it and assess its impact on financial markets. According to the literature focusing on the US markets, between ***40% and 70%*** of the trading volume in the US equity markets stems directly from HFT. The European and Asian-Pacic markets are slightly less exposed to HFT as ***38%*** (for the European markets) and between ***10%-30%*** (for the Asian-Pacic markets) of the traded volume is attributed to HFT.

2 Market Stock Data

- The **weak-form** of the Efficient Market Hypothesis claims that prices on traded assets already reflect all **past** publicly available information.

- The **semi-strong** form of the Efficient Market Hypothesis states both that prices reflect all past publicly available information and **new public** information released.

- The **strong-form** of the Efficient Market Hypothesis claims that prices instantly reflect past and public information and hidden private information.

3 Kyle 1985 Model

Assumptions

- We consider a financial market where a single risky asset is traded at time t=1. A the opening of the trading day (time t=0) **N traders** receive a Signal S about the liquidation value of the risky asset $\tilde{v} \sim N(0, \sigma_v^2)$
- In this setup S=v
- Moreover, noise traders submit an order $\tilde{u} \sim N(0, \sigma_u^2)$ independent of $\tilde{v}$
- Market makers set the price by observing the aggregate order flow

$$\tilde{\omega} = \sum_{i=1}^{N} \tilde{x}_i + \tilde{u}$$

- At the end of the trading day the liquidation value is announced to the market.

Results : Equilibrium

- **<u>Proposition 1.</u>**

There exists a unique linear equilibrium in which the reaction of each agent to his private information is $x = \beta\tilde{v}$

The price is given by $p = E\left[\tilde{v}\middle|\tilde{w}\right] = \lambda\tilde{w}$

Results: The reaction

The reaction is given by:

$$\beta = \frac{1}{\sqrt{N}} \frac{\sigma_u}{\sigma_v}$$

Therefore each insider reacts less aggressively to his private information when the competition is fierce (N, the number of insiders).

The reaction is higher when the insider can better dissimulate his private information (thanks to the noise σ_u)

Results: The liquidity

The liquidity parameter is given by:

$$\lambda = \frac{\sqrt{N}}{N+1}\frac{\sigma_v}{\sigma_u}$$

Therefore the liquidity parameter which measures the adverse selection increases with the noise stemming from the liquidity traders. Indeed, the insiders can hide their private information in this case. And the market makers fear to deal with better informed traders.

However, the competition leads the insiders to reveal their private information. And therefore the adverse selection problem drops with the competition

Results: Informativeness of price

The error variance of price is given by:

$$\Sigma = \mathrm{var}(\tilde{v}|\tilde{w}) = \frac{1}{N+1}\sigma_v^2$$

When N=1, at time t=1 the single insider reveals the half of his private information. So the market is not strong-form efficient.

However, when the competition (N) leads the insiders to reveal almost all private information. The error variance of price tends to 0 when N tends to infinity.

Results: Profits

The individual and aggregate expected profits are given by:

$$\begin{cases} \pi_{individual} = \dfrac{1}{\sqrt{N}(N+1)}\sigma_u\sigma_v \\ \pi_{aggregate} = \dfrac{\sqrt{N}}{N+1}\sigma_u\sigma_v \end{cases}$$

When N=1, at time t=1 the single insider realizes a positive profits which increases with the level of noise stemming from the noise traders(σ_u) . So the market is not strong-form efficient.

However, when the competition (N) increases both individual and aggregate profits tend to zero

III AI and Finance applications

1 History

2 Applications

1 History

Development stage	Driving technology	Main landscape	Inclusive finance	Relationship between technology and finance
Fintech 1.0 (financial IT)	Computer	Credit card, ATM, and CRMS	Low	Technology as a tool
Fintech 2.0 (Internet finance)	Mobile Internet	Marketplace lending, third-party payment, crowdfunding, and Internet insurance	Medium	Technology-driven change
Fintech 3.0 (financial intelligence)	Big data, blockchain, cloud computing, AI, etc	Intelligent finance	High	Deep fusion

1 History

1937 Claude Shannon proposes that Boolean algebra can be used to model electronic circuits

1943 McCulloch & Pitts recognise that Boolean circuits can be used to model brain signals

1950 Alan Turing develops the Turing Test

1950 Minsky and Edmonds build the first neural network computer (the SNARC)

1956 The term "artificial intelligence" is coined by John McCarthy

1956 Newell and Simon create the Logic Machine

1957 Economist Herbert Simon predicts that computers would defeat humans at chess within the following decade

1958 Frank Rosenblatt introduces a new form of neural network known as "perceptron"

1958 Early genetic algorithms experiments

1959 Arthur Samuels demonstrates that a computer can play checkers better than its creator, and even play against itself to practice

1 History

1961 Newell and Simons creates General Problem Solver

1964 Computers understand natural language enough to solve algebraic and word problems

1965 Herbert Dreyfus' report severely criticises the emerging AI field

1967 Marvin Minsky predicts that within a generation the problem of creating "artificial intelligence" would be solved

1969 Bryson and Ho develop a back propagation algorithm

1971 Terry Winograd's program SHRDLU answers questions in natural language

1973 UK Lighthill Report ends British government support for AI research

1974 – 1980 First "AI Winter"

1 History

1980 Expert Systems, or Knowledge Systems, emerge as a new field within AI

1980s Early part of decade – Benioff and Feynman create Quantum Computing

1982 PlanPower is conceptualised by Applied Expert Systems (APEX)

1982 James Simons starts quant investment firm Renaissance Technologies

1984 American Association for AI coins the term "AI Winter"

1987 Personal Financial Planning System (PFPS) used by Chase Lincoln First Bank

1987 – 1993 Second "AI Winter"

1 History

1988 David Shaw founds D.E. Shaw and is an early adopter of AI among its hedge funds

1990s The AI industry shows renewed interest in neural networks

1990 Neural net device reads handwritten digits to determine amounts on bank cheques

1993 FinCen puts FAIS (its AI system) into service to monitor money laundering

1997 Deep Blue defeats Garry Kasparov, world chess champion at the time. IBM's stock price increases by $18 billion

2005 The DARPA 132-mile challenge sees AI applied to autonomous driving

2007 The DARPA Urban Challenge

2009 Google's first self-driving car

2010 Flash Crash occurs on 6 May. In 36 minutes, the S&P crashed 8%, before a rebound

1 History

2012 On 1 August , Knight Capital loses $440 million 45 minutes after deploying unverified trading software

2016 Google's DeepMind AlphaGo applies ML algorithms to win at international Go championship

2017 Two Sigma hedge fund which uses ML, crosses the $50 billion in assets under management

2018 The Merkel government announces €3 billion will be spent on AI capabilities

2018 President Macron announces that all algorithms developed for government use will be made publicly available

2018 Alibaba announces plans to bring AI chips to market the following year

2018 Baidu becomes the first Chinese tech giant to join a US led consortium on AI safeguards

2 Applications

AI technology can improve service efficiency and reduce costs.

Wealth management

1. Financial product recommendation

The key to design a good recommendation system is to characterize the preferences of customers and construct personalized behavior models.

2. Robo-Advisor

The core task of Robo-Advisor is to build an optimized portfolio for asset management using risk models, which helps decrease the cost of portfolio construction while improving quality.

2 Application

Risk management

1. Intelligent credit

Credit evaluation is the basis of risk management. Different from traditional credit information systems, intelligent credit systems, which integrate big data and AI, consider financial, government public service, life, and social data.

2. Risk assessment

Compared with traditional approaches that rely heavily on the availability of rich financial information and the experience of experts, AI technology can automatically identify hidden patterns through heterogeneous data sources and can thus achieve improved user profiling when assessing risks.

2 Applications

Risk management

3. Fraud detection

Fraud detection involves monitoring the behavior of populations of users to estimate, detect, or avoid undesirable behaviors.

4. Bankruptcy prediction

Bankruptcy prediction is the art of forecasting bankruptcy and various measures of financial distress of public firms.

2 Applications

Financial identity authentification

Identity authentification is the key to ensure financial information security. Identifying users through recognition, image recognition, voice print recognition, and optical character recognition (OCR) technology would significantly reduce checking costs and improve user experience.

Smart financial consulting

Recent developments on speech recognition and natural language processing enable machine-to-human communication via interactive smart Q&A interfaces and enhance service experience while reducing costs. Generally, chatbot systems can analyze customers’ goals and are highly responsive to customers with personalized advice or tailored answers, such as investment policies and portfolio strategies.

2 Applications

Blockchain

Although the big data era has arrived, data are often mastered in different institutions. Data ownership, data security, and credit intermediaries thus become barriers to data sharing. A new generation of Internet technology is required to solve the problem of information decentralization. Hence, blockchain was introduced. The concept of blockchain was proposed by Nakamoto (2009). Blockchain, which underpins bitcoin, is a digital currency supported by cryptographic methods (a “cryptocurrency”) and is a distributed, publicly available, and immutable ledger.

In finance, blockchain has broad application prospects in digital currency, payment and settlement, intelligent contracts, and financial transactions. Typical applications include bitcoin, litecoin, and other electronic currencies, more secure and open distributed accounting systems, and payment and settlement systems. Consensus mechanisms and security guarantees are important components of blockchain technology.

Machine Learning

1 Principle and methology

2 Linear regression

3 Clustering

AI Modeling

In everyday life, companies model to predict an action. For example, an estate agency wants to estimate the price of a house for its clients. It will do so according to an often implicit model (what it calls its expertise) that will explain the price of a flat according to the surface area, the neighbourhood, the number of rooms, interest rates, and so on.

Another example is when a bank wants to predict whether a client will be able to repay a loan. It will predict this by modelling this capacity according to the client's disposable income, frequency of withdrawals, general level of indebtedness, assets and so on.

The strength of Artificial Intelligence (AI) is to use explicit predictive models based on powerful statistical and mathematical tools. In addition, the computing capacity and speed of computers make it possible to integrate a very large number of variables into AI predictive models and to process very large databases.

Principle

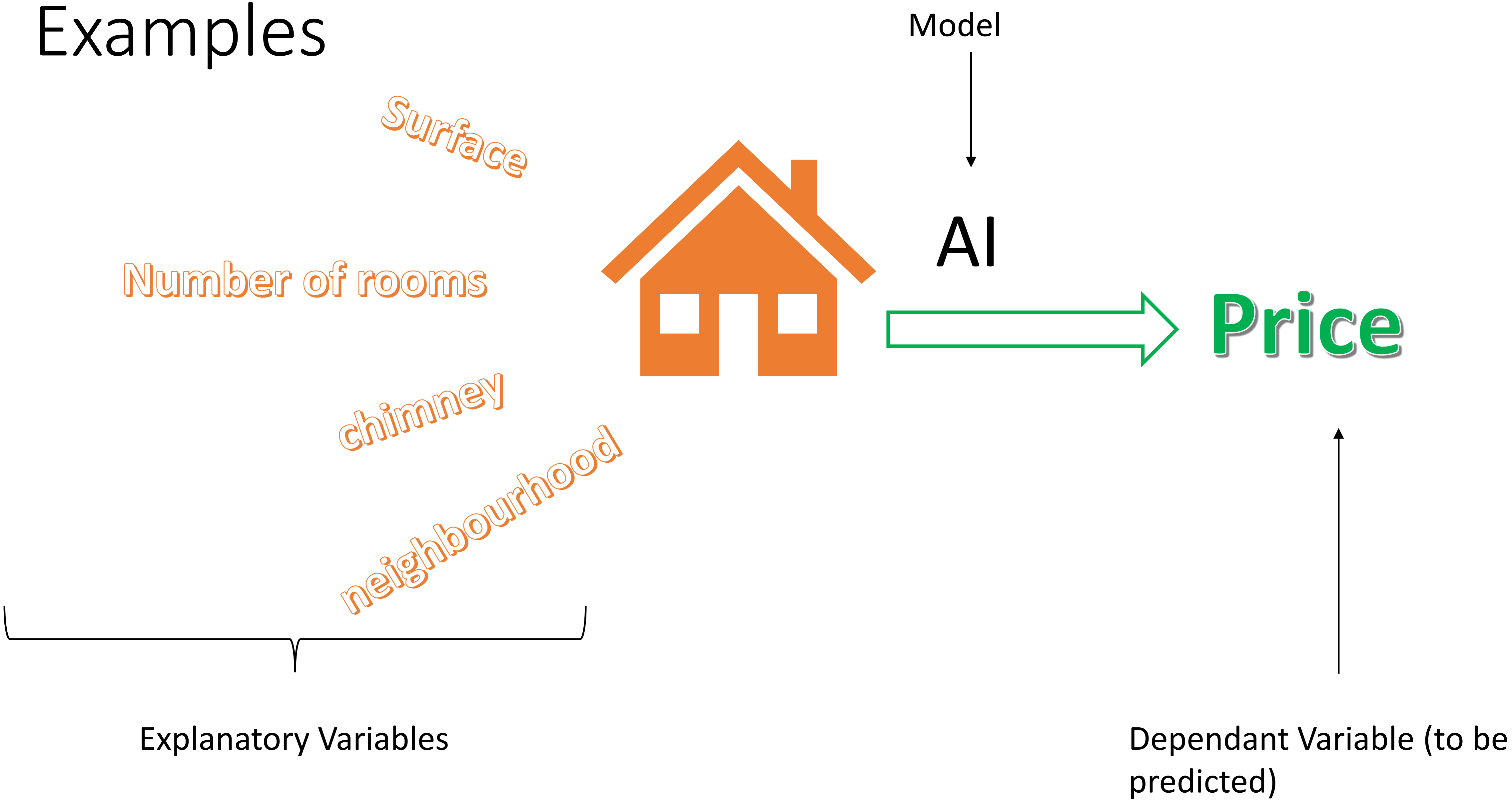
The purpose of an AI robot is to predict a variable (e.g. a sales price, a purchase act, a machine breakdown, the result of an advertising campaign, the evolution of an employee) from other variables (an area, purchase amounts, gender, age, etc.).

