HOW TO MAKE EXPERTS WORK WITH THE ALGORITHM

Albrecht Zimmermann, CODAG, GREYC

PhD Day, Caen

7.11.2023

GREYC
Laboratoire de recherche en sciences du numérique
Normandie Université
UNICAEN
ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE
cnrs

Outline

What are they speaking off when they say « AI » ?

- Learning
  - Supervised, semi-supervised, (generative), unsupervised
- Reinforcement learning
- Data Mining

- Intelligence : « ability to adapt to new circumstances »

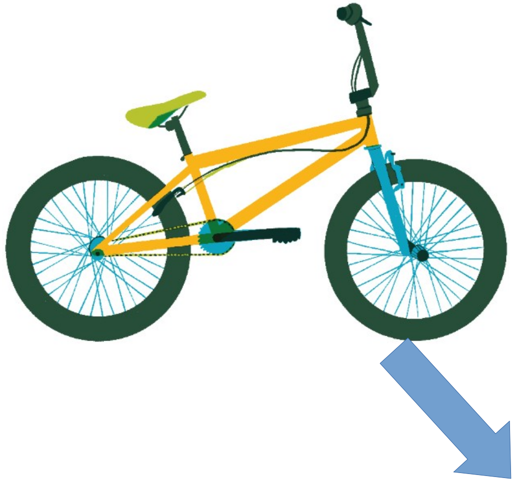
What are they speaking off when they say « AI » ?

- Learning
  - Supervised, semi-supervised, (generative), unsupervised
- Reinforcement learning
- Data Mining

- Intelligence : « ability to adapt to new circumstances »

What are they speaking off when they say « AI » ?

- Learning
  - Supervised, semi-supervised, (generative), unsupervised
- Reinforcement learning
- Data Mining

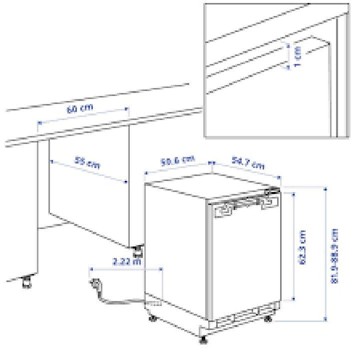


- Intelligence : « ability to adapt to new circumstances »

What are they speaking off when they say « AI » ?

- Learning
  - Supervised, semi-supervised, (generative), unsupervised
- Reinforcement learning
- Data Mining

- Intelligence : « ability to adapt to new circumstances »

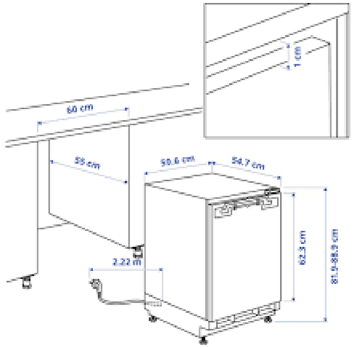


What are they speaking off when they say « AI » ?

- Learning
  - Supervised, semi-supervised, (generative), unsupervised
- Reinforcement learning
- Data Mining

- Intelligence : « ability to adapt to new circumstances »

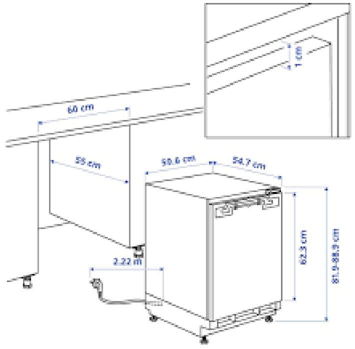
Knowledge transfer



What are they speaking off when they say « AI » ?

- Learning
  - Supervised, semi-supervised, (generative), unsupervised
- Reinforcement learning
- Data Mining

Knowledge transfer



- Intelligence : « ability to adapt to new circumstances »
  - → artificial, yes ! Intelligence, no !

(Semi-)supervised learning

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and
  - A test set
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

(Semi-)supervised learning

Lots of work in the 80s

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and
  - A test set
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

(Semi-)supervised learning

Lots of work in the 80s

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

(Semi-)supervised learning

Lots of work in the 80s

GREYC
Laboratoire de recherche en sciences du numérique

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

(Semi-)supervised learning

Lots of work in the 80s

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

Learning method

(Semi-)supervised learning

Lots of work in the 80s

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

Train → **Learning method**

(Semi-)supervised learning

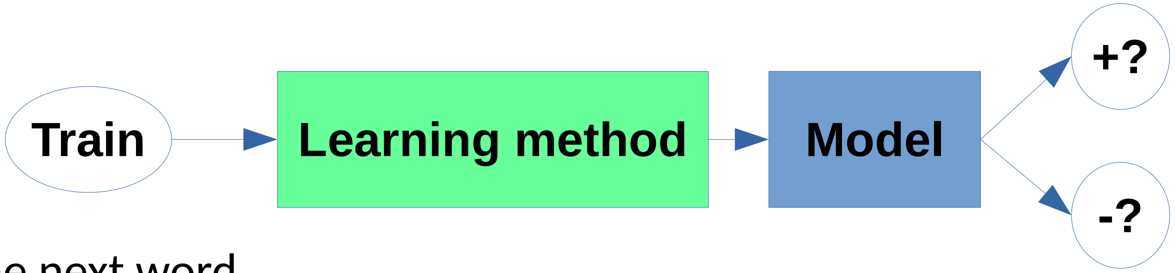
Lots of work in the 80s

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and — To find the parameters
  - A test set — To evaluate whether it works on other data
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

(Semi-)supervised learning

Lots of work in the 80s

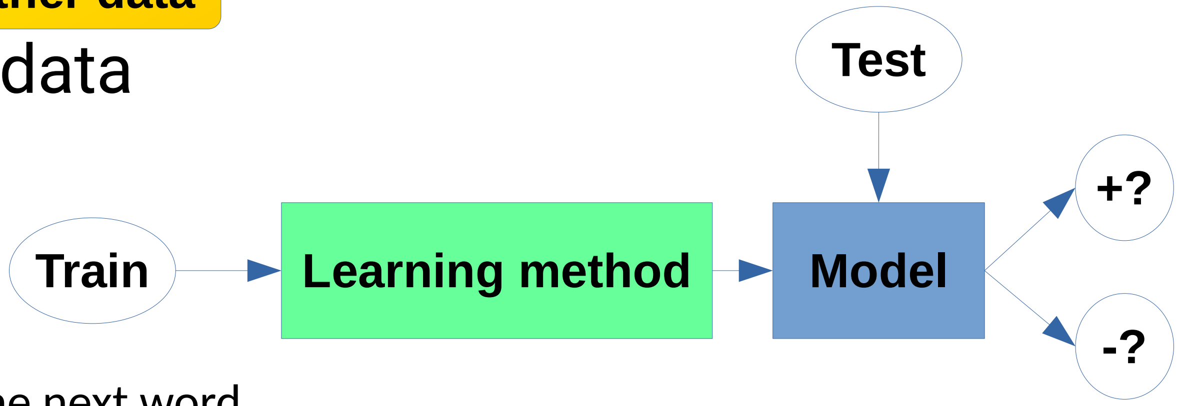
- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match



(Semi-)supervised learning

Lots of work in the 80s

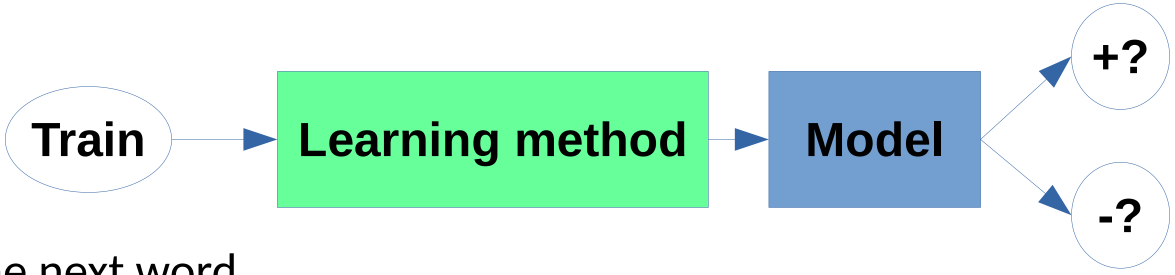
- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match



(Semi-)supervised learning

Lots of work in the 80s

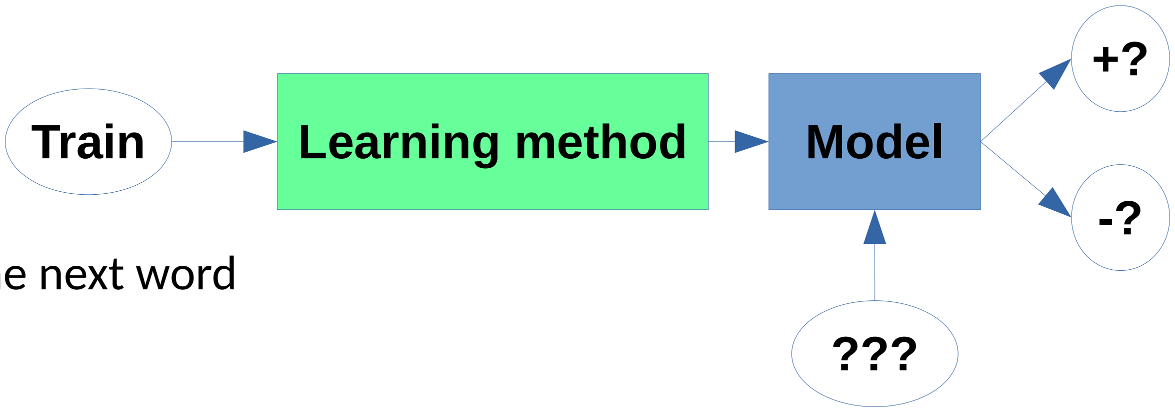
- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match



