

# BIG DATA AND CO

A glimpse on possibilities in Accounting

# WHO AM I?

Assistant Professor of Accounting at Rady

- PhD in Finance - Dauphine University, France
- Master in Finance - Dauphine University, France
- Master in AI - Toulouse University, France
- Master in CS Engineering - INSA, France

# RESEARCH

## BEFORE

- AI - Reinforcement Learning

## RECENTLY

- Banking Regulation
- Impact of data on financing decisions

## NOW

- Real effects implications of accounting rules
- Public disclosures and Information Processing

# INTERNATIONAL EXPERIENCE

Hong Kong University

Harvard

Imperial College Business School

CONTACT

[mmilone@ucsd.edu](mailto:mmilone@ucsd.edu)

And when we will be back on campus:

858-534-8511

Office 3W102

# FINANCIAL ACCOUNTING

# FINANCIAL ACCOUNTING

Aggregation of information into financial statements

# FINANCIAL ACCOUNTING

Aggregation of information into financial statements

Provides information about past transactions



# MANAGERIAL ACCOUNTING

# MANAGERIAL ACCOUNTING

Produces information for management decisions

# MANAGERIAL ACCOUNTING

Produces information for management decisions

Focuses on the future

# MANAGERIAL ACCOUNTING

Produces information for management decisions

Needs a lot of Data

Focuses on the future

# MANAGERIAL ACCOUNTING

Produces information for management decisions

Needs a lot of Data

Focuses on the future

Need for prediction

# ON THE NEED FOR DATA

# ON THE NEED FOR DATA

Cost estimations

# ON THE NEED FOR DATA

Cost estimations

Cost allocations



# ON THE NEED FOR DATA

Cost estimations

Cost allocations

Breakeven Point

# ON THE NEED FOR DATA

Cost estimations

Cost allocations

Breakeven Point

Sensitivity Analysis

# ON THE NEED FOR DATA

Cost estimations

Cost allocations

Breakeven Point

Sensitivity Analysis

Transfer Pricing

# ON THE NEED FOR PREDICTION

Forecast of Cash Flows

Risk Management

...

Cost Analysis

Fraud detection

...

# ON THE NEED FOR PREDICTION

ABOUT THE FUTURE

Forecast of Cash Flows

Risk Management

...

Cost Analysis

Fraud detection

...

# ON THE NEED FOR PREDICTION

ABOUT THE FUTURE

Forecast of Cash Flows

Risk Management

...

ABOUT THINGS WE DO NOT KNOW

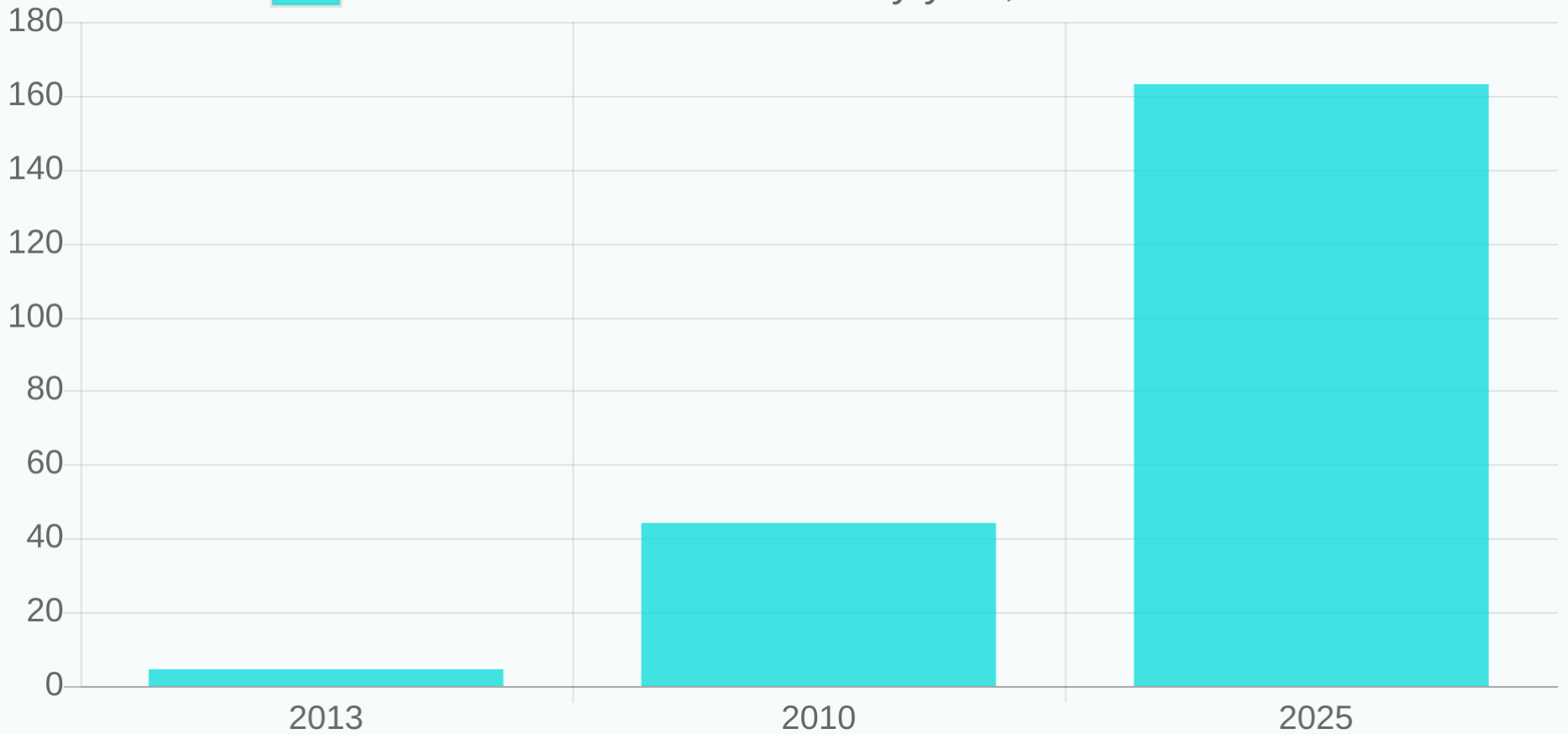
Cost Analysis

Fraud detection

...