In the first approach, we can say that an AI is a "black box" which receives the explanatory variables as input and gives a prediction on the realization of the variable to be explained as output.

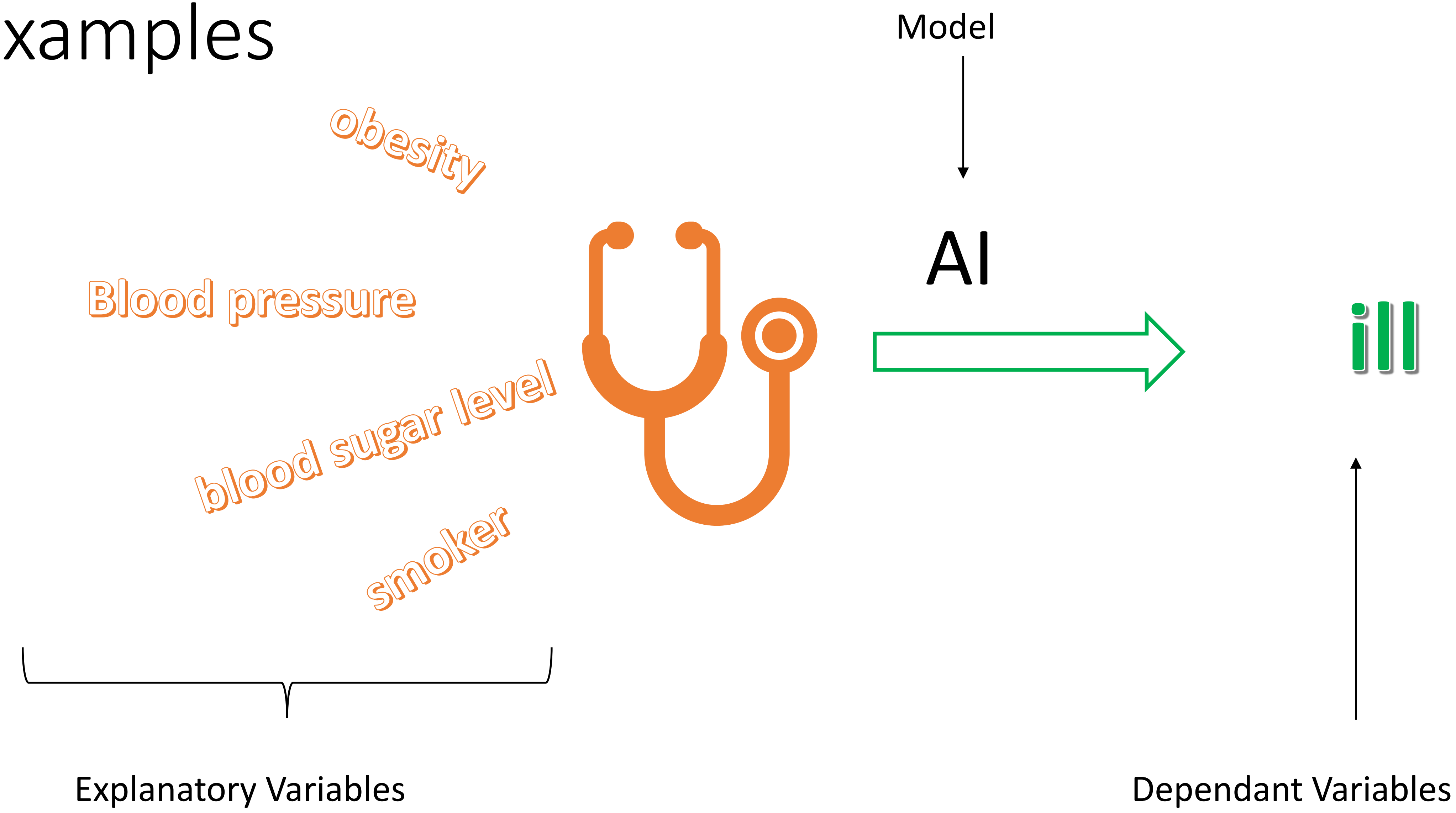
Inside the black box there is a more or less complex architecture which affects a series of weights which weight the importance of each explanatory variable at the input (these weights are called model parameters).

The parameters of the model are determined according to several methods: Deep Learning, Reinforcement Learning, Supervised Learning, Unsupervised Learning

Examples


Model
Surface
Number of rooms
chimney
neighbourhood
AI
Price
Explanatory Variables
Dependant Variable (to be predicted)

Examples


Model
AI
obesity
Blood pressure
blood sugar level
smoker
ill
Explanatory Variables
Dependant Variables

Learning

A human's ability to predict an action or event from a given situation comes from experience, i.e. from examples of situations he has already encountered that are similar or even identical.

AI will also use a database of examples to optimize its internal parameters so that its predictions are as close as possible to reality.

In AI, this example base is called the training data.

When a database is available, it is divided into three parts

- Most of it will be the training data for optimizing the parameters
- Another part will be the test data for checking whether AI with the previously determined parameters statistically succeeds in predicting the correct answers.
- And the third part will be the validation data that allows to compare the differences of different AI models for the same problem.

Machine Learning (Clustering Analysis)

Companies have to deal with diverse and large populations (customers, suppliers, bills, etc.). In order to make the right decisions, they need to divide them into categories.

Automatic classification (Machine Learning, ML) allows them to do this without any a priori on these categories.

In this module we analyze the hierarchical classification which clusters objects by their proximity according to a certain distance. Once the classes have been constituted, they are analyzed in order to carry out an action plan.

Principle

Step 1

We consider a sample consisting of a number of objects, for example 10,000. Each of these objects is characterized by the observations of quantitative variables (qualitative variables need to be recoded into quantitative variables, for example the variable Gender with two modalities becomes the variable Gender with two values 0 and 1).

Step 2

A distance is then built between the objects based on their characteristics. The observations of the variables must be standardized beforehand in order to be comparable.

Step 3

The groups are formed iteratively by bringing the objects closer together one by one according to the distance chosen in step 2

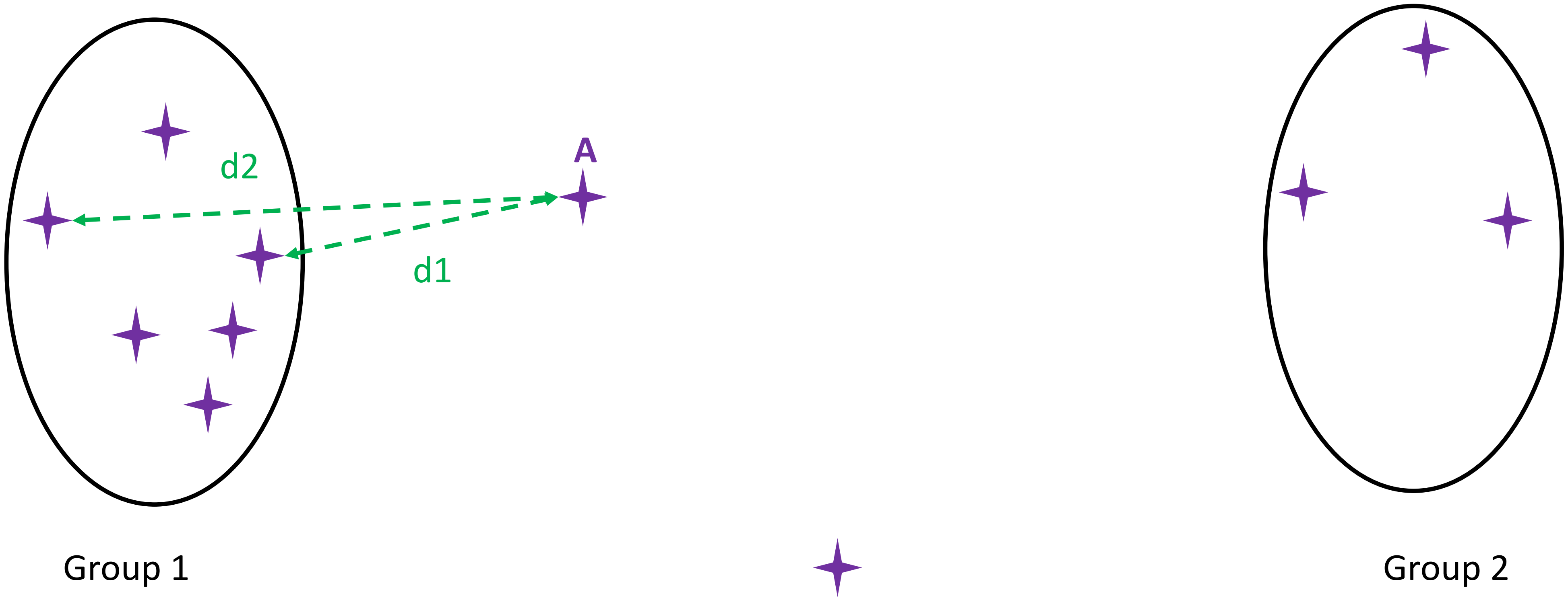
Algorithm

The algorithm is follows:

- Make a table of the distances between all the objects
- Locate the two objects that have the smallest distance that are grouped together in a new class G1.
- Redo a distance table with the new object G1 and all the other objects (this table therefore has one object less since two objects have been replaced by a single G1, so the table has one row and one column less than the previous one).
- Repeat the operation until the desired number of groups has been reached or until all the objects have been put together in the same group.

Remark: the choice of the distance between the group formed by aggregated elements and the other objects is made according to the simple-linkage clustering when the smallest distance between an object and each of the elements of the group is chosen (if the largest distance is chosen, it is called complete-linkage clustering).

Single/Complete-linkage clustering



Single-linkage clustering: we consider the distance d1 between A and Group 1
Complete-linkage clustering : we consider the distance d2 between A and Group 1

Example
Introduction

It is assumed that 4 individuals can be classified according to their purchasing habits.

The table of monthly consumption with monthly salaries

In euros	Food	household appliances	Cleaning products	Leisure	Rent	Salaries
Sophie	800	100	100	50	500	2100
Paul	500	200	90	200	500	1900
Jacques	900	100	100	400	700	2500
Amélie	400	100	150	50	800	1700

Example Solution

1. We will use an ML algorithm (hierarchical clustering). The data must first be standardized. Hence the different observations can be compared, One obtains the following standardized data table:

Standardized data	Food	household appliances	Cleaning products	Leisure	Rent	Salaries
Sophie	0,73	-0,58	-0,43	-0,87	-0,96	0,17
Paul	-0,73	1,73	-0,85	0,17	-0,96	-0,51
Jacques	1,21	-0,58	-0,43	1,57	0,58	1,52
Amélie	-1,21	-0,58	1,71	-0,87	1,35	-1,18

2. We introduce a **distance between individuals** (we choose the Euclidean distance, i.e. we add the differences squared between of each field between two individuals).

Example Solution

3 We have the table of distances :

	Sophie	Paul	Jacques	Amélie
Sophie	0,00	9,18	10,37	15,47
Paul		0,00	17,70	19,00
Jacques			0,00	24,27
Amélie				0,00

End of stage 1: we have a first group formed by Paul and Sophie.

Step 2 we establish a new distance table by replacing Paul and Sophie by Group 1.

Example Solution

We have the new table of distances between individuals

	Group 1	Jacques	Amélie
Group 1	0	10,37	15,47
Jacques		0	24,27
Amélie			0

Thus the following clustering is made between Group 1 and Jacques at the end of stage 2

In step 3, the 4 individuals are grouped together at the end of the algorithm.

4 **Conclusion**: the manager decides the number of classes he is interested in. If he wants 3 then we have (Paul, Sophie) and Jacques and Amélie. If he wants two then we have (Paul, Sophie, Jacques) and Amélie. In the latter case, an analysis shows that unlike the others, Amélie spends most of her money on housing and food and is therefore unlikely to be interested in leisure activities.

Linear Regression

Introduction

- In this chapter we will study two types of regressions: linear regression and multiple regressions.

- Regression is a method which consists to predict one variable from other variables.

- Linear regression consists of finding the best-fitting straight line through the points.