(Semi-)supervised learning

Lots of work in the 80s

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match

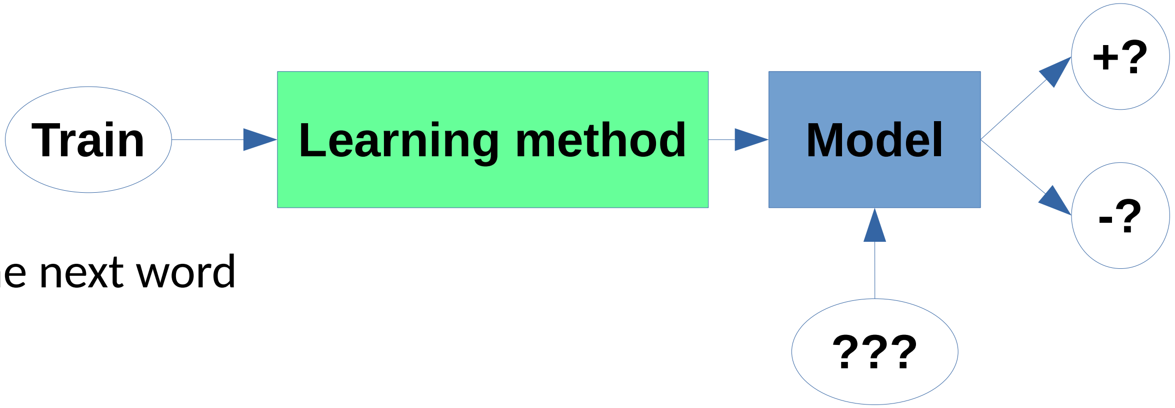


(Semi-)supervised learning

Lots of work in the 80s

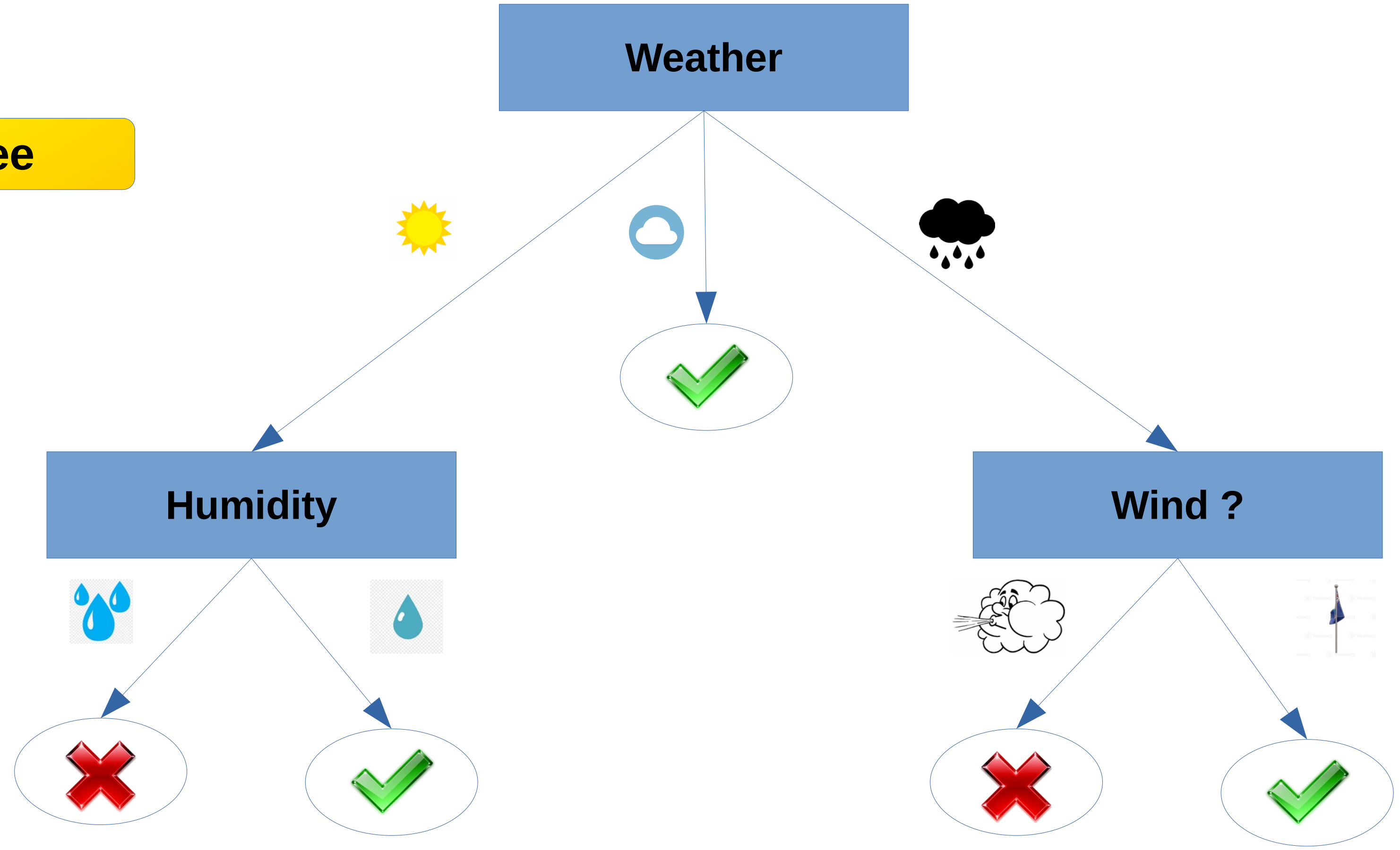
Semi-supervised ? Mix of labeled and unlabeled data

- Each « data point » has a « label »
  - Which guides the learning of a « model »
  - Learning : find good parameters of the model
  - Normally, one uses a
  - Training set and **To find the parameters**
  - A test set **To evaluate whether it works on other data**
- Goal : predict the label of unlabeled data
- For example :
  - Give a loan : yes or no
  - Classify an x-ray : cancer or not
  - Large language models (e.g. [Chat]GPT) : predict the next word
  - Predict who wins a basketball match



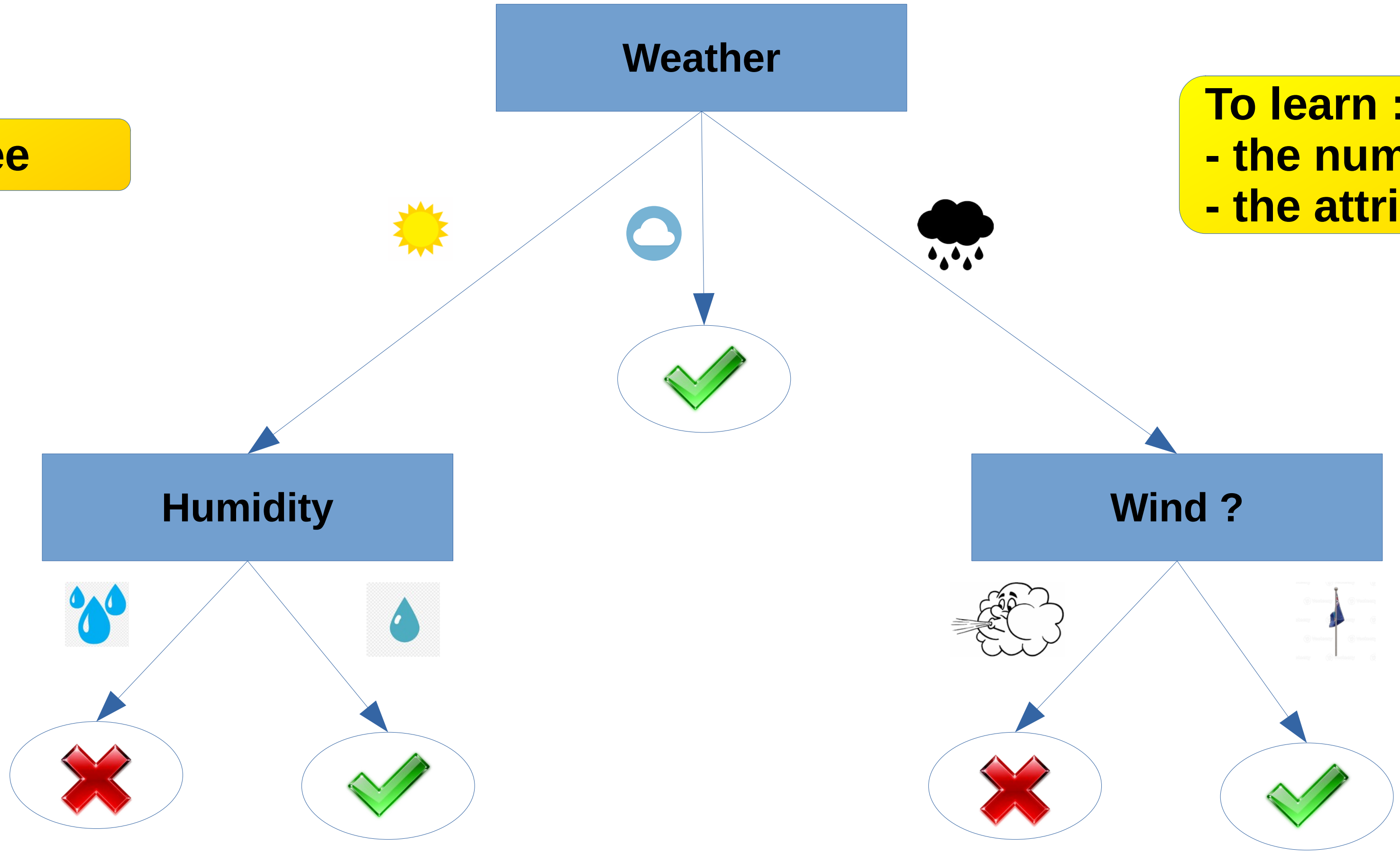
A model ?