DATA IS COMING ...

Amount of data created every year, in ZETABYTES



1 ZB = 1 Billion Terabytes



# BIG DATA

ARE COMPANIES THERE YET?

# WHAT IS BIG DATA?

# WHAT IS BIG DATA?

Just a Big Buzzword?

# WHAT IS BIG DATA?

Just a Big Buzzword?

Just a lot more data?

# WHAT IS BIG DATA?

Just a Big Buzzword?

Just a lot more data?

The combination of previously separate datasets?

# WHAT IS BIG DATA?

Just a Big Buzzword?

Just a lot more data?

The combination of previously separate datasets?

Structured vs unstructured data?

# WHAT ARE COMPANIES DOING TODAY?

Microsoft power BI and co

Slowly Starting

# WHAT ARE COMPANIES DOING TODAY?

BUSINESS ANALYTICS

Microsoft power BI and co

Slowly Starting



# WHAT ARE COMPANIES DOING TODAY?

BUSINESS ANALYTICS

Microsoft power BI and co

PREDICTIVE ANALYTICS

Slowly Starting

# RPA

= ROBOT PROCESSING AUTOMATION

# RPA

= ROBOT PROCESSING AUTOMATION

Automation of routine rules-based business  
processes

# RPA

= ROBOT PROCESSING AUTOMATION

Automation of routine rules-based business  
processes

Embedded into Enterprise Resource Planning (ERP)  
softwares

# RPA

= ROBOT PROCESSING AUTOMATION

Automation of routine rules-based business  
processes

Embedded into Enterprise Resource Planning (ERP)  
softwares

UiPath, Blue Prism, Automation Anywhere

# RPA

## EXAMPLES

Invoice processing

Data validation

Customer Relationship

HR processes (payroll, ...)

# RPA

## EXAMPLES

Invoice processing

Data validation

Customer Relationship

HR processes (payroll, ...)

# RPA

## WHY DOES IT MATTER?



# RPA

## WHY DOES IT MATTER?

### Digitization of processes

# RPA

WHY DOES IT MATTER?

Digitization of processes

Enhances the creation and gathering of data

# RPA

WHY DOES IT MATTER?

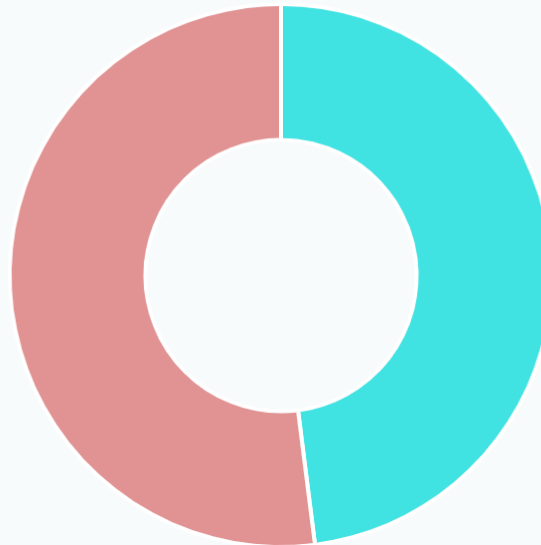
Digitization of processes

Enhances the creation and gathering of data

First step towards *Intelligent Automation*

# RPA

■ Have started RPA ■ Have NOT started RPA

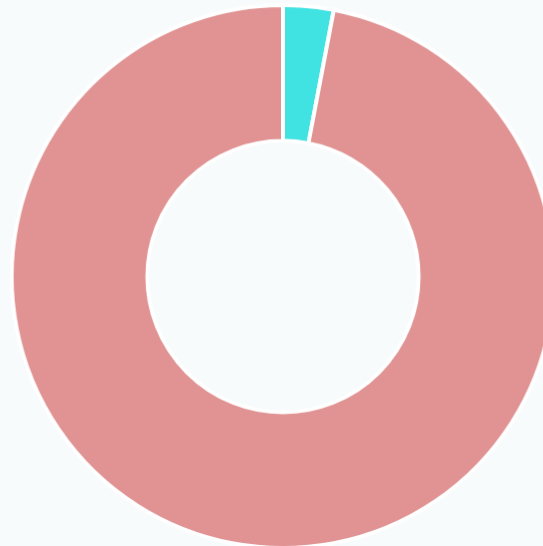


< 50% of companies have started RPA

Data from EY, 2018

# RPA

■ Have implemented RPA at scale ■ Have NOT



3% of enterprises have achieved scale

Data from EY, 2018

# BEYOND RPA

## FROM DATA TO DECISION MAKING

Data is not enough. We also need

- A way to make the data useful
- A way to convey the information learned

# BEYOND RPA

## FROM DATA TO DECISION MAKING

Data is not enough. We also need

- A way to make the data useful
- Prediction
- A way to convey the information learned

# BEYOND RPA

## FROM DATA TO DECISION MAKING

Data is not enough. We also need

- A way to make the data useful
- Prediction
- A way to convey the information learned
- Visualization



# MACHINE LEARNING

## AND HOW IT AFFECTS ACCOUNTING

# HOW DO WE SOLVE PROBLEMS?

PREDICTION PROBLEM?