- **<u>Definition:</u>** Describing and evaluating the relationship between a given variable (called the dependent variable Y) and one or more other variables (usually known as the independent variable(s), X_1, X_2, …X_k)

, t=1, 2, …T

$$Y_t = \beta_1 X_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + U_t$$

Some definitions (Summary)

- Simple linear regression is a statistical method that allows us to summarize and study relationship between two continuous variables.

- In general, in linear regression, researchers use this parameters:

- X is the variable regarded as the predictor, explanatory, exogenous or independent variables

- Y is the variable called the response, endogenous or dependent variable.

- β is a parameter vector

- ε is called the error term, disturbance term or noise.

Simple Linear Regression: Definition

- Linear regression consists of finding the best-fitting straight line through the points.

- To do so, linear regression studies the link between 2 quantitative variables (X and Y), the model is:

$$Y = aX + b + \varepsilon$$

- Where

  Y is the dependent variable

  X is the independent variable or the explanatory one

  ε is the standard error term, noise or disturbance term

  a, b are coefficients which are estimated according to the n observations of the samples.

- → It consists to predict Y according to a value X

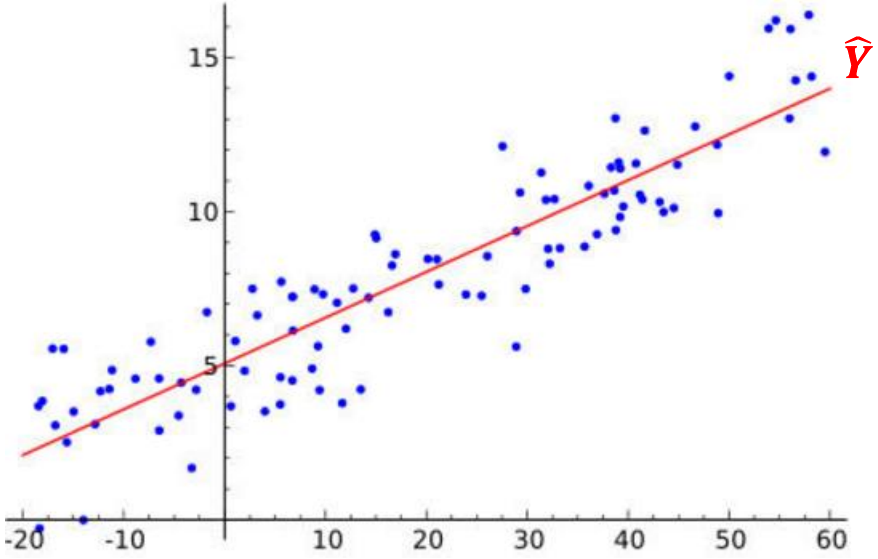
Simple Linear Regression: Scatter Diagram

The scatter Diagram is a good way to see if data follows a **Best Fitting Line.**

The red line is the slope of the Best Fitting Line

The equation of the Best fitting line is :

$$Y = aX + b + \epsilon$$



$\boldsymbol{\hat{Y} = aX + b}$

Simple Linear Regression: Hypothesis of the model

- The standard error estimate is a measure of the accuracy of the predictions
- The errors should be independent of each other.
- The errors should be normally distributed with mean =0 and constant deviation.

Simple Linear Regression: Model Estimation

- The coefficients, a and b are calculated like that :
- First, we have to minimize the equation for the sum of the squared predictions errors:

$$\mathrm{Min}\Sigma(\varepsilon_i^2\ ) = \min\ \Sigma[y_i - (a.x_i + b)]^2$$

- And get the « least squares estimates »:

$$\hat{a} = \mathrm{cov}(X,Y) / Var(X)$$ and $$\hat{b} = \bar{Y} - \hat{a}.\bar{X}$$

- $\hat{Y} = \hat{a}.X + \hat{b}$ are the estimated solution
- $\varepsilon = Y - \hat{Y}$ are called also residuals

Simple Linear Regression Model: Confimation of the Model

- Now, we have seen the model and how we calculate the coefficients a and b.

- The next step of our analysis is to confirm the model. For that we have to do some tests on the coefficient and then we have to study the residu

Simple Linear Regression Model: Coefficient test a

- Here our null hypothesis H_0: a = 0 and our alternative hypothesis is H_1: a ≠ 0

- In this case we have to use Student's test

- The statistic that we will use is: $T = \left(\hat{a} - a\right) \Big/ \sigma_{\hat{a}}$

- Under H_0, T follows a Student distribution T(n-2)
- We reject H_0 when T>2: there is a relationship between X and Y

Simple Linear Regression Model: Coefficient test b

- Here our null hypothesis H_0: b = 0 and our alternative hypothesis is H_1: b ≠ 0
- In this case we have to use Student's test
- The statistic that we will use is: $T = \left(\hat{b} - b\right) / \sigma_{\hat{b}}$
- Under H_0, T follows a Student distribution T(n-2)
- We reject H_0 when T>2: there is a relationship between X and Y

Simple Linear Regression Model: Study of the Residuals

- SSR is the « regression sum of squares » and quantifies how far the estimated sloped regression line, $\hat{y}_i$ is from the horizontal « no relationship line » the sample mean or $\bar{y}$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- SSE is the error sum of squares and quantifies how much the data points vary around the estimated regression line $\hat{y}_i$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- SSTO is the total sum of squaresand quantifies how much the data points vary around their mean. $\bar{y}$

$$SSTO = SSR + SSE$$

Simple Linear Regression Model: Study of the Residuals

- First you should calculate the correlation between the two variables

$$r(x;y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X . \sigma_Y}$$

- $|r(x;y)| = 1$: There is a perfect relationship between X and Y
- r(x;y) = 0 : There is no relationship between X and Y
- If r(x;y) is negative or positive that show us the slope of the relationship

Simple Linear Regression Model: R^2 the coefficient of determination

- We define R^2 / R-Squared or (explained variation)

$$R^2 = \frac{\mathrm{var}(\hat{Y})}{\mathrm{var}(Y)} = 1 - \frac{\mathrm{var}(\varepsilon)}{\mathrm{var}(Y)} = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

- R^2 belongs to [0;1]
- $R^2 = 1$: perfect relationship between X and Y
- $R^2 = 0$: no relationship between X and Y

Simple Linear Regression: Example (1/4)

- On this example we want to understand the relationship between the number of companies and the sales in the industry sector.

- To do so, here
  - X is the number of companies (the independent variable)
  - Y is the sales (dependent variable)

	Number of Companies	Sales
Belgium	216	901,9
Bulgaria	372	1 282,5
Czech Republic	365	3 088,5
Denmark	222	4 725,8
Estonia	143	398,0
Ireland	405	967,4
Greece	371	686,3
Croatia	233	2 721,6
Latvia	277	224,3
Lithuania	122	186,4
Luxembourg	9	71,9
Hungary	446	431,0
Austria	348	2 090,1
Slovenia	97	274,8
Slovakia	176	581,6
Finland	859	1 500,4
Sweden	733	3 688,7
Iceland	33	37,0
Switzerland	212	2 073,5
Macedonia	167	211,2

Simple Linear Regression: Example (2/4)

Summary statistics (Quantitative data):

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Sales	20	0	20	37,000	4725,800	1307,145	1345,390
Number of C	20	0	20	9,000	859,000	290,300	212,710

Correlation matrix:

	Number of Companies	Sales
Number of C	1	0,404
Sales	0,404	1

Regression of variable Sales:

Goodness of fit statistics (Sales):

Observation:	20,000
Sum of weig	20,000
DF	18,000
R²	0,164
Adjusted R²	0,117
MSE	1598032,929
RMSE	1264,133
MAPE	244,286
DW	2,040
Cp	2,000
AIC	287,578
SBC	289,570
PC	1,022

R^2 is close to 0 so the relationship between Sales and the number of companies is not strong.

Simple Linear Regression: Example (3/4)

Analysis of variance (Sales):

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	1	5626837,219	5626837,219	3,521	0,077
Error	18	28764592,730	1598032,929		
Corrected To	19	34391429,950			

Computed against model Y=Mean(Y)

Model parameters (Sales):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	564,442	486,374	1,161	0,261	-457,392	1586,275
Number of C	2,558	1,363	1,876	0,077	-0,306	5,423

Equation of the model (Sales):

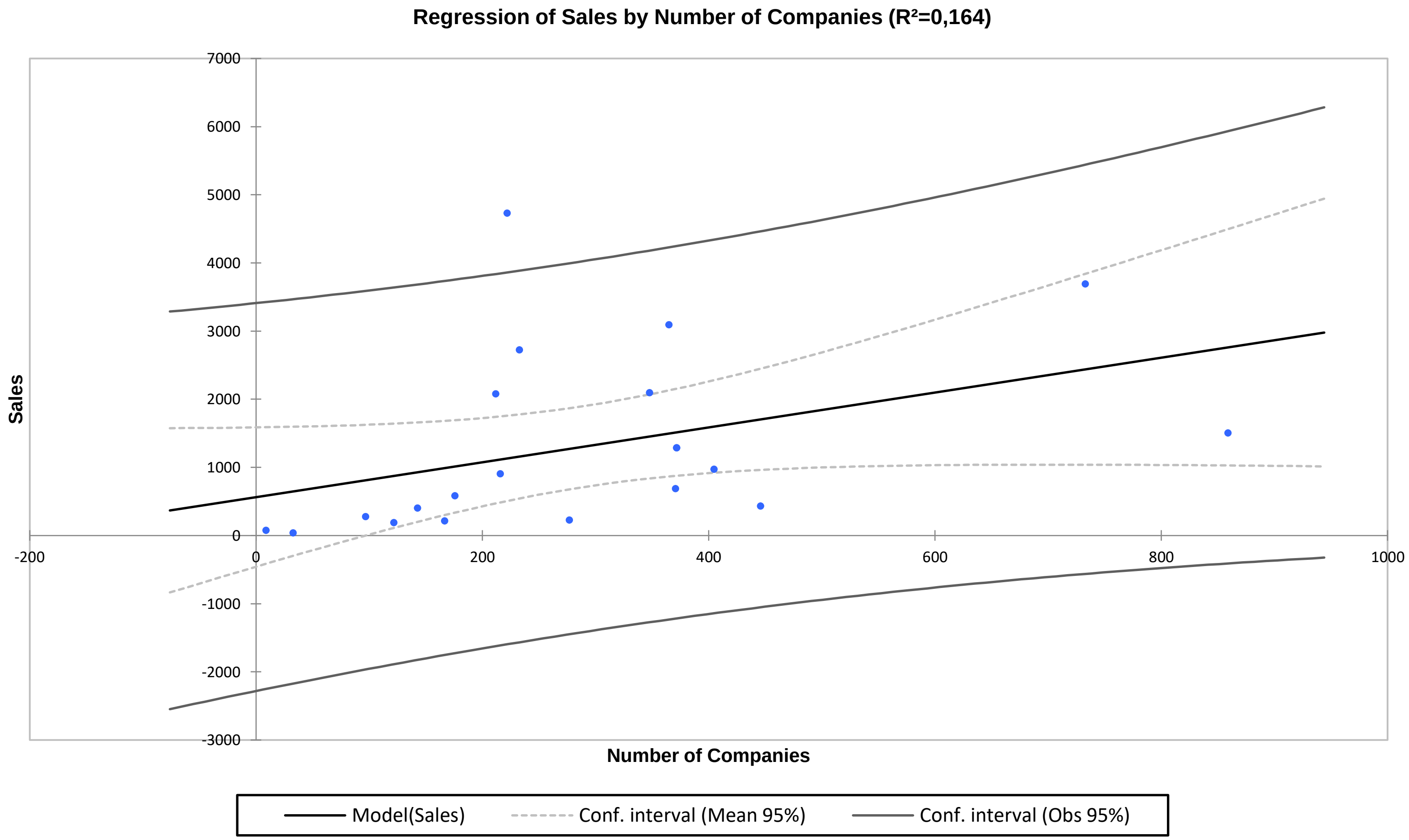
Sales = 564,441798266807+2,55839890366239*Number of Companies

Sales = a * number of companies+ Intercept

Standardized coefficients (Sales):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Number of C	0,404	0,216	1,876	0,077	-0,048	0,857

Simple Linear Regression: Example (4/4)



Multiple Regression: Definition

Multiple regression studies the link between one quantitative variable (Y) and a set of independent quantitative variables X_1 , X_2 , …$X_{n,}$ the model is:

$$Y_t = a_1 X_1 + a_2 X_{2t} + a_3 X_3 + \ldots + a_p X_p + b + \varepsilon$$

With: Y: endogeneous variable

X_1 , X_2 , …X_p : Explanatory variables

b,a_1 , a_2 , …ap : (p+1) coefficients which have to be estimated according to the n observations

- → It is a Generalization of the Linear Regression.

Multiple Regression Regression: Hypothesis of the model

- Hypothesis in the Standard Error Estimate are the same as the linear regression model
  - The standard error estimate is a measure of the accuracy of the predictions
  - The errors should be independent of each other.
  - The errors should be normally distributed with mean =0 and constant deviation.