Decision tree



A model ?

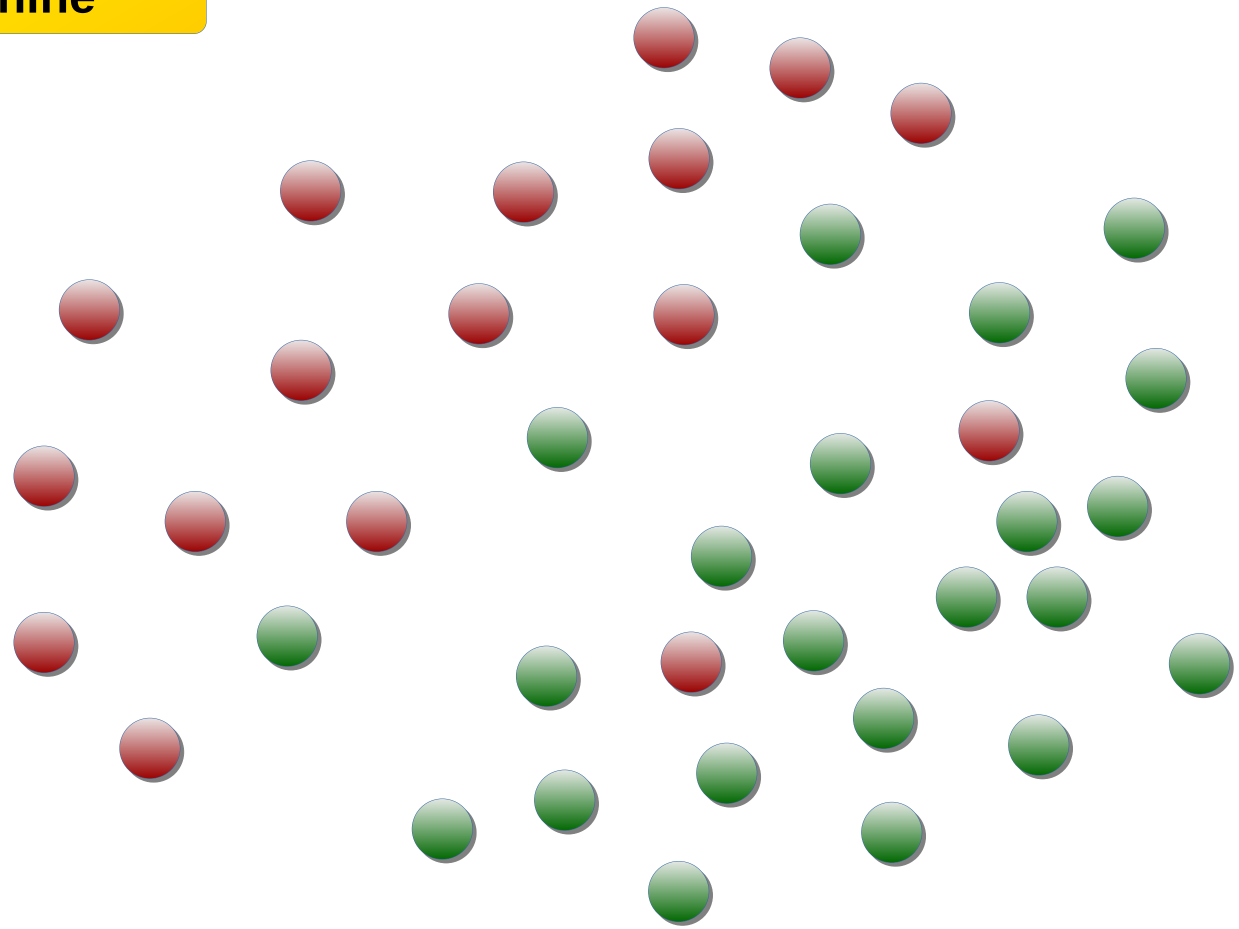
Decision tree



To learn :
- the number of tests
- the attributes to test

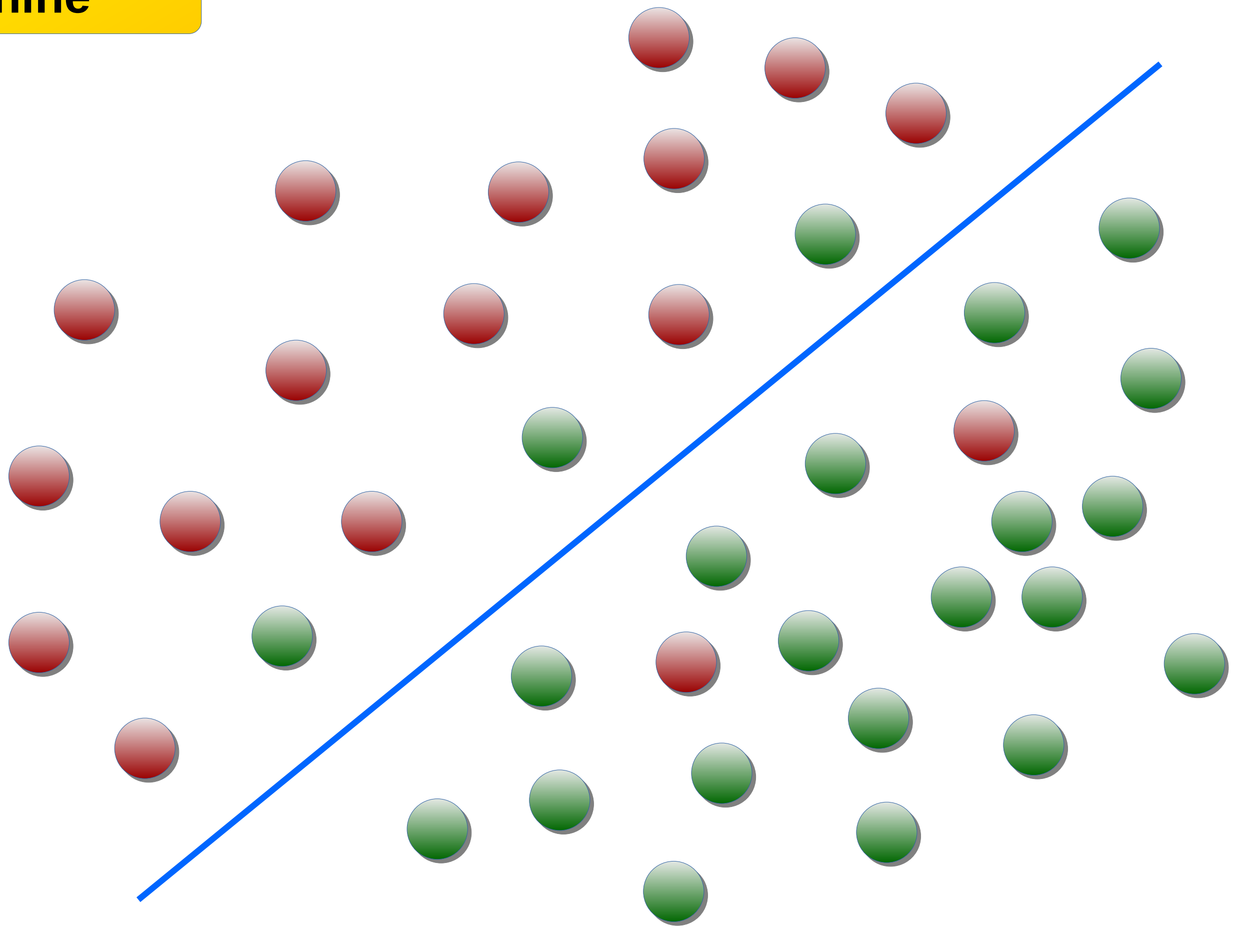
A model ? (2)