DECISION PROBLEM?

# HOW DO WE SOLVE PROBLEMS?

PREDICTION PROBLEM?

Need to find a relationship between  $X$  and  $Y$

DECISION PROBLEM?

# HOW DO WE SOLVE PROBLEMS?

PREDICTION PROBLEM?

Need to find a relationship between X and Y

DECISION PROBLEM?

Need to find an *Algorithm*

# HOW DO WE SOLVE PROBLEMS?

PREDICTION PROBLEM?

Need to find a relationship between X and Y

DECISION PROBLEM?

Need to find an *Algorithm*

= Steps to solve a problem

# HOW ABOUT ML?

Supervised / Unsupervised Learning

Reinforcement Learning

# HOW ABOUT ML?

We want the machine to find the relationship or the algorithm by itself

Supervised / Unsupervised Learning

Reinforcement Learning

# PREDICTION AND VISUALIZATION

How to gain insight using big data analytics with Power BI





# HOW ABOUT ML?

We want the machine to find the relationship or  
the algorithm by itself

HOW DO WE DO THAT?

Supervised / Unsupervised Learning

Reinforcement Learning

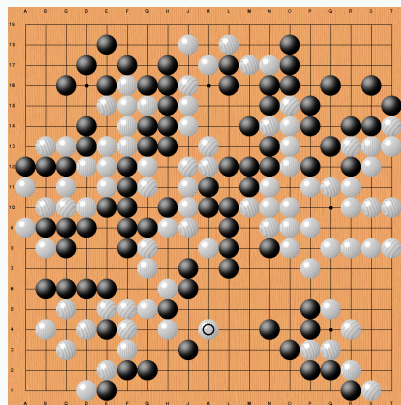
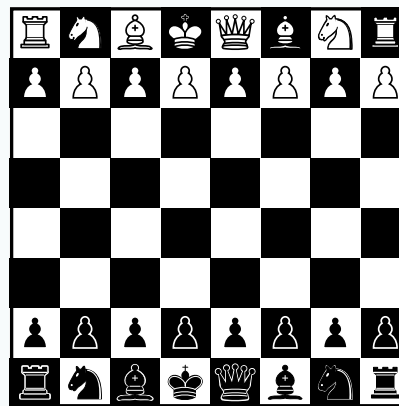
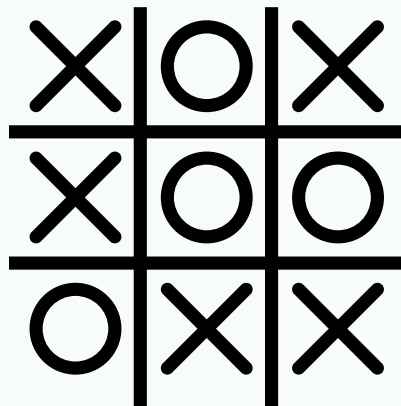
# HOW ABOUT ML?

We want the machine to find the relationship or the algorithm by itself

HOW DO WE DO THAT?

Supervised / Unsupervised Learning

Reinforcement Learning



5	0	4	1	9	2	1	3	1	4
3	5	3	6	1	7	2	8	6	9
4	0	9	1	1	2	4	3	2	7
3	8	6	9	0	5	6	0	7	6
1	8	7	9	3	9	8	5	9	3
3	0	7	4	9	8	0	9	4	1
4	4	6	0	4	5	6	1	0	0
1	7	1	6	3	0	2	1	1	7
9	0	2	6	7	8	3	9	0	4
6	7	4	6	8	0	7	8	3	1

# HOW DOES THE MACHINE LEARN?

Input: board, image

Output: move, number

Highly non-linear math **function**

# HOW DOES THE MACHINE LEARN?

NEED TO MAP AN INPUT TO AN OUTPUT

Input: board, image

Output: move, number

Highly non-linear math function

# HOW DOES THE MACHINE LEARN?

NEED TO MAP AN INPUT TO AN OUTPUT

Input: board, image

Output: move, number

HOW DO WE MAP THESE?!

Highly non-linear math function

HIGHLY WHAT?!