Multiple Regression Regression: Model Estimation

- The estimates of the a coefficients are the values that minimize the sum of squared errors for the sample
- MSE is the Mean Squared Error: $MSE = \frac{SSE}{n-p}$

estimates the variance of the errors : n is the $S = \sqrt{MSE}$ sample size, p is the number of coefficient

- is known as the residual standard error.

Multiple Regression: Model Estimation

- The coefficients, a_1 , a_2 ,… are calculated like that :
- First, we have to minimize the equation for the sum of the squared predictions errors:

$$\mathrm{Min}\Sigma(\varepsilon_i^2) = \min \Sigma[y_i - (a1.x_i + \ldots + ap.xp + b)]^2$$

- The solutions are $\hat{a}_1, \ldots, \hat{a}_p, \hat{b}$

are the estimated solution

$$\hat{Y} = \hat{a}_1.X_1 + \ldots + \hat{a}_p.X_p + \hat{b}$$

- are called also residuals

$$\varepsilon = Y - \hat{Y}$$

Multiple Regression: Model Estimation

- The same way we did for the linear regression model, here we have to realize some tests on the coefficients and also on the residuals to validate de model.

Multiple Regression: Tests of ANOVA

- For this tests we will use the test of ANOVA with the hypothesis :

$H0: \hat{a}1 = \ldots = \hat{a}p = 0$
$H1:$

there is at least one a ≠ 0

- The only issue here is that if we accept the hypothesis H_1 we do not know what is the coefficient equal to 0

Multiple Regression: Tests of ANOVA

- This is the ANOVA table:

Source	Df	SS	MS	F
Regression	p-1	SSR	MSR=SSR/(p-1)	MSR/MSE
Error	n-p	SSE	MSE=SSE/(n-p)	
Total	n-1	SSTO		

- If the p-value is < 5% then you reject the null hypothesis.
- Here the p-value is $P(F(p;n-p-1)>F\alpha)$

Simple Linear Regression Model: Coefficient test a

- Here our null hypothesis H_0: $a_i = 0$ and our alternative hypothesis is H_1: $a_i \neq 0$

- In this case we have to use Student's test

- The statistic that we will use is: $$T = \left(\hat{a}_i - a_i \right) \Big/ \sigma_{\hat{a}_i}$$

- Under H_0, T follow a Student distribution T(n-p-1)

- We reject H_0 when T>Limit value of the t-test: there is a relationship between X and Y

Multiple Regression Model: Coefficient test b

- Here our null hypothesis H_0: b = 0 and our alternative hypothesis is H1: b ≠ 0

- In this case we have to use Student's test

- The statistic that we will use is: $T = \left(\hat{b} - b\right) \Big/ \sigma_{\hat{b}}$
- Under H_0, T follow a Student distribution T(n-p-1)

- We reject H_0 when T> limit of the value of the t-test: there is a relationship between X and Y

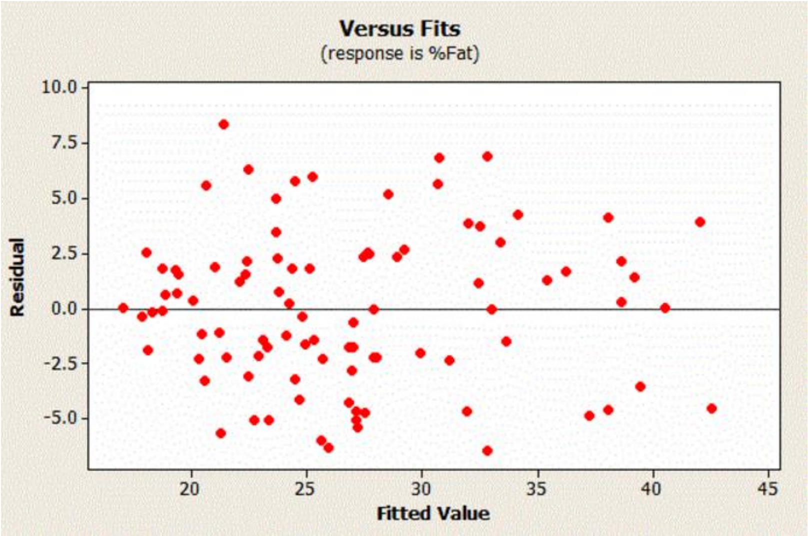
Multiple Regression Model

- To choose the best X_i which fits better to Y you should :

1. Test the model with all the variables.
2. Eliminate one buy one the variables according to the results of the Coefficients tests.

Multiple Regression Model: Study of the Residuals.

- First you should check if the residuals are randomly distributed using graphs such as histograms, scattergrams, etc.



Here, we can see that the residuals are randomly distributed.

Multiple Regression Model: Study of the Residuals

- First you should calculate the correlation between the two variables

$$r(x;y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X . \sigma_Y}$$

- $|r(x;y)| = 1$: There is a perfect relationship between X and Y
- r(x;y) = 0 : There is no relationship between X and Y
- If r(x;y) is negative or positive that show us the slope of the relationship

Multiple Regression Model: R^2 the coefficient of determination

- We define R^2 / R-Squared or (explained variation)

$$R^2 = \frac{\text{var}(\hat{Y})}{\text{var}(Y)} = 1 - \frac{\text{var}(\varepsilon)}{\text{var}(Y)} = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

- R^2 belongs to [0;1]
- R^2 = 1 : perfect relationship between X and Y
- R^2 = 0 : no relationship between X and Y

You can interpret the R^2 as : x% of the variation of Y are explained by the model.

Multiple Regression: Example (1/6)

- On this example we want to understand the relationship between the number of companies, the turnover and the gross direct premium writtens in the financial sector.

- To do so, here
  - X_2 is the gross direct premium writtens
  - X_1 is the number of companies (the independent variable)
  - Y is the turnover(dependent variable)

$$Y = a_1 X_1 + a_2 X_2 + b$$

Multiple Regression: Example (2/6)

GEO/TIME	Number of Enterprises	Turnover	Gross Direct premium writtens
Belgium	0	0	0
Bulgaria	0	0	809,9
Czech Republic	54	6 726,8	6 150,9
Denmark	0	0	0
Germany	669	237 506,0	0
Estonia	13	335,8	325,5
Ireland	0	0	0
Greece	57	3 779,9	3 744,5
Spain	271	60 822,0	56 052,0
France	306	218 255,0	184 299,0
Italy	137	108 362,0	105 310,7
Cyprus	0	0	0
Latvia	9	389,9	348,4
Lithuania	11	331,2	330,5
Luxembourg	0	0	0
Hungary	30	2 586,4	2 562,5
Netherlands	197	39 903,0	0
Austria	48	18 686,4	16 327,9
Poland	0	0	0
Portugal	44	5 983,7	5 544,5
Romania	41	1 816,2	1 809,7
Slovenia	17	2 191,7	1 922,5
Slovakia	18	1 956,9	1 937,2
Finland	0	0	0
Sweden	187	19 436,8	17 908,6
United Kingdom	0	0	0
Iceland	0	0	0
Norway	51	16 177,8	16 036,0
Switzerland	185	90 817,2	44 320,1

Multiple Regression: Example (3/6)

Summary statistics (Quantitative data):

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Turnover	28	0	28	0,000	218255,000	22396,811	47363,369
Number of En	28	0	28	0,000	306,000	65,964	90,895
Gross Direct	28	0	28	0,000	184299,000	17585,668	39901,745

Summary statistics (Quantitative data / Validation):

Variable	Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Std. deviation
Turnover	1	0	1	335,800	335,800	335,800	0,000
Number of En	1	0	1	13,000	13,000	13,000	0,000
Gross Direct	1	0	1	325,500	325,500	325,500	0,000

Correlation matrix:

	Number of Enterprises	Direct premium wri	Turnover
Number of En	**1**	0,730	0,791
Gross Direct	0,730	**1**	0,973
Turnover	0,791	0,973	**1**

Multicolinearity statistics:

	Number of Enterprises	Direct premium writtens
Tolerance	0,466	0,466
VIF	2,144	2,144

Multiple Regression: Example (4/6)

Regression of variable Turnover:

Summary of the variables selection Turnover:

Nbr. of variable:	Variables	MSE	R²	Adjusted R²	Mallows' Cp	Akaike's AIC	Schwarz's SBC	Amemiya's PC
2	Number of Enterpri	97391264,897	0,960	**0,957**	3,000	517,866	521,862	0,046

The best model for the selected selection criterion is displayed in blue

Goodness of fit statistics (Turnover):

Statistic	Training set	Validation set
Observations	28,000	1,000
Sum of weigh	28,000	1,000
DF	25,000	-2,000
R²	0,960	65535,000
Adjusted R²	0,957	
MSE	97391264,897	
RMSE	9868,701	
MAPE	47,904	0,000
DW	2,256	
Cp	3,000	
AIC	517,866	
SBC	521,862	
PC	0,050	
Press	4207437694,133	
Q²	0,931	0,000

R² is R² is close to 1 so the relationship between Sales and the number of companies is strong.

Analysis of variance (Turnover):

Source	DF	Sum of squares	Mean squares	F	Pr > F
Model	2	58134013148,785	29067006574,392	298,456	**< 0,0001**
Error	25	2434781622,422	97391264,897		
Corrected To	27	60568794771,207			

Computed against model Y=Mean(Y)

Type I Sum of Squares analysis (Turnover):

Source	DF	Sum of squares	Mean squares	F	Pr > F
Number of Er	1	37890999951,987	37890999951,987	389,060	**< 0,0001**
Gross Direct	1	20243013196,798	20243013196,798	207,852	**< 0,0001**

Multiple Regression: Example (5/6)

Model parameters (Turnover):

Source	Value	Standard error	t	Pr > |t|	›wer bound (95›	›per bound (95%)
Intercept	-1205,667	2332,344	-0,517	0,610	-6009,218	3597,885
Number of Er	89,933	30,596	2,939	**0,007**	26,920	152,945
Gross Direct	1,005	0,070	14,417	**< 0,0001**	0,861	1,148

Equation of the model (Turnover):

Turnover = -1205,66650936702+89,9325374704178*Number of Enterprises+1,00480355767474*Gross Direct premium writtens

Sales = a1 * number of companies+a2*gross direct premium+ intercept

Multiple Regression: Example (6/6)

- XLSTAT do the interpretation of the regression.
- You can find it at the end of all the regression test.

Interpretation (Turnover):

Using the Best model variables selection method, 2 variables have been retained in the model.

Given the R2, 96% of the variability of the dependent variable Turnover is explained by the 2 explanatory variables.

Given the p-value of the F statistic computed in the ANOVA table, and given the significance level of 5%, the information brought by the explanatory variables is significantly better than what a basic mean would bring.

Based on the Type III sum of squares, the following variables bring significant information to explain the variability of the dependent variable Turnover: Number of EnterprisesGross Direct premium writtens.
Among the explanatory variables, based on the Type III sum of squares, variable Gross Direct premium writtens is the most influential.

Methodology

- Define the variables of interest : $Y, X_1, X_2, ..X_p$
- Global reliability of the model
- Calculation of model coefficients
- Reliability of each model coefficient
- Goodness of fit
- Assumptions to be checked on residuals of the model
- Conclusion

SVM

SVM (Support Vector Machine)

Since 1995, the development of algorithms on SVMs has given extremely powerful results, sometimes superior to those obtained by neural networks.

SVMs are the most powerful Machine Learning techniques for automatically classifying individuals between those who are, for example, granted a loan or not between those who are loyal customers or not between those who will be sensitive to an advertising campaign or not.

This is why SVMs are the AI techniques used primarily in bioinformatics, in finance, computer vision.

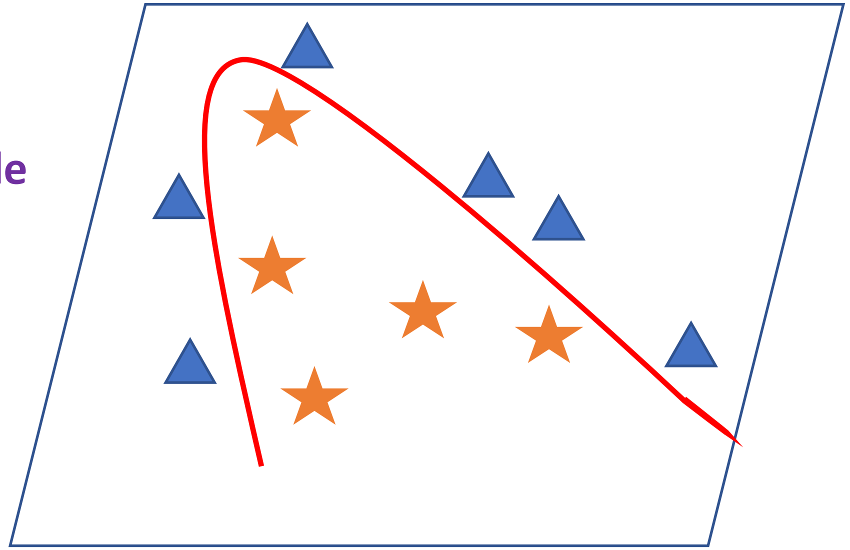
The advantage of SVMs is that they adapt directly to empirical data without assuming a prior architecture (as in neural networks) nor assuming a probabilistic distribution of observations (as in regressions).