Support vector machine



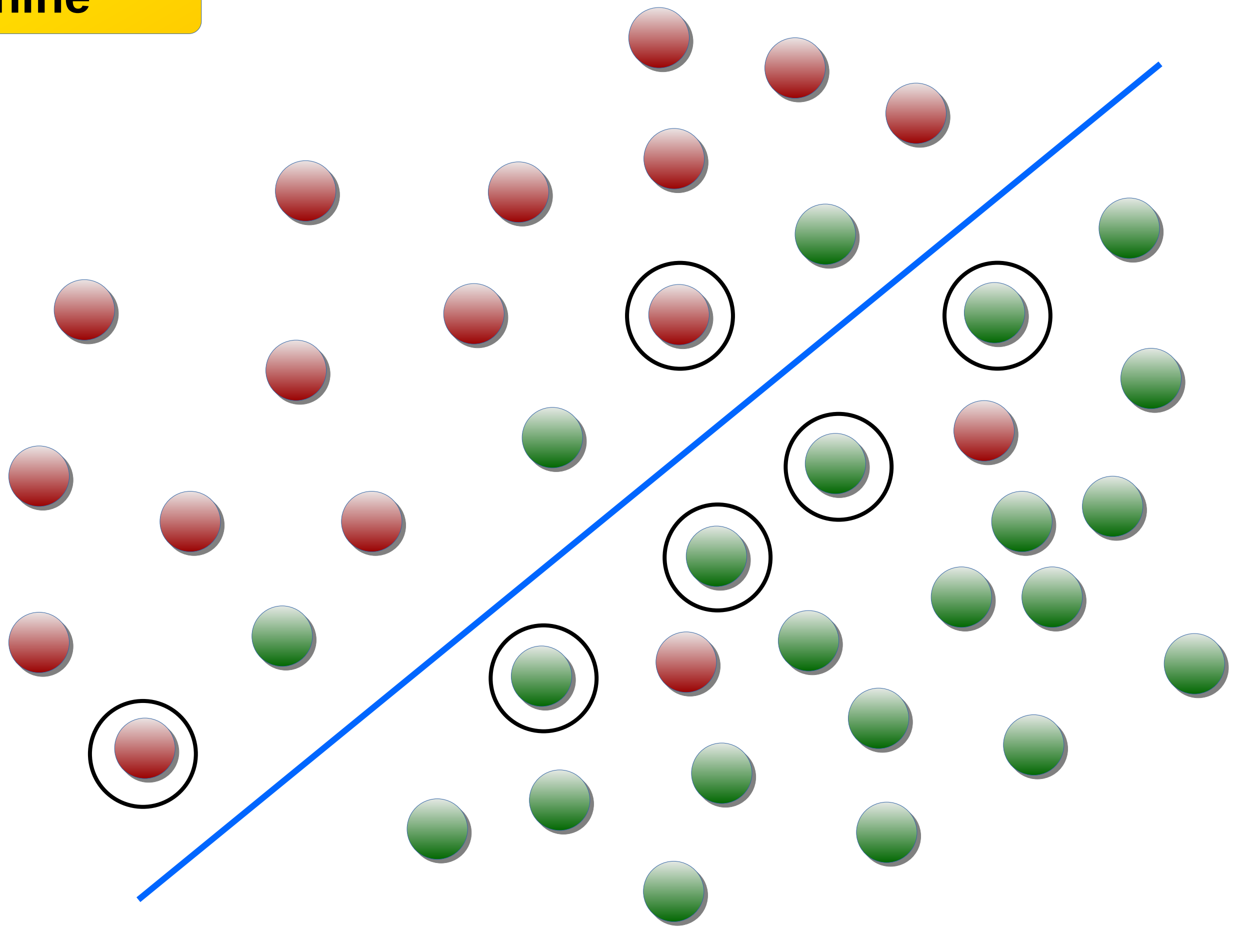
A model ? (2)

Support vector machine



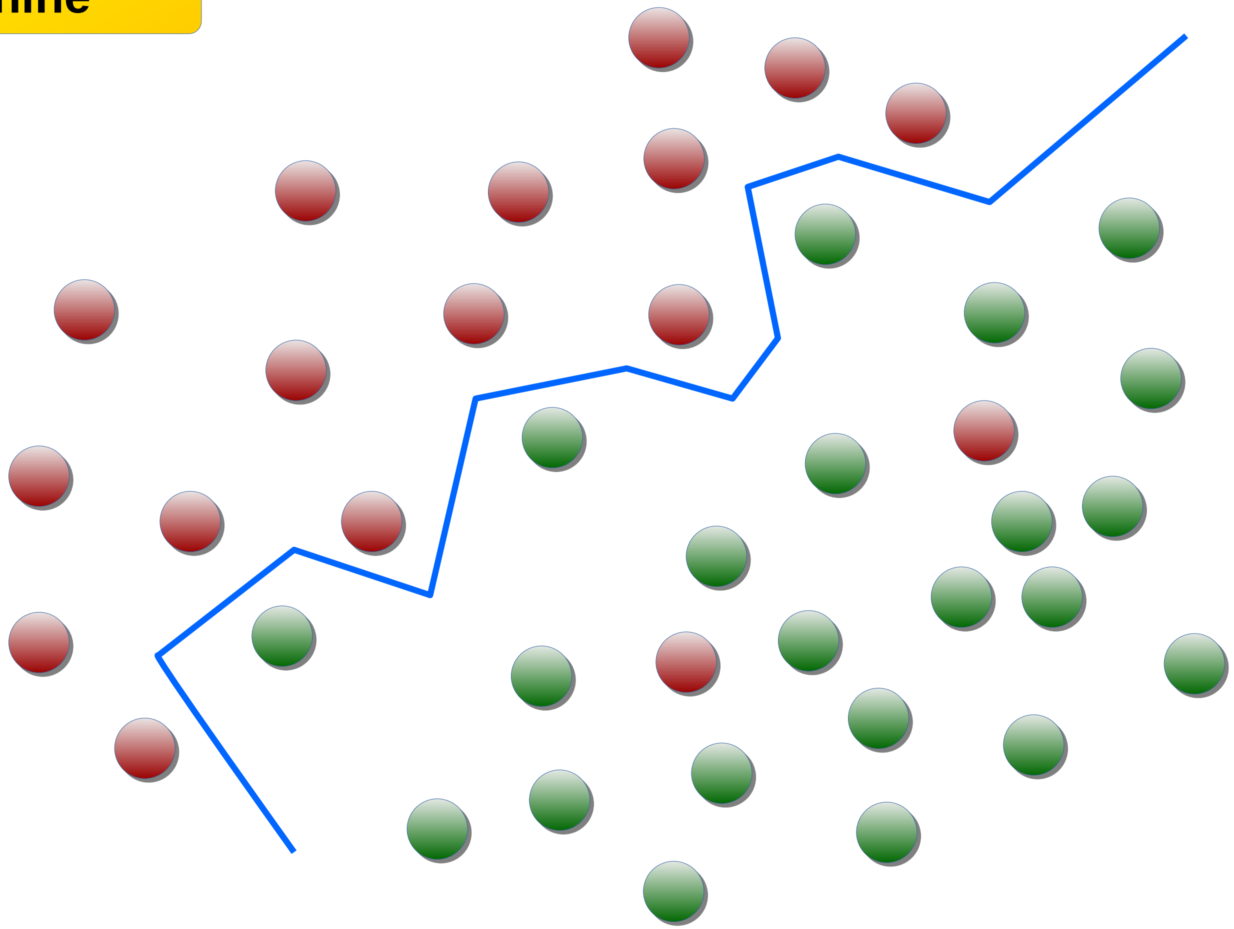
A model ? (2)