# HIGHLY WHAT?!

## POSSIBLE METHODS



# HIGHLY WHAT?!

## POSSIBLE METHODS

### Regression / Classification Trees

# HIGHLY WHAT?!

## POSSIBLE METHODS

Regression / Classification Trees

Random Forests

# HIGHLY WHAT?!

## POSSIBLE METHODS

Regression / Classification Trees

Random Forests

K-nearest neighbor

# HIGHLY WHAT?!

## POSSIBLE METHODS

Regression / Classification Trees

Random Forests

K-nearest neighbor

Neural Networks

# HIGHLY WHAT?!

## POSSIBLE METHODS

Regression / Classification Trees

Random Forests

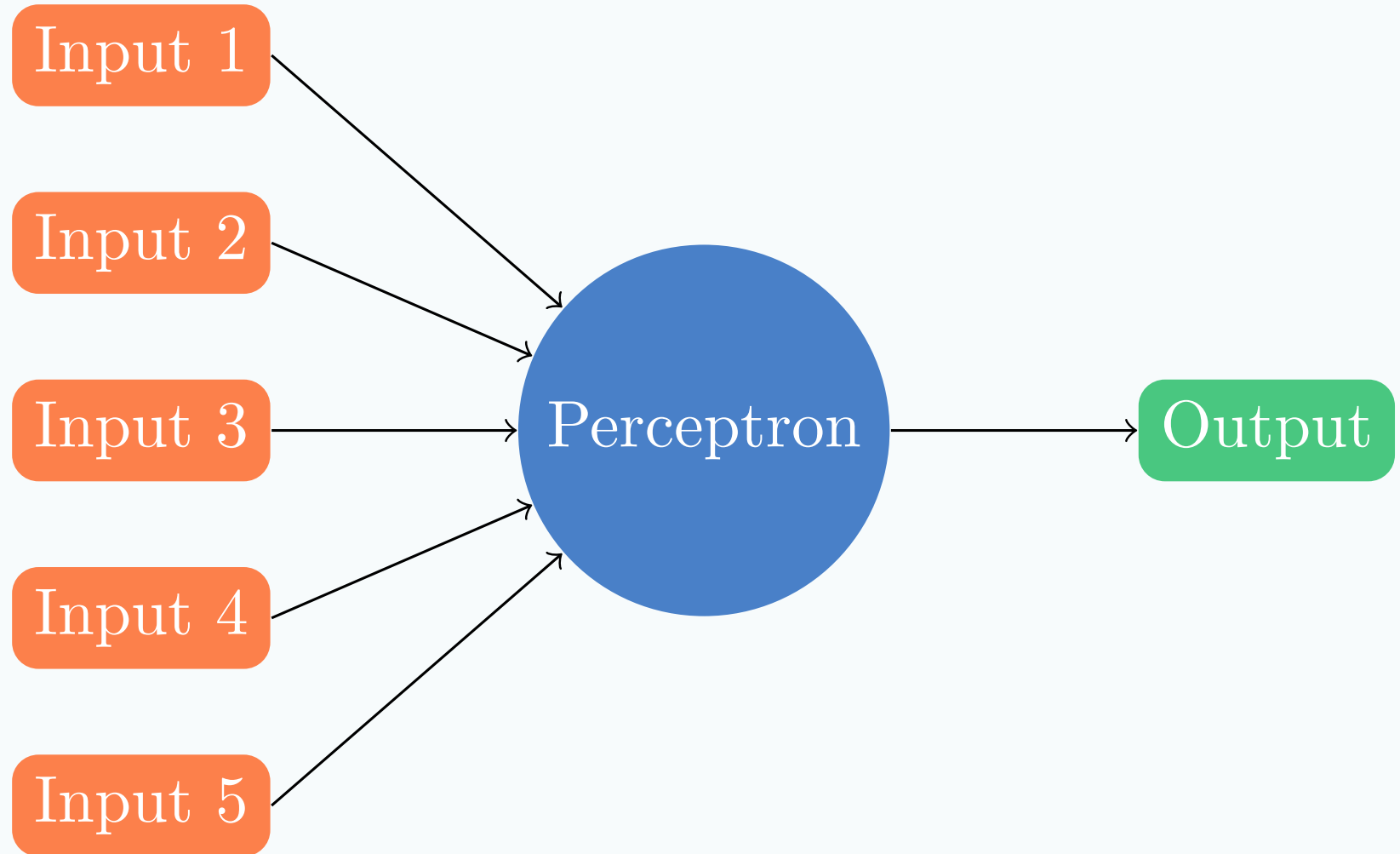
K-nearest neighbor

Neural Networks

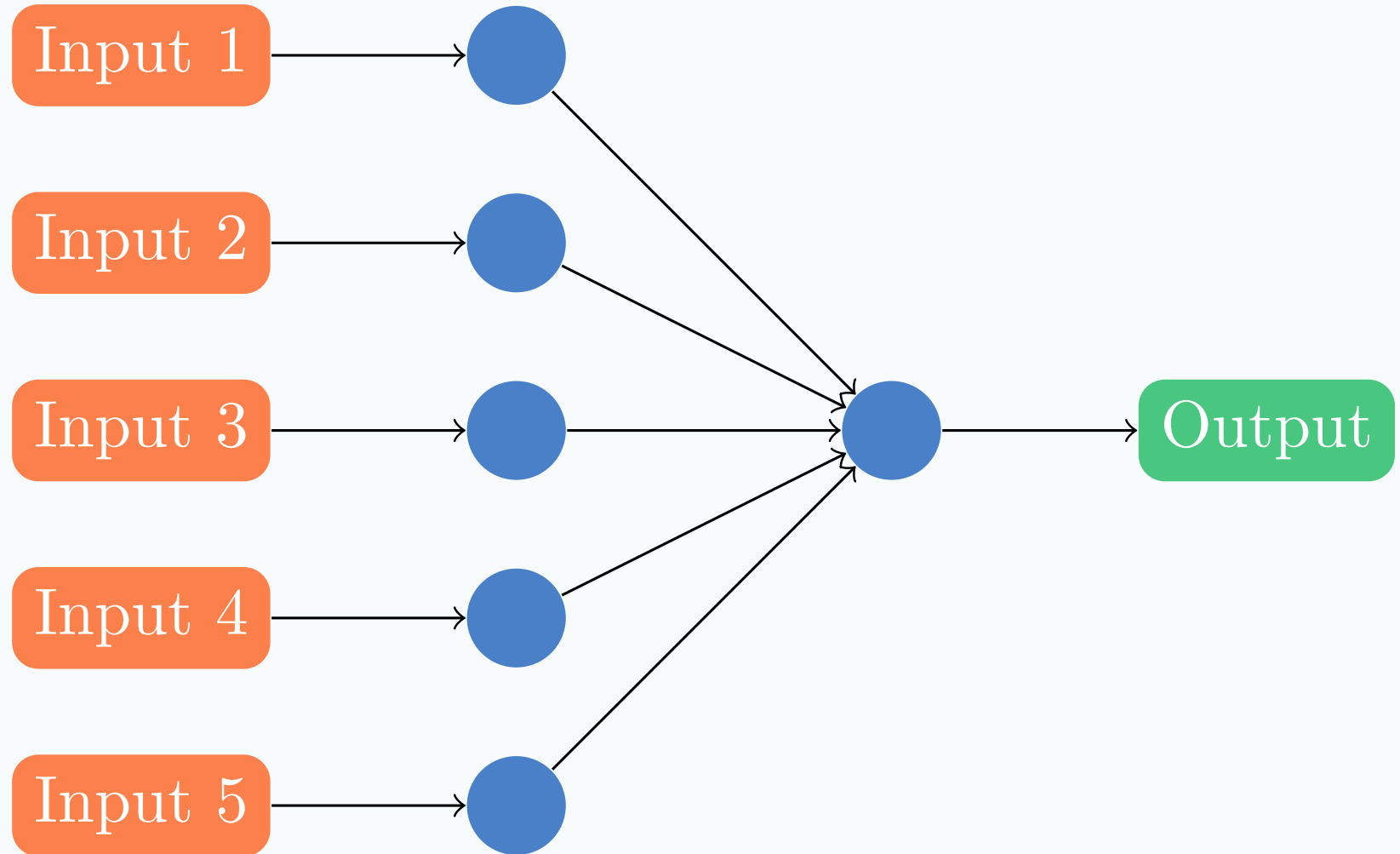
...