Principle

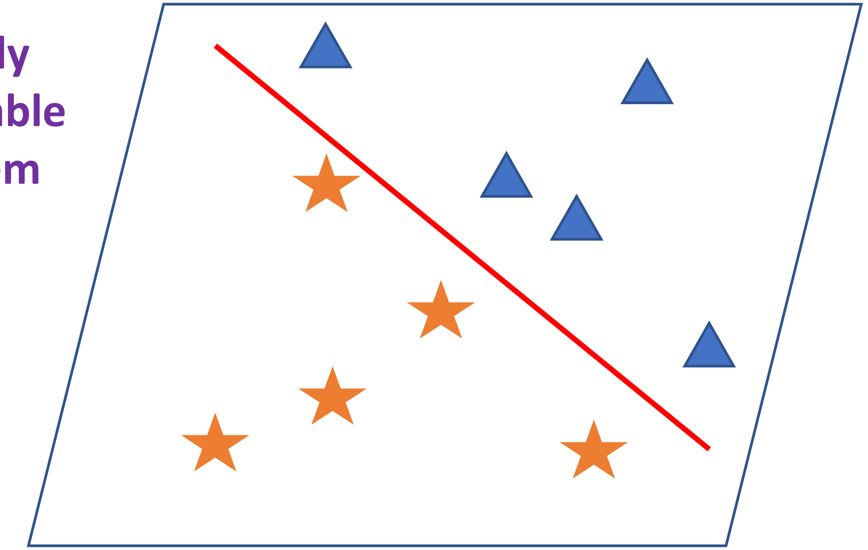
Robust and efficient optimization algorithms use quadratic functions to classify individuals. However, these functions can only solve linearly separable classification problems.

SVMs use powerful algorithms for any type of classification problem by transforming, if necessary, the data to make them linearly separable.

Non-linearly separable problem



Linearly separable problem



Algorithm

Step 1

Change of space: Feature Space

The observations of the variables of all the individuals in the database constitute the initial space (space whose dimension is equal to the number of variables)

A kernel function transforms the original data into new data which are entered in a new space called the feature space. This space has a larger dimension than the original space. In the feature space the individuals are now linearly separable.

The kernel functions generally used are :

- Linear kernels
- Gaussian kernels
- Polynomial kernels

Principle

Step 2

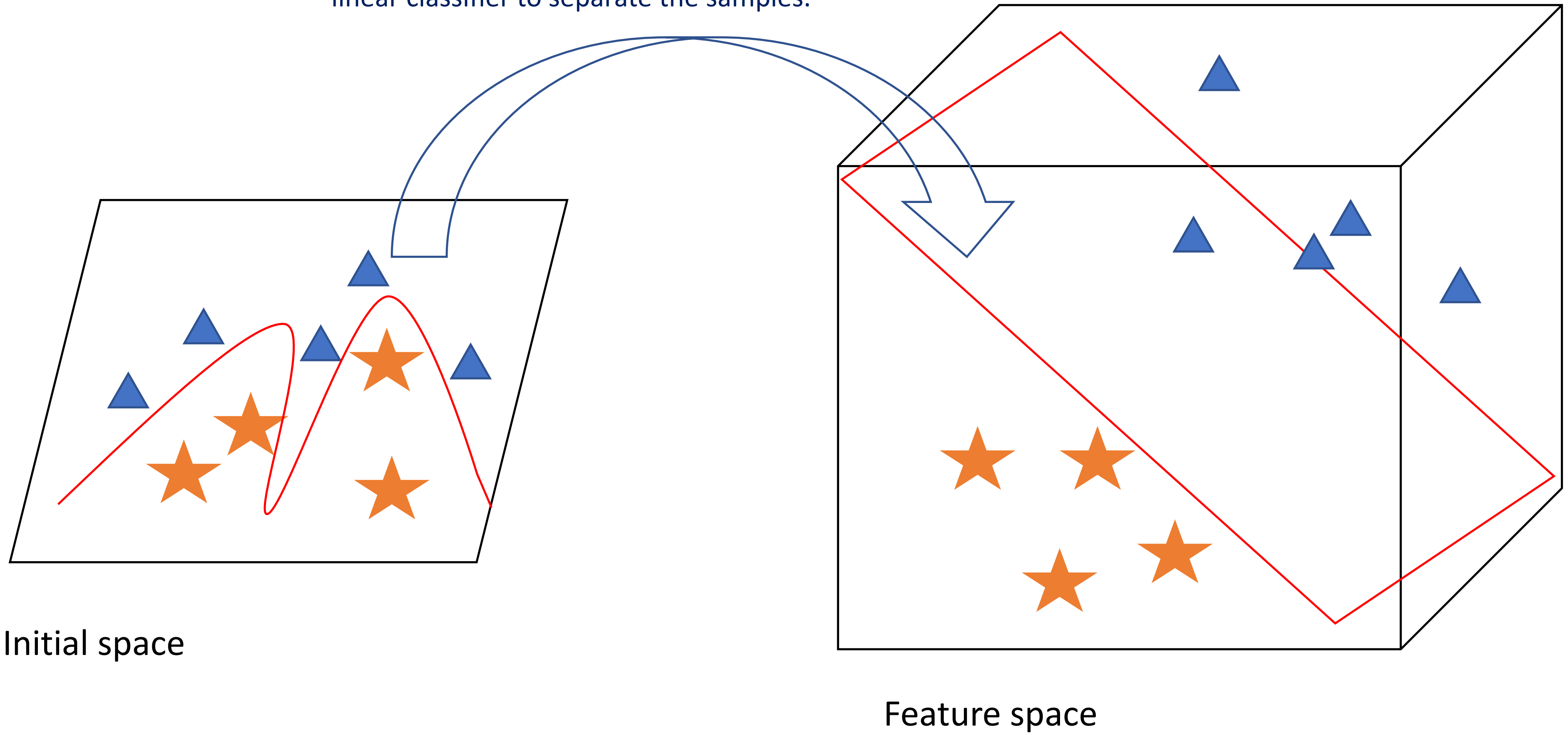
Determination of the largest margin that separates individuals in the feature space. At this stage we use a trick known as the "Kernel Trick" which allows us to calculate the distances (and therefore the margins) in the new space without knowing precisely the transformation from the starting space to the feature space and keeping the small dimension of the initial set and using the observations of the initial set.

Moreover, the procedure does not need all the individuals in the database to determine the separator plane - they only consider those on the margins (they are called support vectors).

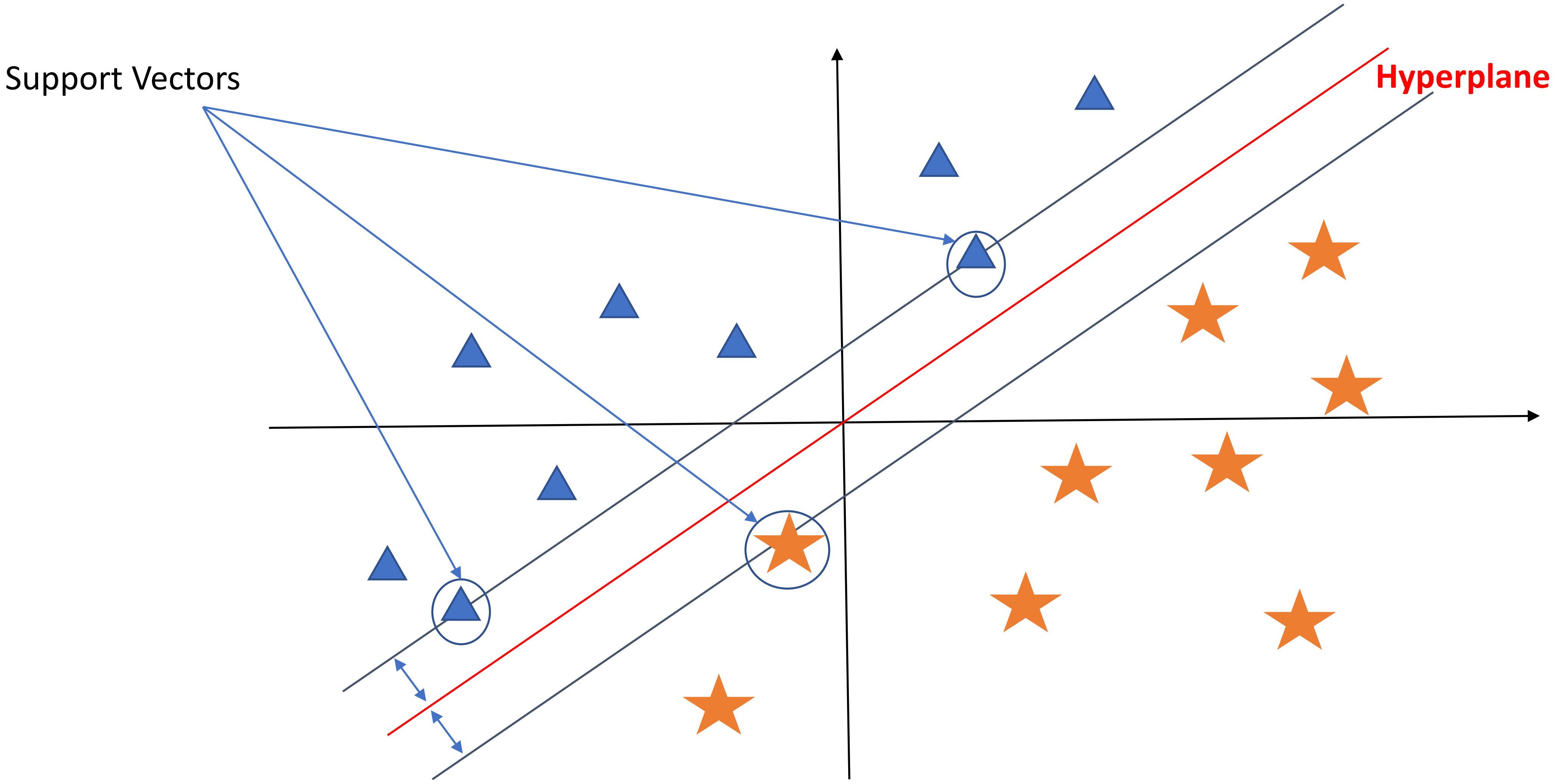
Choice: the user of an SVM has a choice on the kernel function as well as on the optimization program to determine the largest margin

Change of space

Transformation of the plane space into a larger space that allows the use of a linear classifier to separate the samples.



Maximum-margin Hyperplane



Example

A financial advisor wants to be able to give immediate advice to clients on whether or not to obtain credit. To do so, he has a database of individuals who have already applied for credit and the result of their application (0 if credit refused; 1 if credit accepted).

Nom	Genre	nombre de compte	nombre de retraits	Montant sur le compte courant (en euros)	Nombre de découverts	PEA	PEL	Crédit accordé
Antoine	Masculin	1	20	3 000	0	0	1	1
Paul	Masculin	2	35	25 000	0	1	1	1
Sophie	Féminin	1	12	250	7	0	0	0
Ophélie	Féminin	1	62	300	12	0	0	0
Marc	Masculin	2	50	500	13	0	0	0
Pierre	Masculin	3	42	260	8	0	0	0
Virginie	Féminin	1	12	5 000	0	0	1	1
Edouard	Masculin	1	30	1 000	3	0	1	0
Anthony	Masculin	2	15	4 500	0	0	1	1
Gaspard	Masculin	1	17	17 000	0	1	1	1
Isabelle	Féminin	1	15	12 000	0	1	1	1
Corinne	Féminin	3	16	5 000	1	1	0	0
Nadine	Féminin	1	18	4 000	0	0	1	1
Anne	Féminin	4	13	17 000	0	1	1	1
Xavier	Masculin	1	15	10 000	0	1	0	1
François	Masculin	2	12	12 000	0	1	1	1
Marie	Féminin	2	14	1 200	1	1	0	0
Louis	Masculin	3	18	7 800	0	1	1	1
Jules	Masculin	1	19	5 000	1	1	1	1
Delphine	Féminin	3	32	1 300	2	0	0	1

SVM Model

The software determines the optimal separation margins between individuals who have had credit accepted and those who have been refused credit in the new representation space. We have the following confusion matrix:

de \ Vers	0	1	Total	% correct
0	6	1	7	85,71%
1	1	12	13	92,31%
Total	7	13	20	90,00%

In total, the parameters are such that 90% of the observations are classified identically to reality.

Prediction

The financial advisor has another basis on which to test the model already trained.