Support vector machine



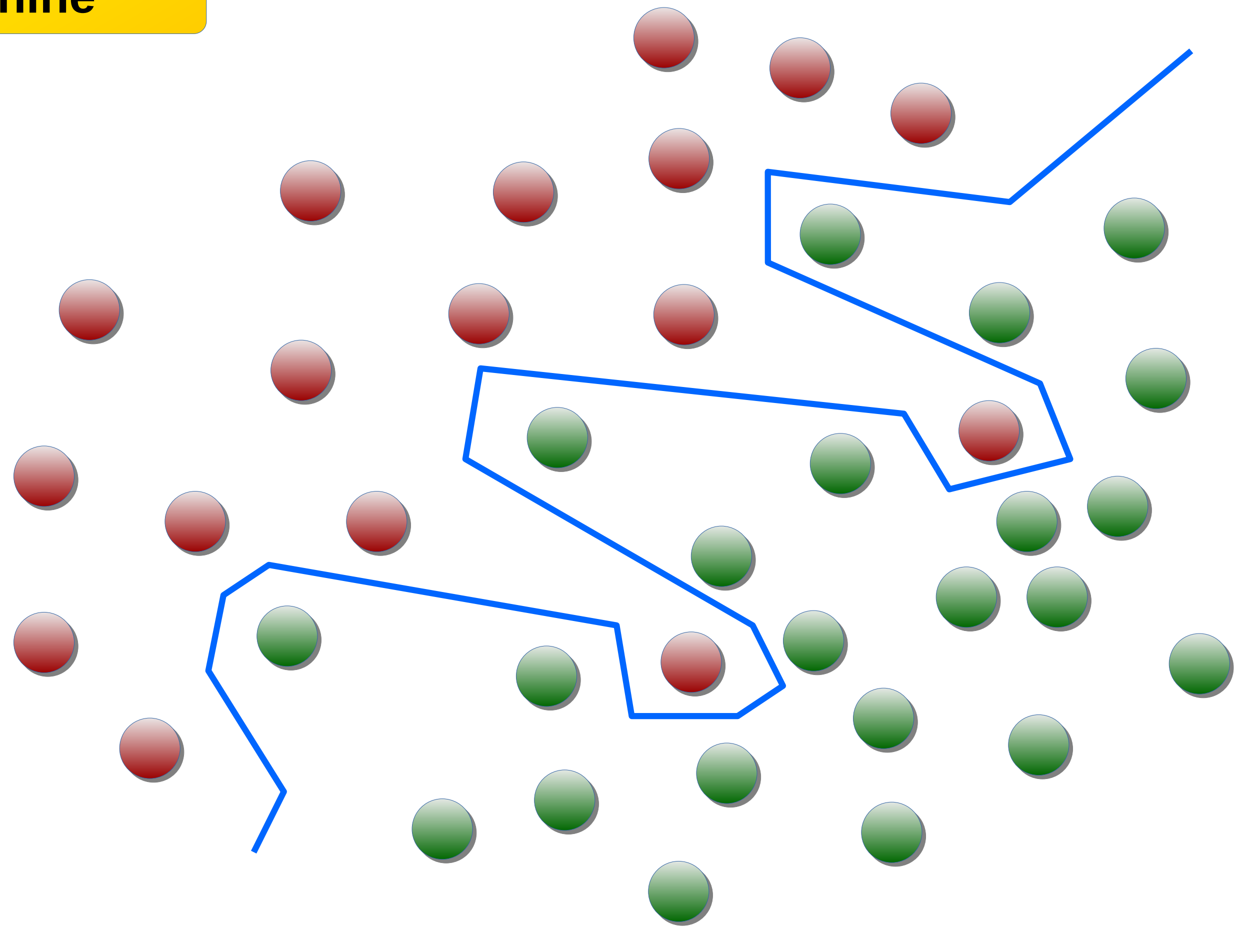
A model ? (2)

Support vector machine



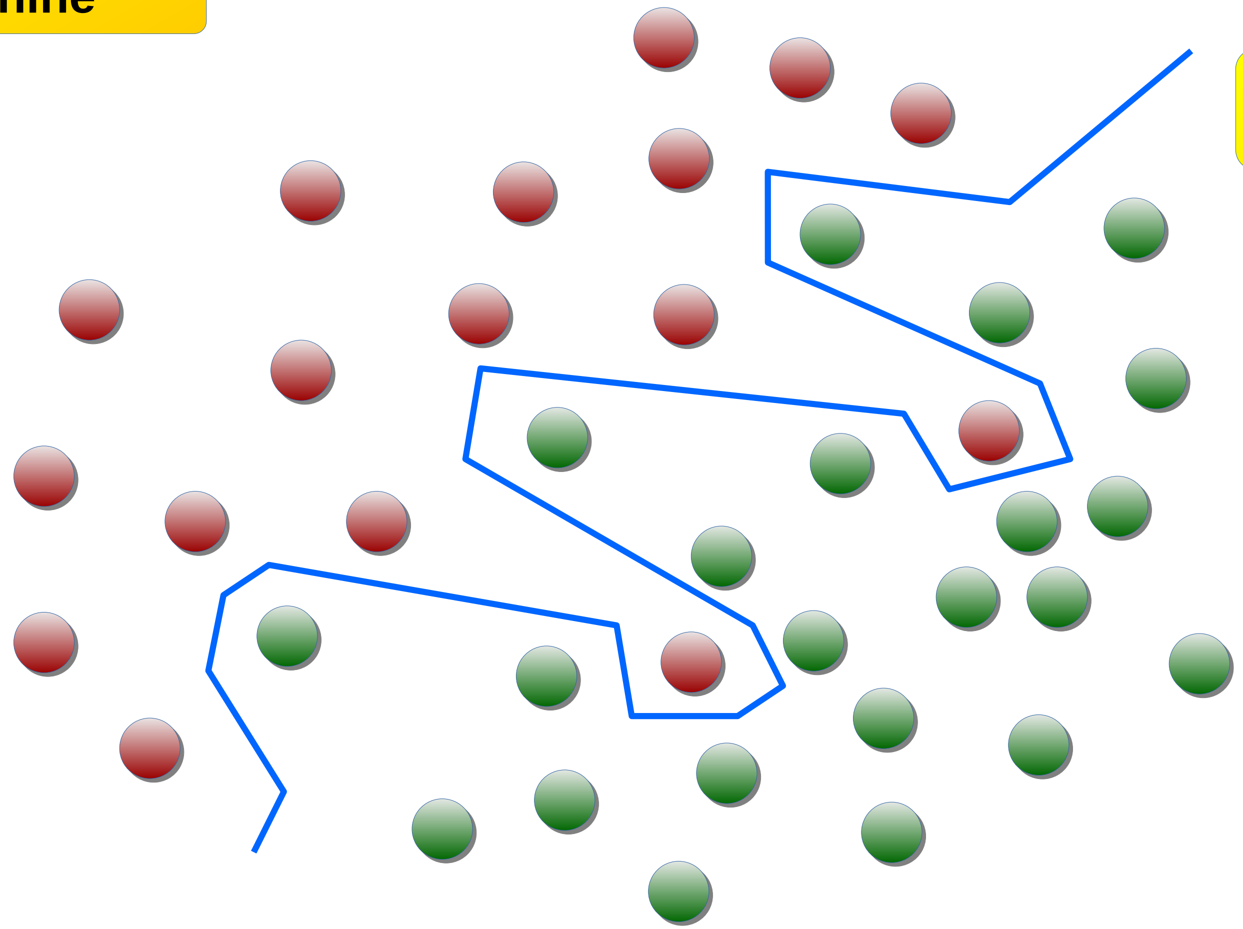
A model ? (2)

Support vector machine



A model ? (2)

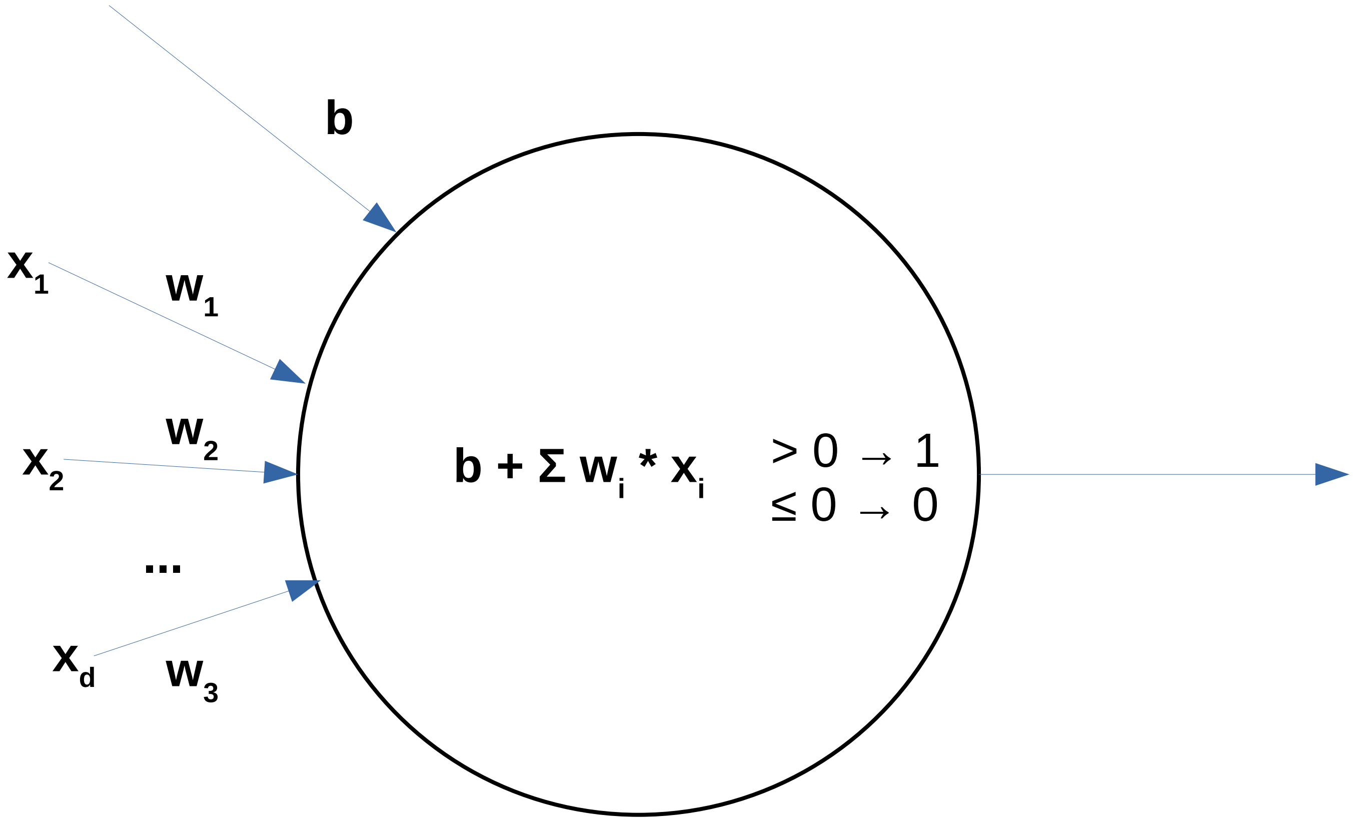
Support vector machine



To learn : the shape of the decision surface

A model ? (3)