# NEURAL NETWORKS

# NEURAL NETWORKS

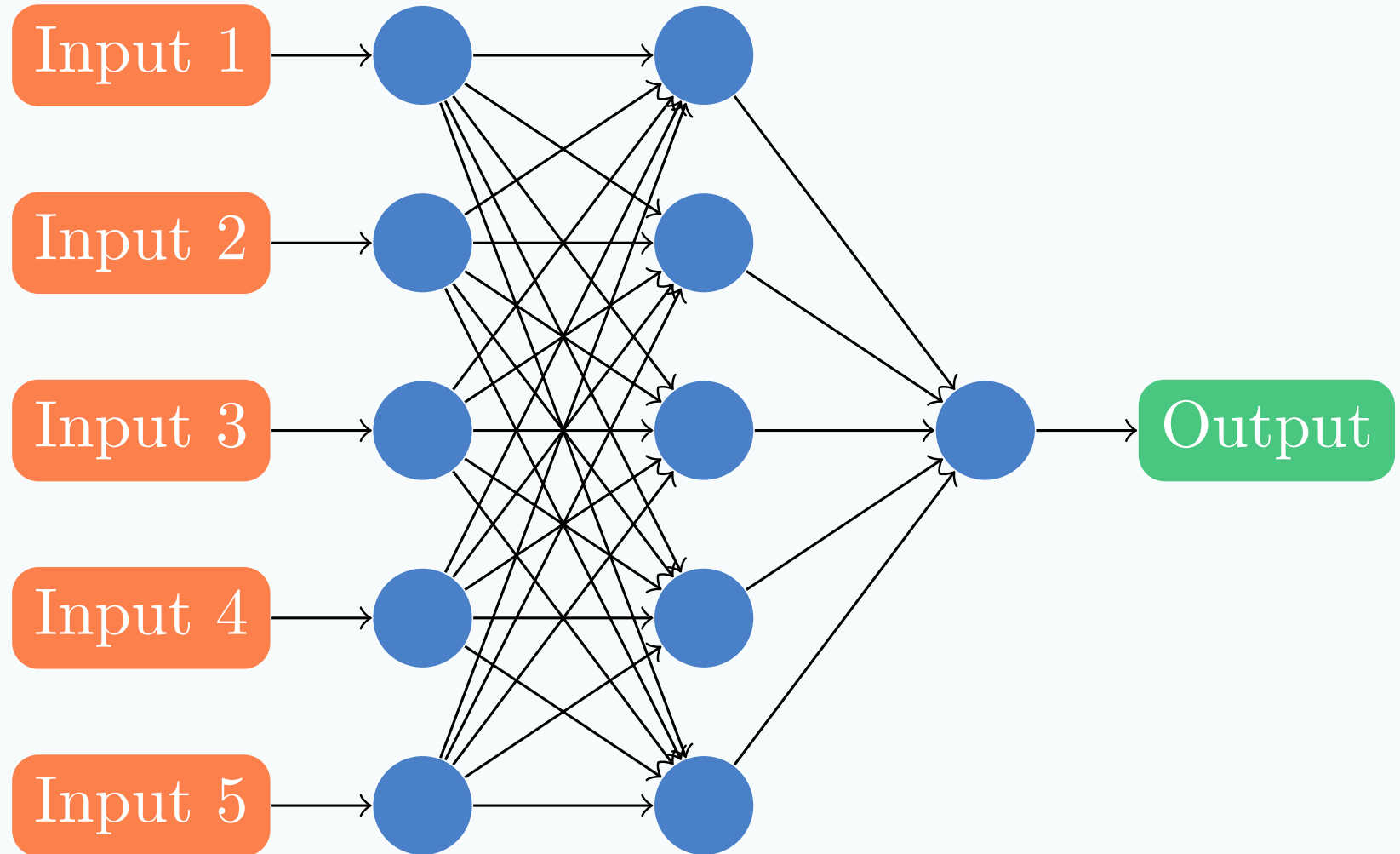


# NEURAL NETWORKS

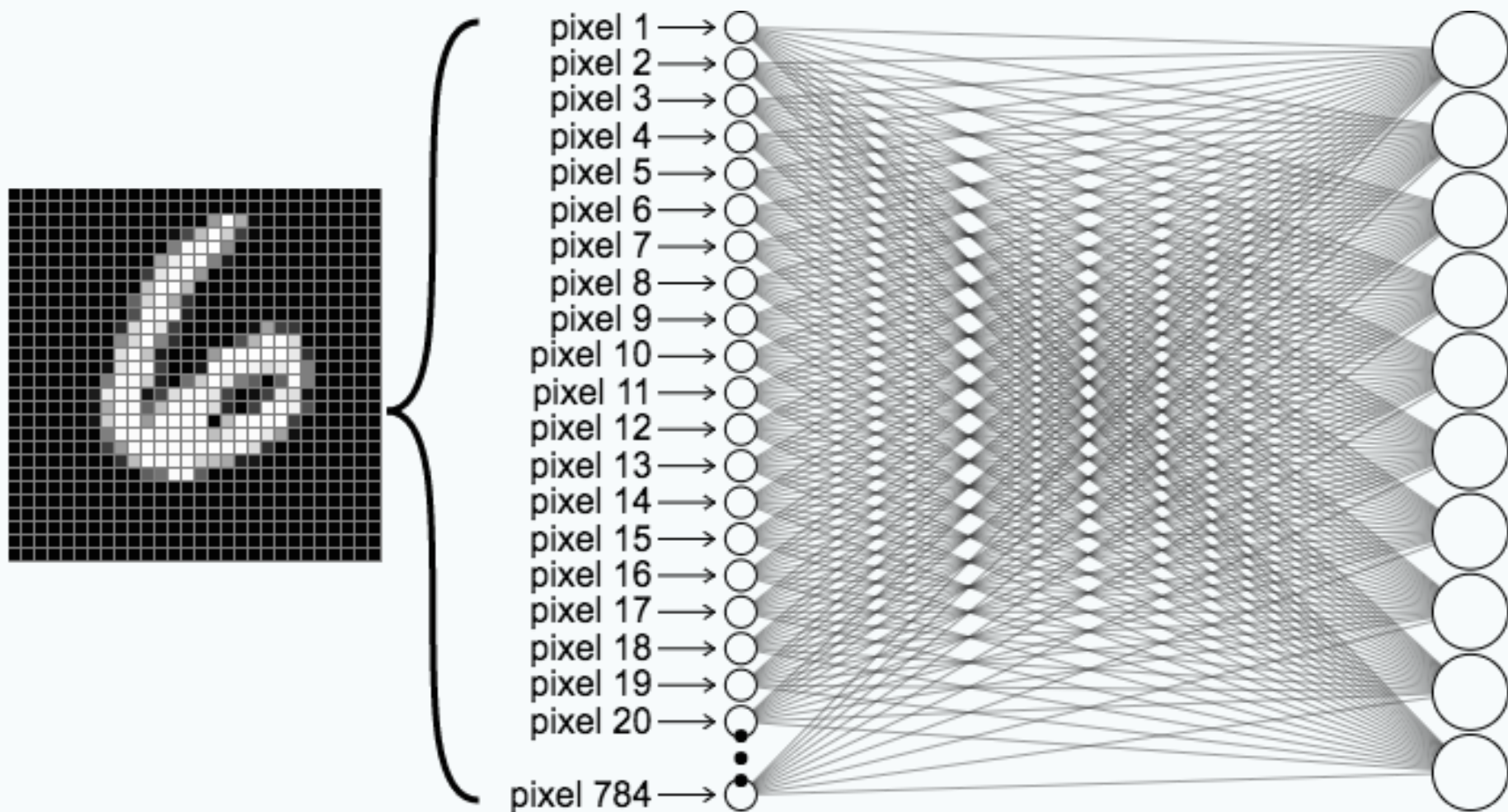




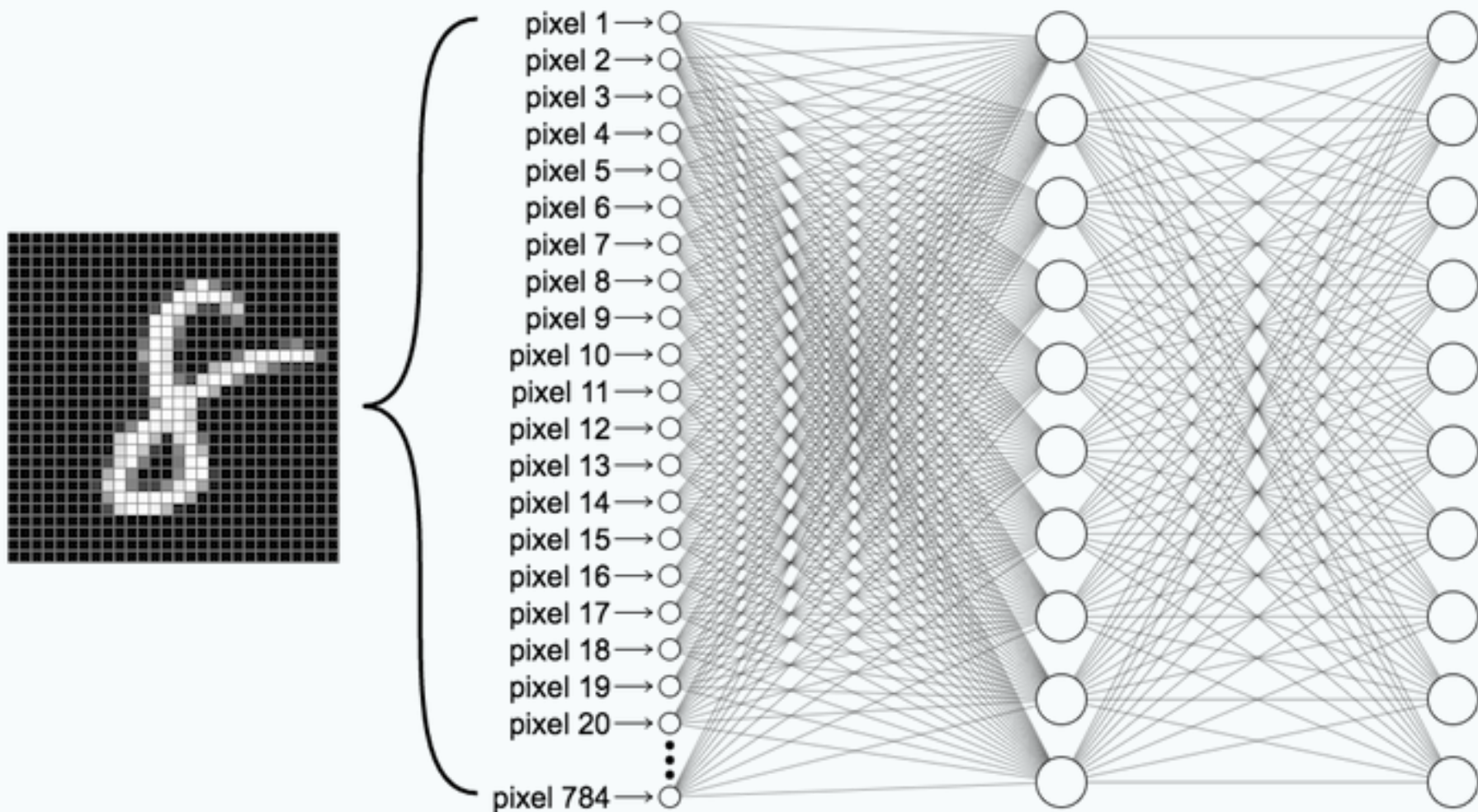
# NEURAL NETWORKS



# NEURAL NETWORKS

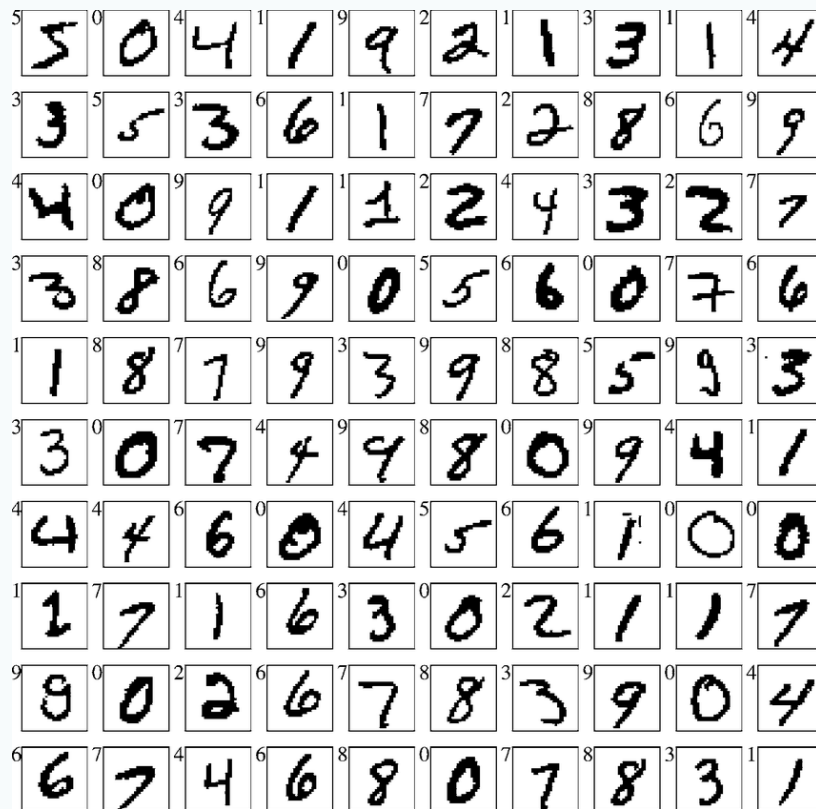


# NEURAL NETWORKS

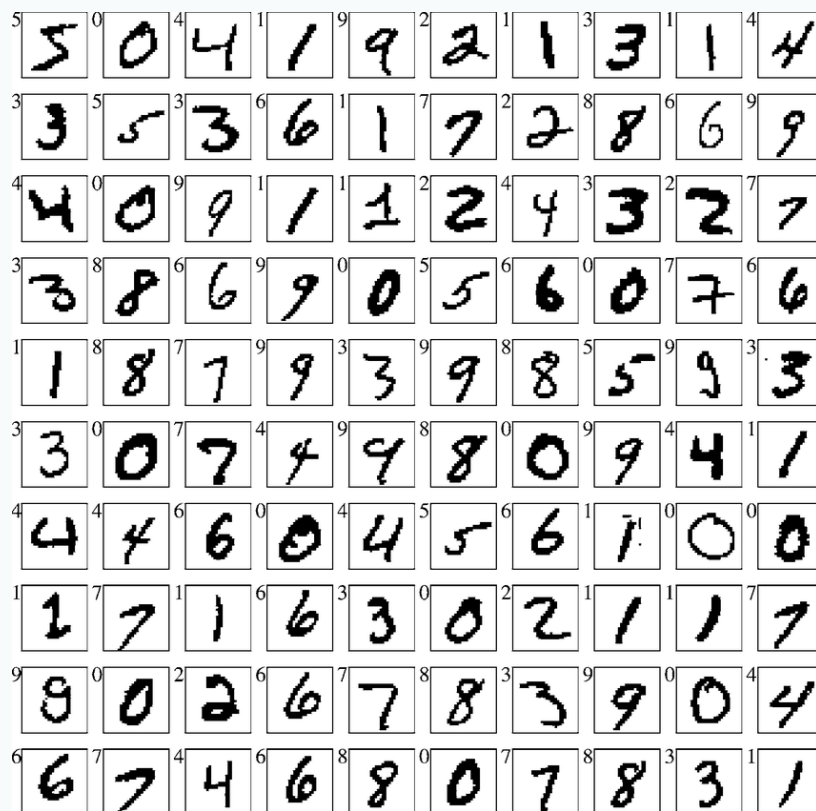


SURE, BUT HOW DOES IT LEARN?!

# SURE, BUT HOW DOES IT LEARN?!



# SURE, BUT HOW DOES IT LEARN?!



## Supervised Learning

# WHY DO WE CARE?

Real life problems are highly non-linear

Difficult to find rule based algorithms

# EXAMPLES

Image recognition

Translations

Self-Driving

Natural Language Processing

...



# MACHINE LEARNING IN ACCOUNTING

Error detections

High risk transactions detection

Transaction allocations

Fraud detections

# MACHINE LEARNING IN ACCOUNTING

Error detections

High risk transactions detection

Transaction allocations

Fraud detections

# MACHINE LEARNING IN PUBLIC ACCOUNTING

# MACHINE LEARNING IN PUBLIC ACCOUNTING

ML requires a lot of data to learn

# MACHINE LEARNING IN PUBLIC ACCOUNTING

ML requires a lot of data to learn

Audit firms observe many firms over time

# MACHINE LEARNING IN PUBLIC ACCOUNTING

ML requires a lot of data to learn