Nom	Genre	nombre d	nombre de retraits	Montant sur le compte courant (en euros)	Nombre de découverts	PEA	PEL	Crédit accordé
Jordane	Masculin	1	25	2 500	0	0	1	1
Sylvie	Féminin	1	27	500	4	0	1	0
Léa	Féminin	3	28	320	5	0	0	0
Christophe	Masculin	4	16	14 000	0	1	1	1
Pénélope	Féminin	1	17	9 000	0	1	0	1
Alain	Masculin	1	24	3 500	1	1	1	0
Hélène	Féminin	2	15	2 100	1	0	0	0
Thomas	Masculin	1	14	36 000	0	1	1	1
Karine	Féminin	2	19	3 800	0	0	1	1
Ludovic	Masculin	2	22	7 000	0	1	1	1

Results

The model's predictions are compared with reality and two misclassifications are obtained (underlined in red)

	Classe prédite	Crédit accordé
PredObs1	1	1
PredObs2	1	0
PredObs3	0	0
PredObs4	1	1
PredObs5	1	1
PredObs6	1	0
PredObs7	0	0
PredObs8	1	1
PredObs9	1	1
PredObs10	1	1

When we look at observation 6, which corresponds to Alain, he has all the indicators that should normally lead him to accept a loan, and the SVM model, moreover, grants him a loan on this basis. However, in reality he did not obtain the loan for exogenous reasons.
The error on observation 2 (Sylvie) seems to indicate that the possession of a PEL is a determining factor in obtaining a loan.

Artificial Neural Network

Deep Learning: A high-performance model

Since 2006, Artificial Intelligence (AI) has made considerable progress and even experienced a qualitative leap thanks to Deep Learning. Deep Learning is extremely efficient in several areas such as image recognition, natural language and involves important investments in finance, marketing and logistics.

Deep Learning uses artificial neural networks. These networks were introduced in the 1940s to generalize linear regression to non-linear problems.

Algorithms and architectures of neural networks were developed in the 1980's and 1990's and are fully effective today because they benefit from :

- Larger databases that allow statistical generalization
- Much more neural architecture thanks to more powerful computers
- Slight modifications to the algorithms for finding optimal solutions

Principle

neural networks are a set of interconnected neurons. Each neuron processes each of the information it receives from the other neurons by assigning a connection weight to each of them in a linear fashion and then emitting an output signal. Each neuron has an activation function that is based on biological considerations:

For certain information the neuron is inactive. Above a certain threshold the neuron is activated and sends a signal.

Thanks to the activation functions, the neural architecture makes multiple linear regression more complex.

In fact, for each neuron, the combination of activation function and linear assignment of connection weights is called a Rectified Linear unit (ReLu).

Principle

The neurons are distributed over several layers:

- The input layer receives the information (it is said that it is written on the input layer)
- One or more hidden layers process information
- The output layer gives the response of the neural network.

The number of layers gives the **depth** of the network (hence the Deep Learning).

The number of neurons on each layer gives the **width** of the neural network.

The set of connections between neurons determines the **complexity** of the neural network (thanks to the calculation power of computers, the most powerful networks have as many connections as there are synapses in the human brain).

Diffusion Process

How to process information characterizes the neural network. It is the means to determine the best suitable model.

There are 3 forms of information processing:

- Forward propagation (feedforward) when the information is processed in the direction from the input to the output of the model: each layer processes the information sent to it by the previous layer and sends a signal back to the next layer and this process from the input layer to the output layer.
- The Recurrent Neural Network when the information from the input layer to the output layer is also partially communicated between neurons on the same hidden layer.
- The LSTM (Long Short Term Memory) network when a part of the responses is dropped and part of the processed information goes back to hidden layers.

Artificial Neuron

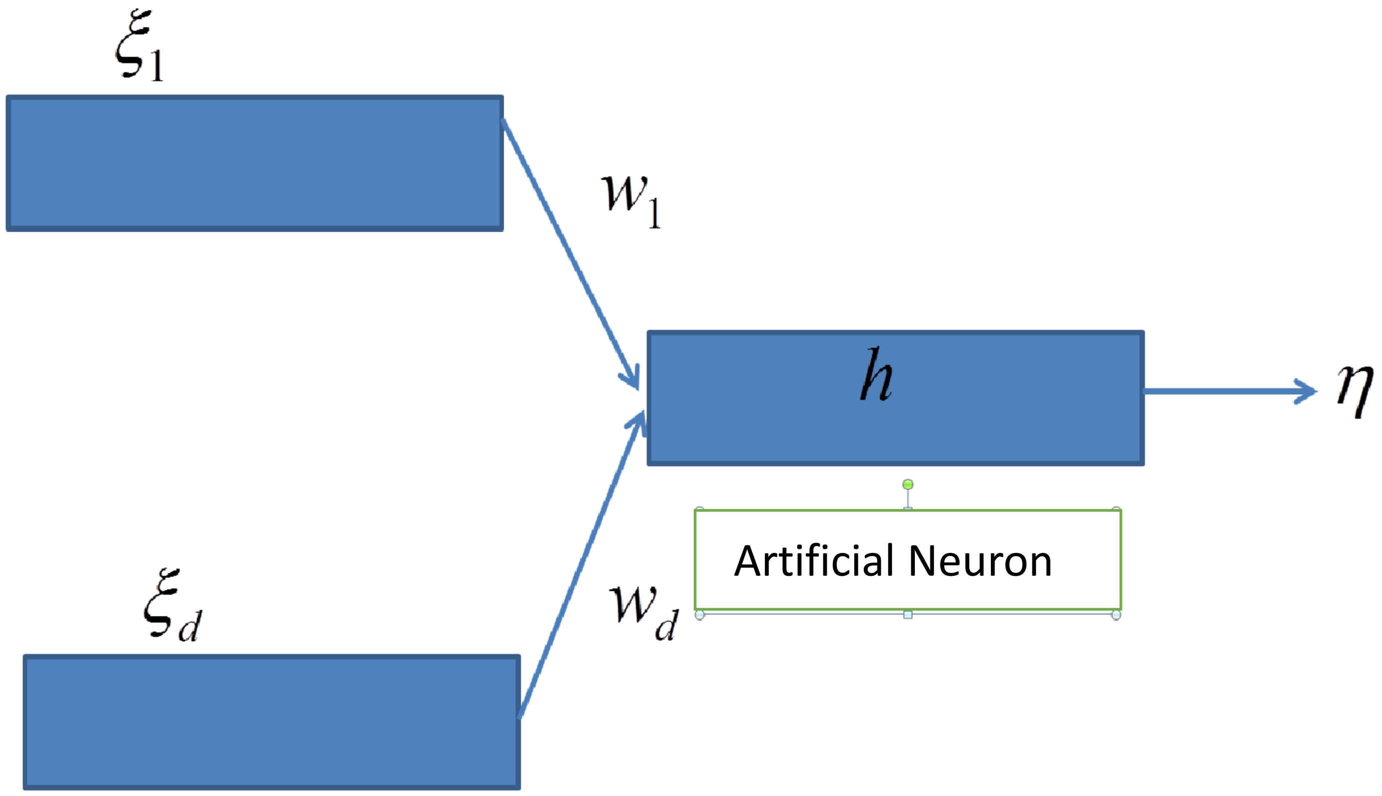
Definition: An artificial neuron is an elementary process connected to other processes that receives a set of signals as input.

$$\xi = \begin{vmatrix} \xi_1 \\ \vdots \\ \xi_d \end{vmatrix}$$

and provides an output decision η (a figure, a purchase decision, an individual's gender, consumer behaviour, suitability for a job, etc.).

Artificial Neuron

The artificial neuron is represented by its h-transfer function.
In input the data are written. Each of the data is weighted by a weight.
Through its transfer function the neuron gives an output signal.



Artificial Neuron

- w_i are the weights of the connections between the neuron in question and the other neurons.

- ξ_i are the signals emitted by each neuron connected to the neuron in question.

- h is a function that characterizes the neuron and is called the neuron transfer function.

Artificial Neuron

$\eta = h(\xi, w)$ neuron's response. It therefore depends on the input signals and connection weights.

$$\xi = \left| \begin{matrix} \xi_1 \\ \vdots \\ \xi_d \end{matrix} \right. \qquad\qquad w = \left| \begin{matrix} w_1 \\ \vdots \\ w_d \end{matrix} \right.$$

Artificial Intelligence consists of optimizing the weights so that the errors of the decisional process are as small as possible.

Sentiment Analysis

Introduction

Stock market prediction methods are divided into two main categories: technical and fundamental analysis.

Technical analysis focuses on analyzing historical stock prices to predict future stock values (i.e. it focuses on the direction of prices).

Fundamental analysis relies mostly on analyzing unstructured textual information like financial news and earning reports.

Technical Analysis

Technical analysis focuses mainly on the stock price movements and on the trading volume.

It can be used on any security with historical trading data. This includes stocks, futures, commodities, fixed-income, currencies, and other securities.

This prediction method can depart from the fundamental value of the risky asset. It relies on the structured data produces by the financial markets.

Technical Analysis

There are three benchmark models :

- Kyle (1985) model
- Black Scholes model
- Time series model

They can predict the future stock prices by analysing the yield curve

Definition and Useful Vocabulary

- A derivative product is a financial security whose value depends on 1 or several UA

$$\pi_t = f\left(t, S_t^0, S_t^1, \ldots, S_t^N\right)$$

3 main product families:

- Forward, Futures,
- Options
- Swaps

Definition and Useful Vocabulary

Where are derivatives are traded ?

- Exchange-traded Markets : Futures and Options
- Liffe, Eurex, CBOT, NYMEX
- Standardized products
- Trading floor or electronic platform (becoming the standard)
- Over-The-Counter (OTC)
- No-standard-product (customized)
  - Phone/dealer
  - Credit Risk

Definition and Useful Vocabulary

The trading of a financial asset involves at least four discrete steps

- A buyer and a seller must locate one another and agree on a price
- The trade must be cleared (the obligations of each party are specified)
- The trade must be settled (the buyer and the seller must deliver the cash or securities to satisfy their obligations in the required period of time)
- Ownership records are updated

Definition and Useful Vocabulary

There are at least four different measures of a market and its activity

- **Trading volume:** this measures the number of financial claims that change hands
- **Market value:** market value is the sum of the value of the claims that could be traded without regard to whether they have traded
- **Notional value:** this measures the scale of a position
- **Open interest:** it measures the total number of contracts for which counterparties have a future obligation to perform

Definition and Useful Vocabulary

Transaction costs and the bid ask spread

- **Buying and selling a financial asset**
- **Brokers: commission (fees)**
- **Market Makers (bid-ask) : the order processing cost; the inventory holding cost; the adverse selection cost.**

- **Short-selling when the price of an asset is expected to fall**
  - **First: borrow and sell an asset (get $$)**
  - **Then: buy back and return the asset (pay $)**

→ **If price fell in the mean-time: profit=$$-$**

Uses

- **Hedging (Risk Management)**
- **Speculation: Directional views on the future of the market**
- **To cash in an arbitrage profit**
- **To change the nature of the liability (fixed rate-floating rate, rate-share, etc.)**
- **To reduce the transaction costs**
- **To bypass regulations**
- **To create structured products (Financial Engineering)**

Uses

- **Risk-management : irrationality :**

- **Risk aversion**
- **Feedback trading**
- **Herding**
- **Overconfidence**

Pricing Principles

Hypotheses:

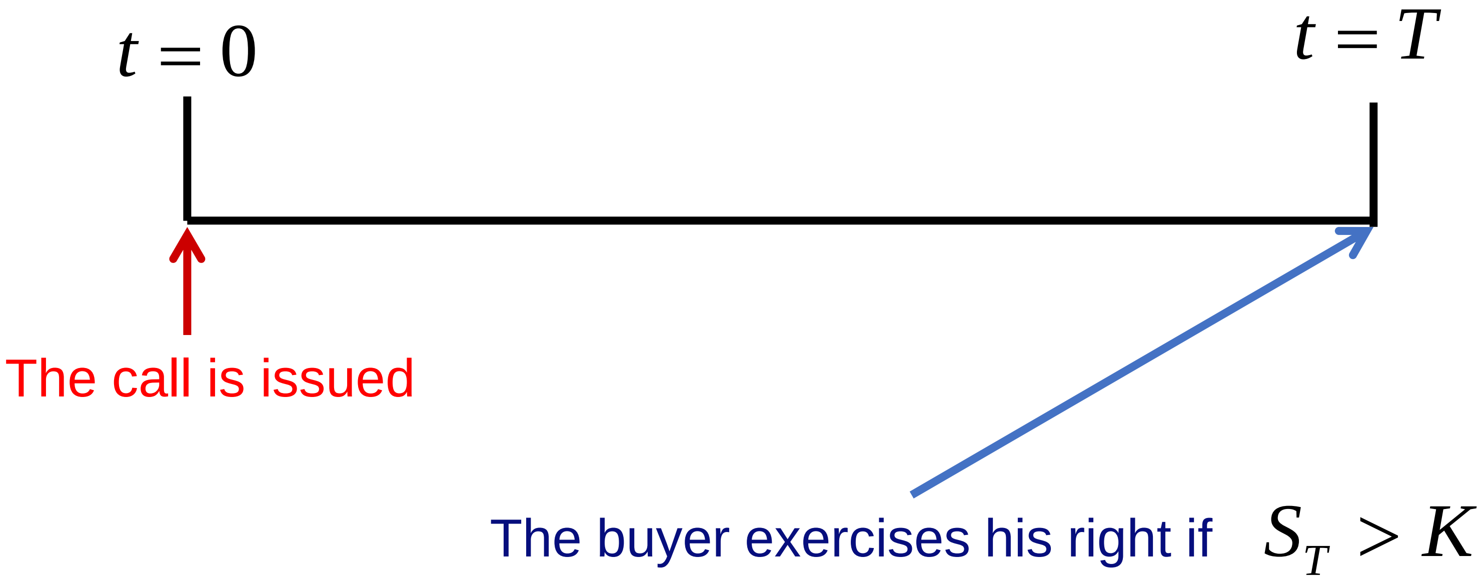
- **H1:** Frictionless Market (no transaction costs, no bid-ask spread, no margin calls, no restrictions on short-selling, no taxes)
- **H2:** no counterparty risk
- **H3:** Competitive markets (traders are price takers)
- **H4:** AAO