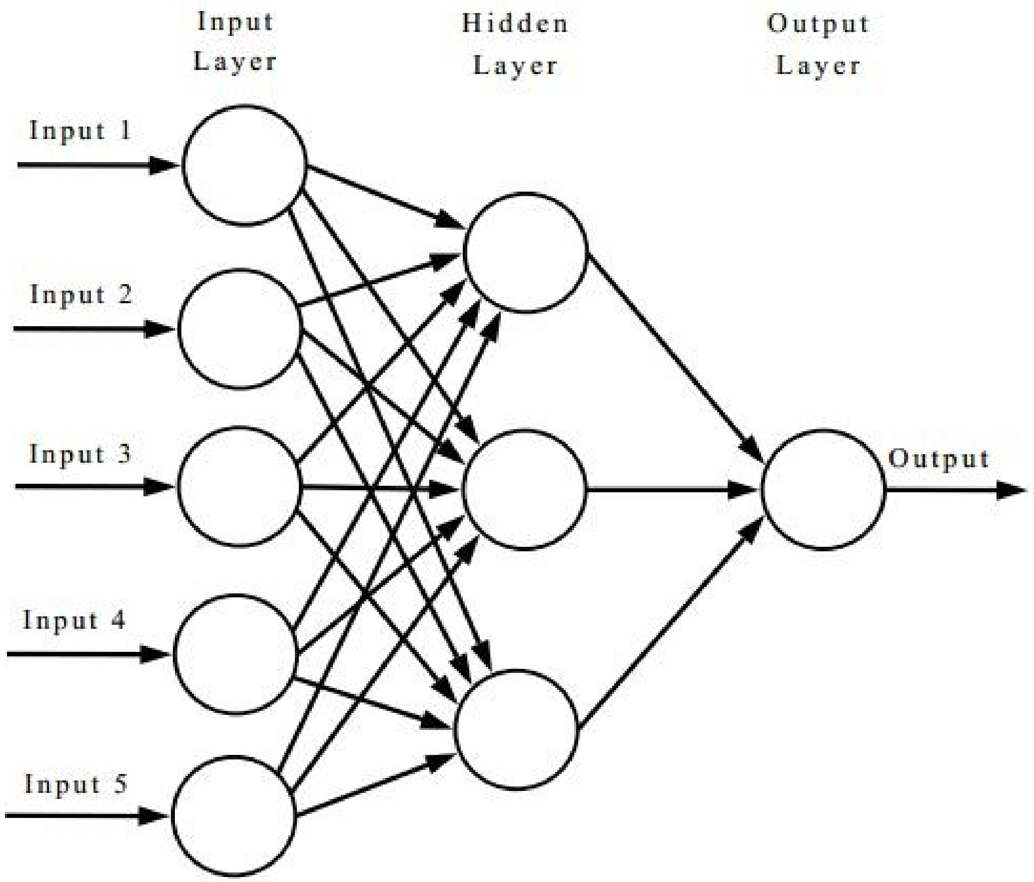
The perceptron : a simplified neuron



A model ? (4)

Neural network

Representation



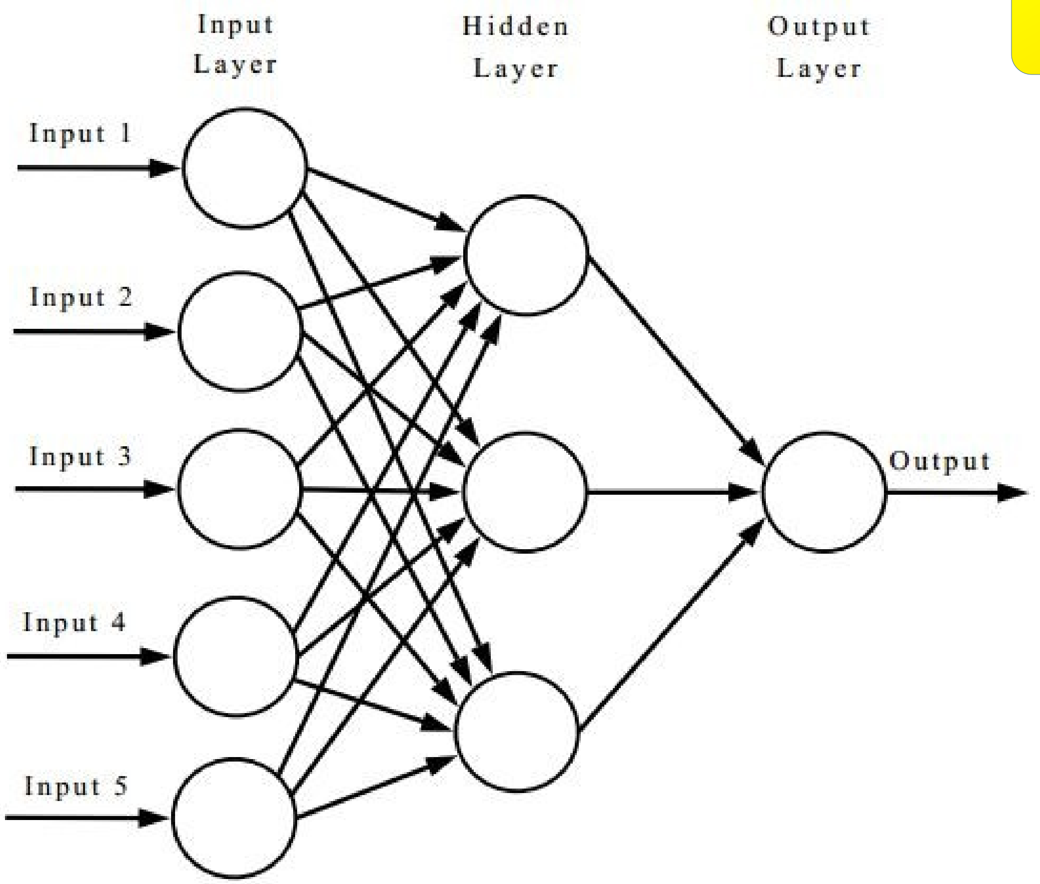
Prediction

A model ? (4)

Neural network

To learn : the connection weights

Representation



Prediction

Unsupervised learning

- No labels
- Grouping « similar » data
  - Similarity : function that takes two data points and returns a number

Unsupervised learning

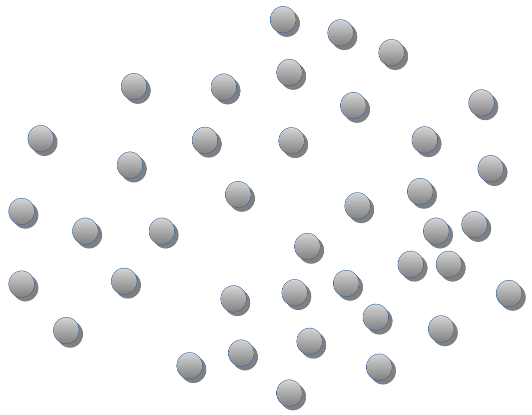
Earliest work in the 60s

- No labels
- Grouping « similar » data
  - Similarity : function that takes two data points and returns a number

Unsupervised learning

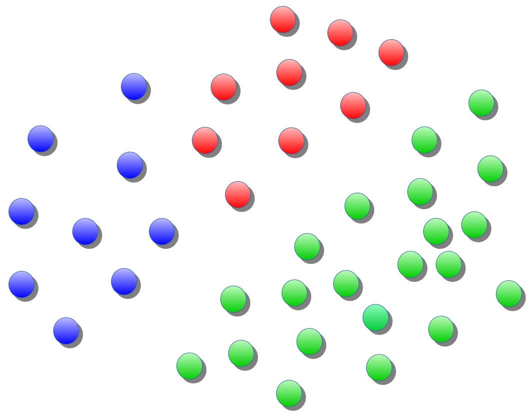
Earliest work in the 60s

- No labels
- Grouping « similar » data
  - Similarity : function that takes two data points and returns a number