Audit firms observe many firms over time

They observe errors in many different situations

# MACHINE LEARNING IN PUBLIC ACCOUNTING

ML requires a lot of data to learn

Audit firms observe many firms over time

They observe errors in many different situations

Increased ability to use ML methods

RPA + ML = INTELLIGENT AUTOMATION



THIS IS AWESOME!

THIS IS AWESOME!

NOT REALLY...

# Overfitting problems

## Regularization methods

Techniques allow to tackle complex problems

But we do not yet have enough ressources

COMPUTERS ARE "TOO" SMART ...

Overfitting problems

Regularization methods

Techniques allow to tackle complex problems

But we do not yet have enough ressources

COMPUTERS ARE "TOO" SMART ...

Overfitting problems

Regularization methods

OR NOT GOOD ENOUGH...

Techniques allow to tackle complex problems

But we do not yet have enough ressources

COMPUTERS ARE "TOO" SMART ...

Overfitting problems

Regularization methods

OR NOT GOOD ENOUGH...

Techniques allow to tackle complex problems

But we do not yet have enough ressources

BASICALLY, IT IS HARD TO DESIGN ...

# MACHINE LEARNING PITFALLS (1)

# MACHINE LEARNING PITFALLS (1)

ML (supervised) requires a lot of data to learn