Call option (1/5)

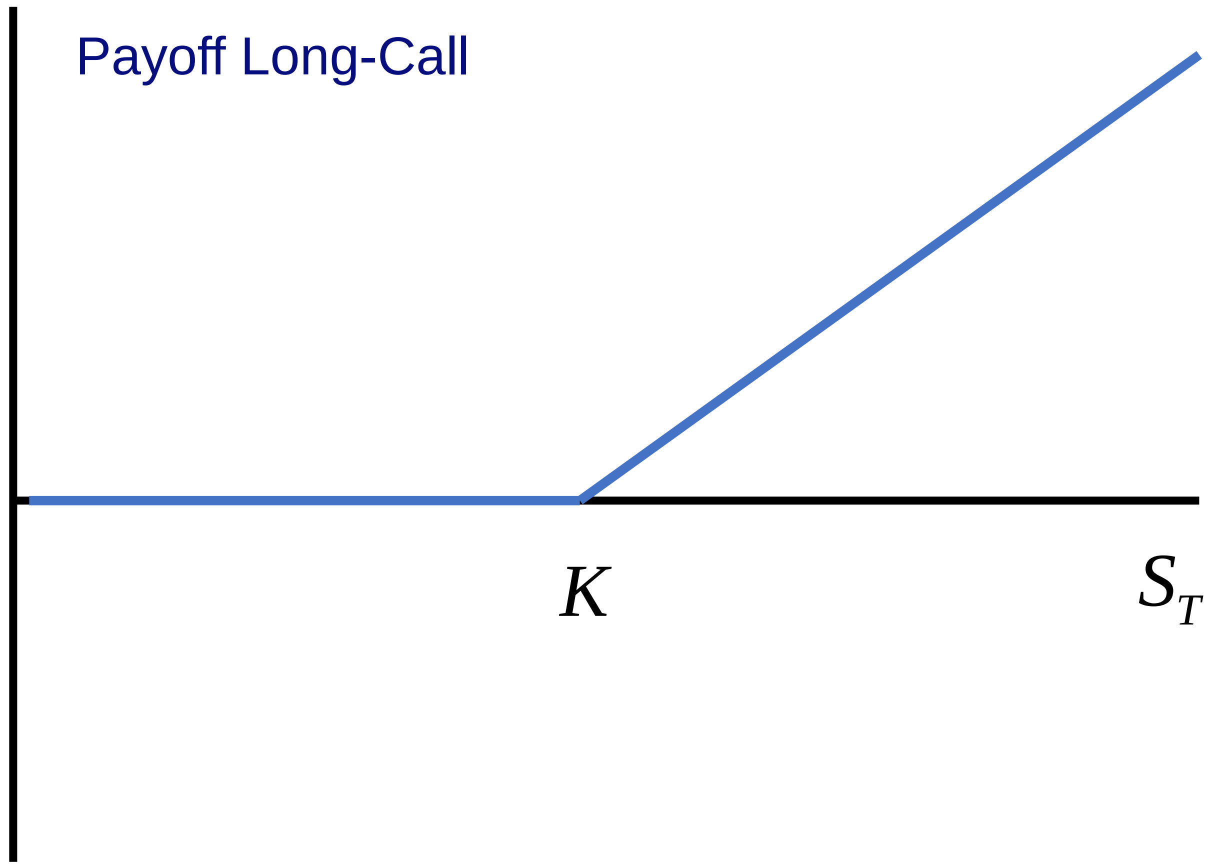
Definition: a Call is the right but not the obligation to buy the UA at a specified future date T (expiry or maturity) and at a predetermined price K (strike price)

- The seller has the corresponding obligation to fulfill the transaction—that is to sell —if the buyer (owner) "exercises" the option.
- The buyer pays a premium to the seller for this right

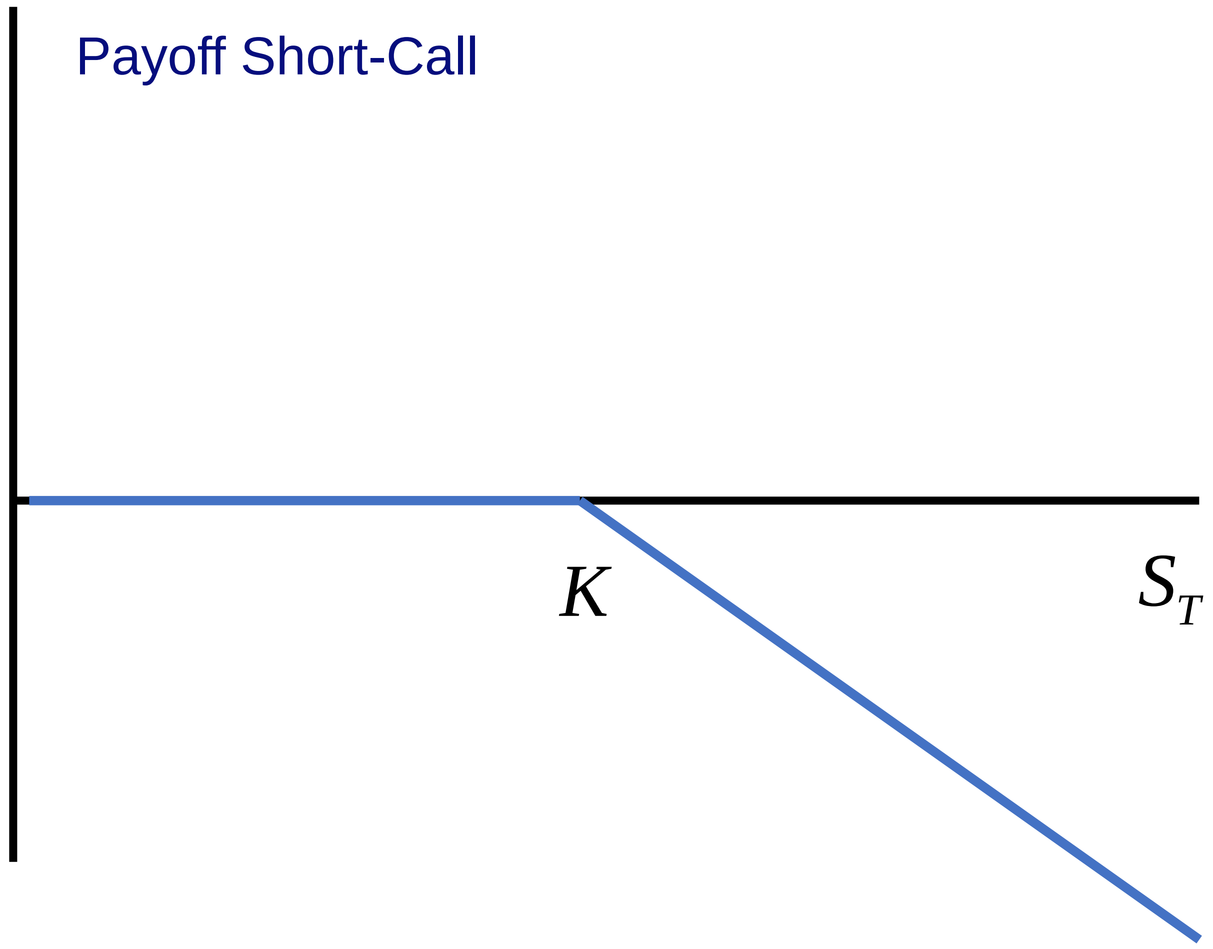
Call option (2/5)



Call option (3/5)



Call option (4/5)



Call option (5/5)

Exercise 1: At time $t = 0$

Mr Douglas owns 100$ he buys an underlying asset $S_0 = 100\$$

Sir Heath chooses to buy 10 calls, the value of each call is equal to $C = 10\$$; strike price K=$100

a) At time $t = T$, $S_T = 150\$$ calculate the different gains

b) At time $t = T$, $S_T = 99\$$ calculate the different gains

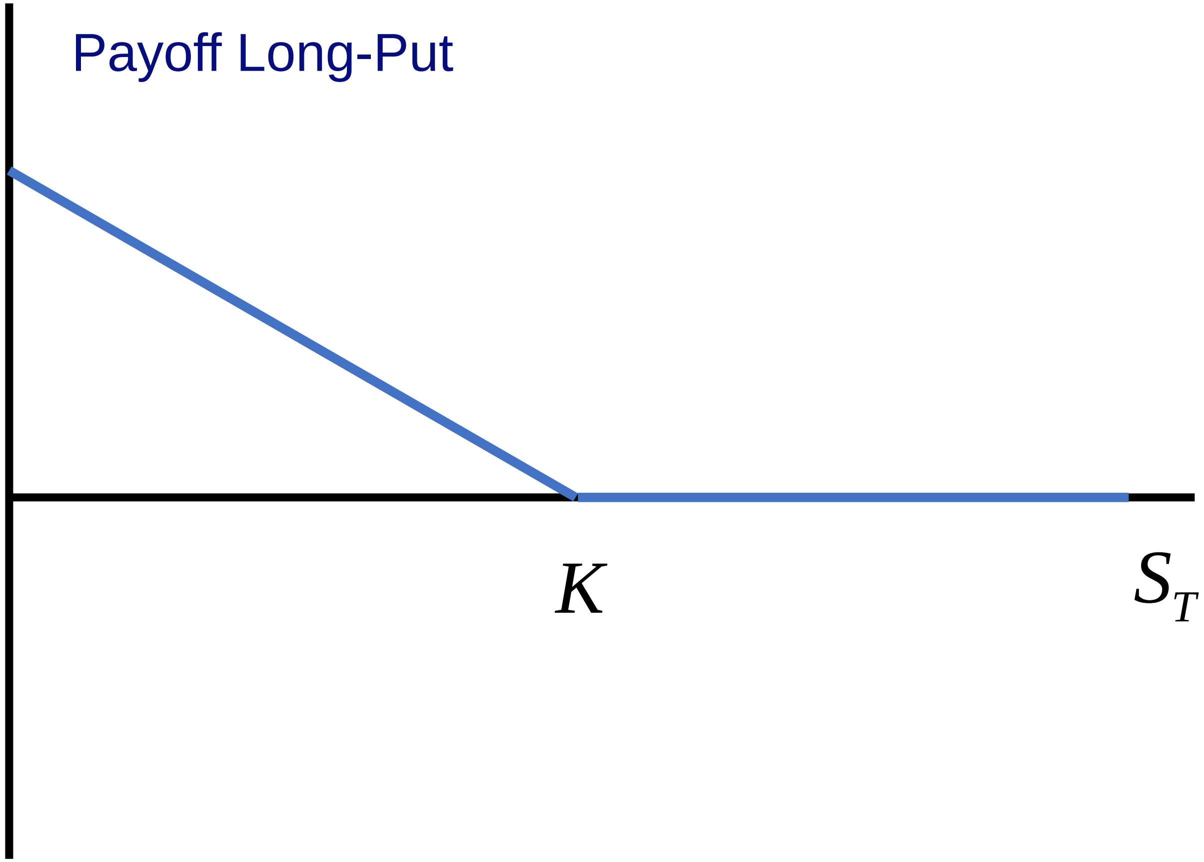
c) Make concluding remarks

Put option (1/3)