Unsupervised learning

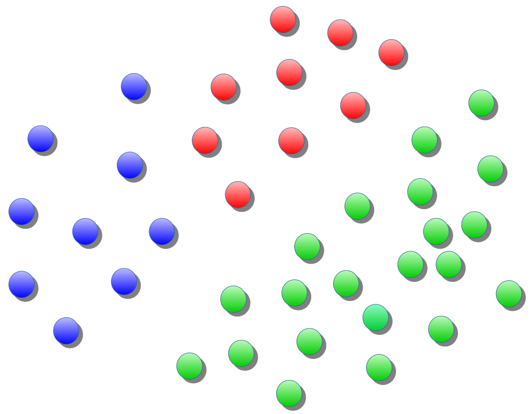
- No labels
- Grouping « similar » data



Unsupervised learning

Can be done with NNs

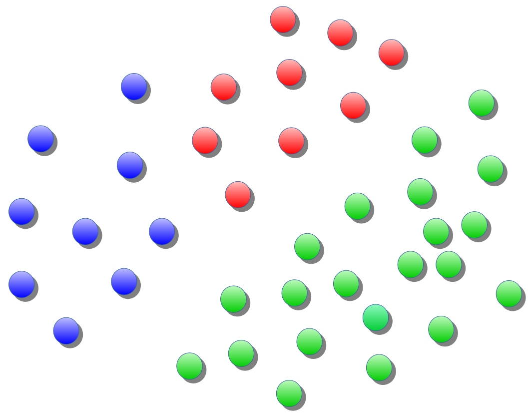
- No labels
- Grouping « similar » data

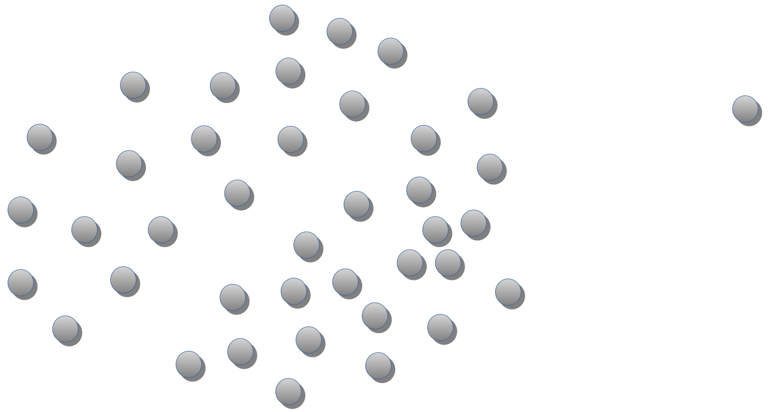


Unsupervised learning

Can be done with NNs

- No labels
- Grouping « similar » data
  - Similarity : function that takes two data points and returns a number

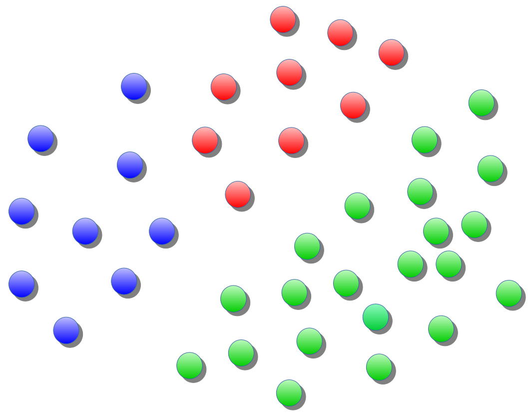




Unsupervised learning

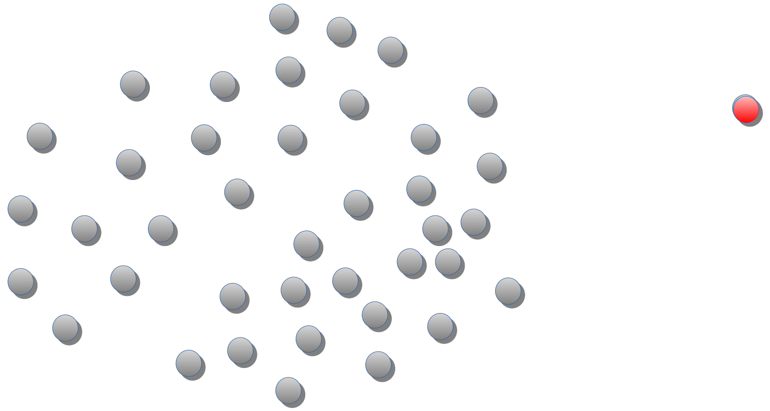
Can be done with NNs

- No labels
- Grouping « similar » data
  - Similarity : function that takes two data points and returns a number



  - Understand new data, identify categories

- Finding « anomalies »
  - Identify a network intrusion
  - Identify credit card fraud



Reinforcement learning

Can involve NNs

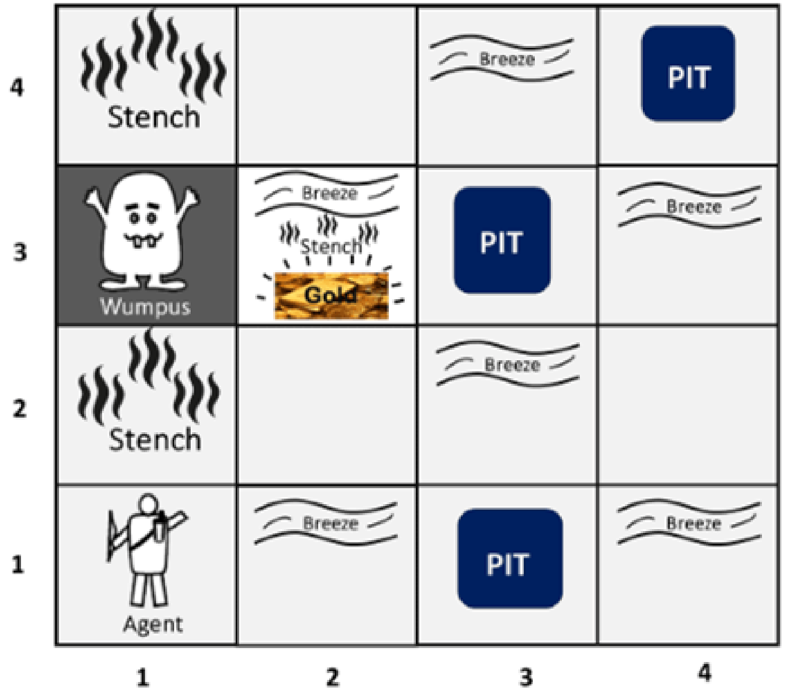
- A final goal : win a game, move to a place
- One can't define concrete rules
- Method :
  - Act in a ± random manner (exploration)
  - Once one reaches a final « state », receive a reward (a positive number) or a punishment (a negative number)
  - Assign a part of this number to prior actions (more recent actions get a higher reward/punishment)
  - Repeat

- Learn to play Go
- Learn to play Mario
- Teach a Roomba (or rather robot) to move around
- Teach drones to fly and hover

Reinforcement learning

Can involve NNs

- A final goal : win a game, move to a place
- One can't define concrete rules
- Method :
  - Act in a ± random manner (exploration)
  - Once one reaches a final « state », receive a reward (a positive number) or a punishment (a negative number)
  - Assign a part of this number to prior actions (more recent actions get a higher reward/punishment)
  - Repeat

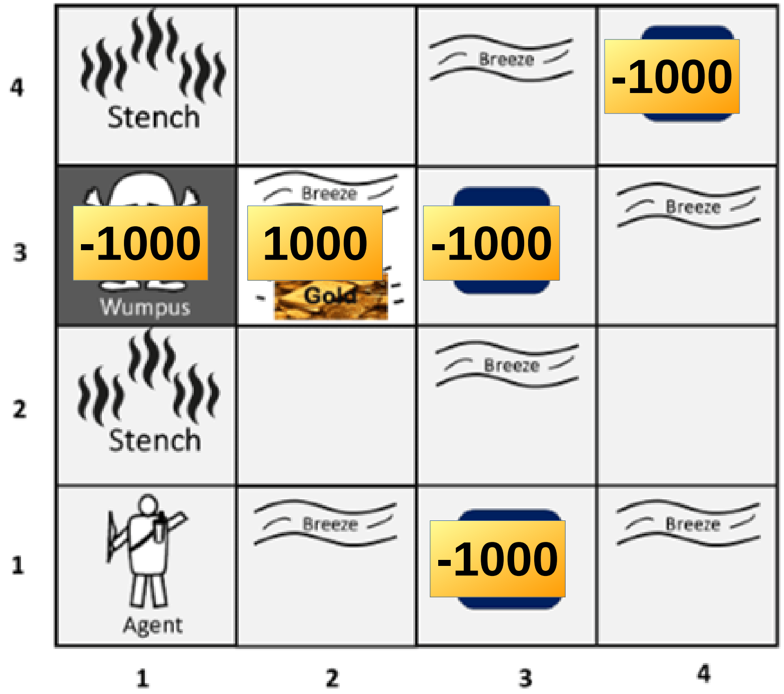


- Learn to play Go
- Learn to play Mario
- Teach a Roomba (or rather robot) to move around
- Teach drones to fly and hover