# MACHINE LEARNING PITFALLS (2)

# MACHINE LEARNING PITFALLS (2)

Can only learn information from the given data

# MACHINE LEARNING PITFALLS (2)

Can only learn information from the given data

Garbage In = Garbage Out

# MACHINE LEARNING PITFALLS (2)

Can only learn information from the given data

Garbage In = Garbage Out

*Cannot* adapt to unknown situations

# MACHINE LEARNING PITFALLS (3)

# MACHINE LEARNING PITFALLS (3)

Need to have a way to improve ...

# MACHINE LEARNING PITFALLS (3)

Need to have a way to improve ...  
... still need humans

IS THAT ALL?



Accounting is about recording transactions

BLOCKCHAIN

# BLOCKCHAIN

Created to solve the double spending problem  
without trusted third party

# TRADITIONALLY

Third party records all transactions

When asked for, can say if a new transaction is possible (i.e. pay)

All parties involved trust the third party

# BLOCKCHAIN IN ACCOUNTING

Like Accounting, Blockchain is about registering transactions

It seems likely to disrupt accounting

# BLOCKCHAIN IN ACCOUNTING

Audit is about verifying transactions

Blockchain provides a way to trust the information

Implementations started in China

SHOULD WE BE WORRIED?

NOT REALLY ...

Accounting involves a lot of reasoning

REMEMBER



# REMEMBER

## 1. Identifying Transactions

# REMEMBER

1. Identifying Transactions
2. Recording / Measuring -- Needs concepts (GAPP)

# REMEMBER

1. Identifying Transactions
2. Recording / Measuring -- Needs concepts (GAPP)
3. Communicating (BS, OE, IS, SCF)

MAYBE ONE DAY

MAYBE ONE DAY

BLOCKCHAIN + ML

MAYBE ONE DAY

BLOCKCHAIN + ML

But we should have some time

YOUR JOB

# YOUR JOB

Learn as many *concepts* as you can now



# YOUR JOB

Learn as many *concepts* as you can now

Understand the big picture and how things fit together

# YOUR JOB

Learn as many *concepts* as you can now

Understand the big picture and how things fit together

Rely on this knowledge to adapt

# YOUR JOB

Learn as many *concepts* as you can now

Understand the big picture and how things fit together

Rely on this knowledge to adapt

The world will look a lot different in 30 years

# YOUR JOB

Learn as many *concepts* as you can now

Understand the big picture and how things fit together

Rely on this knowledge to adapt

The world will look a lot different in 30 years

Never stop learning!

THANK YOU