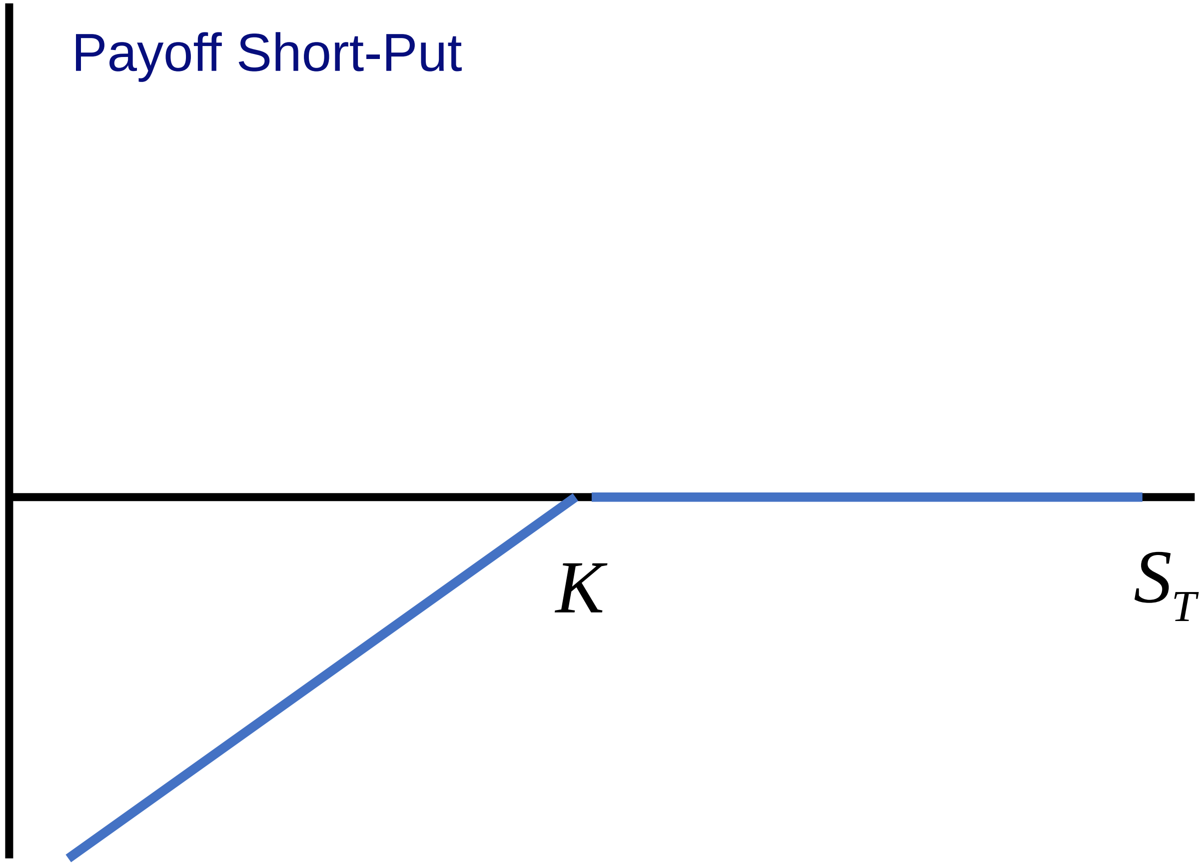
Definition: a Put is the right but not the obligation to sell the UA at a specified future date T (expiry or maturity) and at a predetermined price K (strike price)

- The seller has the corresponding obligation to fulfill the transaction—that is to buy —if the buyer (owner) "exercises" the option.
- The buyer pays a premium to the seller for this right

Put option (2/3)



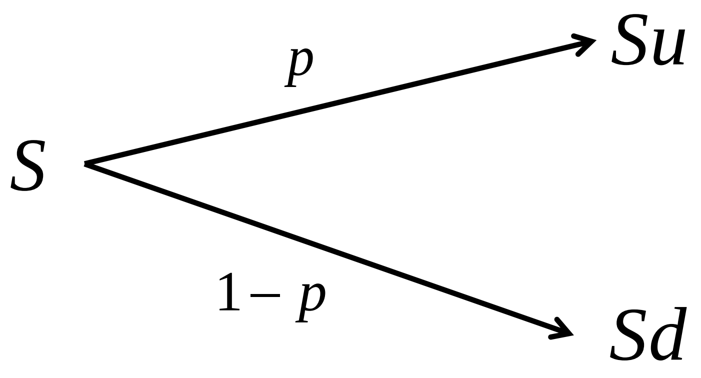
Put option (3/3)



CRR (1/10)

Hypotheses: We consider a financial market, with only one auction and where two assets are traded

- A risk-free asset with a return r
- A risky asset yields a return $\tilde{R} = \dfrac{P_1}{P_0}$



CRR (2/10)

First property: show that the AAO implies that :

$$d < r < u$$

- Deduce the existence of risk-neutral probability measure such that :

$$E_{P^*}\left[\tilde{R}\right] = r \quad \text{with} \qquad P^*\left(\tilde{R} = u\right) = q$$

$$\begin{tikzcd} & Su \\ S \arrow[ur, "q"] \arrow[dr, "1-q"'] & \\ & Sd \end{tikzcd}$$

CRR (3/10)

By using the replicating portfolio find the value of the call with strike K and the maturity T=1.

At time t=0, the seller receives the premium C and buys Δ risky asset

At time t=1, we have:

$$\begin{cases} (C-\Delta S)r + \Delta S u = Su - K \\ (C-\Delta S)r + \Delta S d = 0 \end{cases}$$

CRR (4/10)

Proposition: The absence of arbitrage opportunities implies that there exists a risk- neutral probability measure.

Proposition: The price of a derivative is the present value of the expectation of the payoff of that derivative under the risk-neutral probability measure

CRR (5/10)

Definition: A complete market is a market in which the payoff of any derivative could be replicated by using the existing underlying assets.

Proposition: In a complete market with the absence of arbitrage opportunities there exists a unique risk-neutral probability measure.

CRR (6/10)

Exercise 2: We consider a static market model in which two assets are traded. A riskless asset guarantees a return $r = 1{,}05$, the strike K=$100

At the commencement of the trading game the value of the risky asset is $S_0 = 100\$$

At the end of the trading day the value of the risky asset is either $S_1 = 90\$$ or $S_1 = 110\$$

1) Calculate the return of the risky asset

CRR (7/10)

The historical probability measure gives

$$P\left(\tilde{R} = u\right) = P\left(\tilde{R} = d\right) = \frac{1}{2}$$

2) Calculate the expectation of the risky asset return.

3) Determine the risk-neutral probability measure.

CRR (8/10)

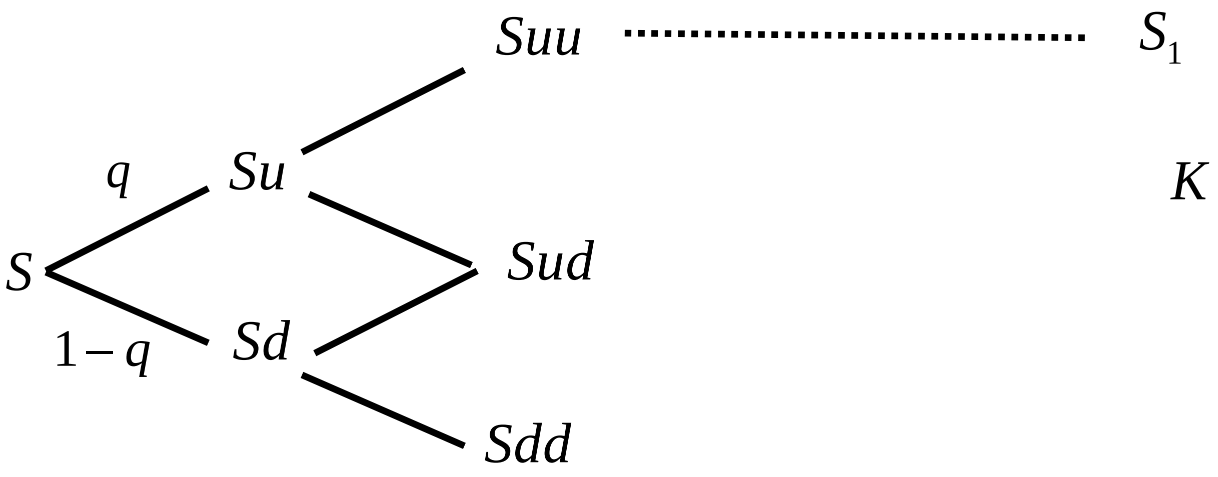
We consider at the opening of the market a call option on the risky asset with a maturity date $T = 1$ and a strike price $K = 100$

4) Determine the hedging strategy.

5) Determine the premium of that call.

CRR (9/10)

We consider a sequential-auction market with T periods



CRR (10/10)

The probability to have done *n* climbs:

$$\binom{T}{n} q^n (1-q)^{T-n}$$

$$n^* = \inf\{n \geq 1, S_1 \geq K\}$$

And

The premium

$$C = S\left\{\sum_{n=n^*}^{T} \frac{\binom{T}{n}(qu)^n((1-q)d)^{T-n}}{r^T}\right\} - \frac{K}{r^T}\left\{\sum_{n=n^*}^{T}\binom{T}{n} q^n (1-q)^{T-n}\right\}$$

Stochastic Processes

- In deterministic processes the future value of state variables is known. In other words, the state variable is only a function of time (at each future date it is only likely to take on a unique value). $S_t = S(t)$
- The value of the variable S is only time-dependent.
- Constant Value

$$\left.\begin{array}{l} dS = 0 \\ S(0) = S_0 \end{array}\right\} \Leftrightarrow S_t = S_0 \ \ \forall t$$

Stochastic Processes

- Variable S increases continuously over each period.
- ∀

$$\left.\begin{aligned} dS &= \mu dt \\ S(0) &= S_0 \end{aligned}\right\} \Leftrightarrow S_t = S_0 + \mu t,\ \forall t$$

- This evolution corresponds to a declining growth rate of S as S rises. If S represents the evolution of a capital with capitalised interest, the annual interest rate is declining and tends towards zero.

$$S_{t+1} - S_t = \mu\ \forall t$$

- The model therefore does not represent the evolution of a capital invested at constant interest.

Stochastic Processes

- Constant growth rate (or interest rate)

$$\left.\begin{aligned} & dS = \mu S dt \;\; ou \;\; \frac{dS/dt}{S} = \mu \\ & S(0) = S_0 \end{aligned}\right\} \Leftrightarrow S_t = S_0 e^{\mu t} \;\; ou \;\; \ln(S_t) = \ln(S_0) + \mu t, \; \forall t$$

- μ represents the continuous annual growth rate or the continuous annual interest rate. If, for example, b(t) represents the evolution of a capital invested at the continuous annual risk-free interest rate r, assumed constant, we would have: db = r.b.dt
- One could imagine a more general deterministic process of the following kind
- dS = μ (t)Sdt where the annual interest rate would evolve in a deterministic way over time

Brownian Motion

- **Brownian Motion**: it has been used in physics to describe the motion of a particle that is subject to a large number of small molecular shocks.

Properties:

- 1) the change dB_t during a small period of time dt is

$dB_t \sim \sqrt{dt}\,N(0,1)$ where $N(0,1)$ is a standardized normal distribution

- The changes for any two different short intervals of time are independent.

Black-Scholes Hypothesis

- The dynamic of the underlying asset is given by

$$dS_t = \underbrace{rS_t dt}_{\text{The trend}} + \underbrace{\sigma S_t dB_t}_{\text{The random walk}}$$

Sentiment Analysis

Definition:

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space.

Remark: There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc.

Sentiment Analysis

A new context

These days more and more critical information about the stock market has become available on the Web. Examples include BBC, Bloomberg, and Yahoo Finance. It is hard to manually extract useful information out of these resources. This draws a picture of the significance of text mining techniques to automatically extract meaningful information for analyzing the stock market.

Sentiment analysis

Sentiment analysis uses text mining, natural language processing, and computational techniques to automatically extract sentiments from a text. It aims to classify the polarity of a given text at the sentence level or class level, whether it reflects a positive, negative, or neutral view. In stock market prediction task, two important sources of the text are used either social media mainly using Twitter data or online financial news article.

Sources

Twitter sentiment

Twitter is a significant source of data, and many researchers have examined its relationship with stock market movements. While each tweet is restricted to 140 characters, it is believed that the information can accurately reflect public mood.

Online financial news sentiment

Financial news articles are perceived to be a more consistent and reliable source of information. Many researchers suggested that the financial news articles have a strong relationship with stock market fluctuation; therefore, analyzing financial news reports can help in predicting the stock market movements.

Textual Data Preprocessing

Textual data need to be prepared before used by the machine learning algorithm for sentiment analysis task using these methods.

Feature extraction

Feature extraction or sometimes called attribute selection aim to select features, attributes, or piece of text that is more relevant to the prediction task. Many methods have been used for feature selection. The commonly used feature selection procedure for document or sentence classification task is the bag-of-words (BOW) approach.

In the mentioned model, each word in a text or document will be treated as a feature neglecting the grammar or word order and only preserving the abundance.

The second most popular method used recently for the feature selection process is Word2vec. In this technique, the aim is to learn word embedding using a two layer neural network. The input to that neural network is a text, and the output is a group of vectors (i.e. the input is a corpus and the output is a vector of words).

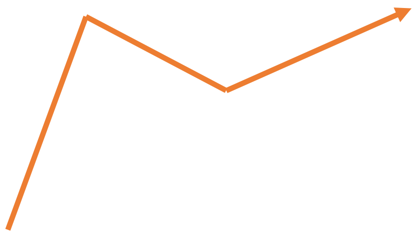
ML and DL for Sentiment Analysis

Textual data

- Social Media
- Financial News
- Blog Forum

Market values

Historical stock prices



Preprocessing

- Feature Selections
- Feature Representations

Classification Model

- ML
- DL

Market Prediction