Reinforcement learning

Can involve NNs

- A final goal : win a game, move to a place
- One can't define concrete rules
- Method :
  - Act in a ± random manner (exploration)
  - Once one reaches a final « state », receive a reward (a positive number) or a punishment (a negative number)
  - Assign a part of this number to prior actions (more recent actions get a higher reward/punishment)
  - Repeat



- Learn to play Go
- Learn to play Mario
- Teach a Roomba (or rather robot) to move around
- Teach drones to fly and hover

Data mining

- Given a large mass of data
- Find « patterns » : sub-structures in the data
  - Regularities
  - Surprising aspects

- Find the products bought together in a supermarket
- Find molecular substructures in molecules that cure an illness
- Find an unexpected regularity
- Find descriptors for learning

Data mining

Began in the mid-90s

- Given a large mass of data
- Find « patterns » : sub-structures in the data
  - Regularities
  - Surprising aspects

- Find the products bought together in a supermarket
- Find molecular substructures in molecules that cure an illness
- Find an unexpected regularity
- Find descriptors for learning

Data mining

Began in the mid-90s

- Given a large mass of data
- Find « patterns » : sub-structures in the data
  - Regularities
  - Surprising aspects

Reorganise the shop

- Find the products bought together in a supermarket
- Find molecular substructures in molecules that cure an illness
- Find an unexpected regularity
- Find descriptors for learning

Data mining

Began in the mid-90s

- Given a large mass of data
- Find « patterns » : sub-structures in the data
  - Regularities
  - Surprising aspects

Reorganise the shop

- Find the products bought together in a supermarket
- Find molecular substructures in molecules that cure an illness

Develop drugs

- Find an unexpected regularity
- Find descriptors for learning

Data mining

Began in the mid-90s

- Given a large mass of data
- Find « patterns » : sub-structures in the data
  - Regularities
  - Surprising aspects

Reorganise the shop

- Find the products bought together in a supermarket
- Find molecular substructures in molecules that cure an illness

Develop drugs

- Find an unexpected regularity **Define a scientific hypothesis**
- Find descriptors for learning

Involving (or not) the human

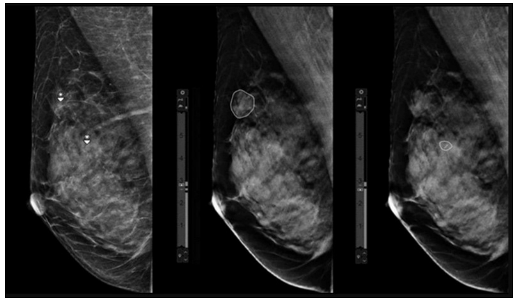
- In the beginning : we automate everything because humans see patterns everywhere and can't process large amounts (100k, 1M) of data

- BUT :

Involving (or not) the human

- In the beginning : we automate everything because humans see patterns everywhere and can’t process large amounts (100k, 1M) of data

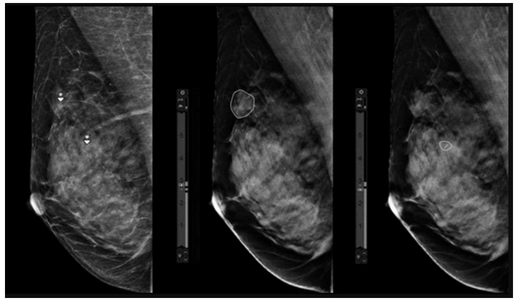
- BUT :
  - Trust : result explanation – predict presence of cancerous lesions via presence of a ruler



Involving (or not) the human

- In the beginning : we automate everything because humans see patterns everywhere and can’t process large amounts (100k, 1M) of data

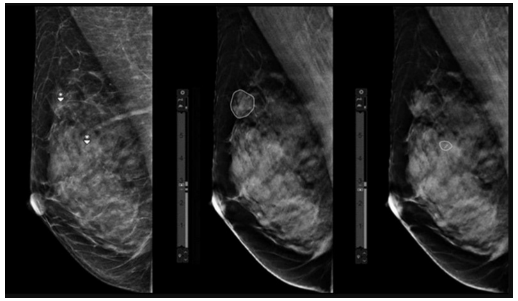
- BUT :
  - Trust : result explanation – predict presence of cancerous lesions via presence of a ruler
  - The model might learn historical discrimination – refuse a loan because the person’s black



Involving (or not) the human

- In the beginning : we automate everything because humans see patterns everywhere and can’t process large amounts (100k, 1M) of data

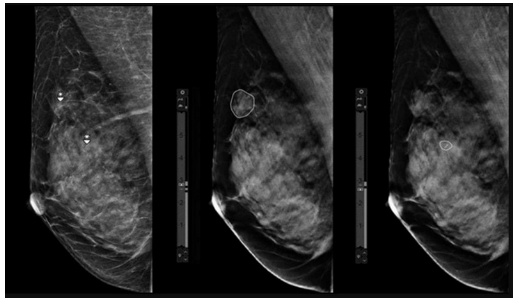
- BUT :
  - Trust : result explanation – predict presence of cancerous lesions via presence of a ruler
  - The model might learn historical discrimination – refuse a loan because the person’s black
  - Hard to modify the result if an expert’s not convinced/satisfied – modify the data or constraints on parameters



Involving (or not) the human

- In the beginning : we automate everything because humans see patterns everywhere and can't process large amounts (100k, 1M) of data

- BUT :
  - Trust : result explanation – predict presence of cancerous lesions via presence of a ruler
  - The model might learn historical discrimination – refuse a loan because the person's black
  - Hard to modify the result if an expert's not convinced/satisfied – modify the data or constraints on parameters

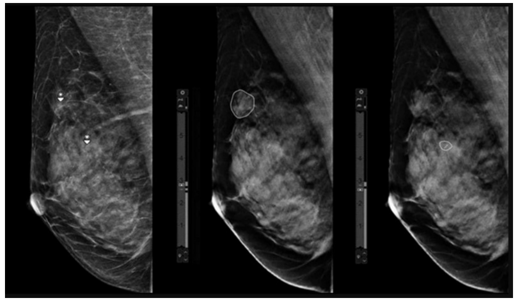


Indirect action

Involving (or not) the human

- In the beginning : we automate everything because humans see patterns everywhere and can’t process large amounts (100k, 1M) of data

- BUT :
  - Trust : result explanation – predict presence of cancerous lesions via presence of a ruler
  - The model might learn historical discrimination – refuse a loan because the person’s black
  - Hard to modify the result if an expert’s not convinced/satisfied – modify the data or constraints on parameters
  - Too many patterns – hard to interpret

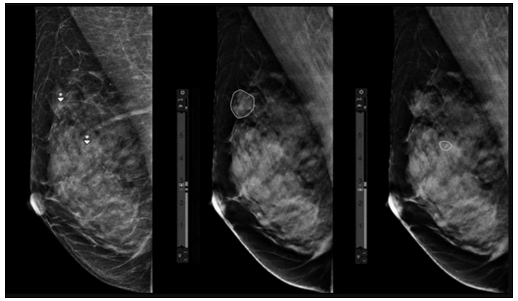


Indirect action

Involving (or not) the human

- In the beginning : we automate everything because humans see patterns everywhere and can't process large amounts (100k, 1M) of data

- BUT :
  - Trust : result explanation – predict presence of cancerous lesions via presence of a ruler
  - The model might learn historical discrimination – refuse a loan because the person's black
  - Hard to modify the result if an expert's not convinced/satisfied – modify the data or constraints on parameters
  - Too many patterns – hard to interpret
  - Certain actions damage or destroy the robot !



Indirect action

The expert in (semi-)supervised learning

1) Labeling

- Not always an « expert » - labeling images (captchas)
- Its own service sector : Kenya, Amérique latine, Asie (badly paid)
- Often still expert knowledge necessary

2) Active learning :

- Semi-supervised setting : identify data for which prediction is not very confident (knowing the true label would add a lot of information)
- Ask the expert to supply the label

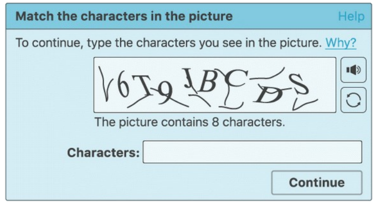
3) Correction :

- Identify prediction errors

4) User feedback :

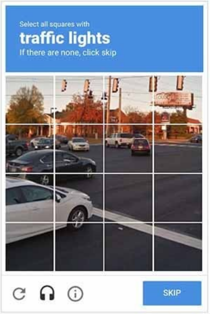
- Google translate
- LLMs

The expert in (semi-)supervised learning



1) Labeling

- Not always an « expert » - labeling images (captchas)
- Its own service sector : Kenya, Amérique latine, Asie (badly paid)
- Often still expert knowledge necessary



2) Active learning :

- Semi-supervised setting : identify data for which prediction is not very confident (knowing the true label would add a lot of information)
- Ask the expert to supply the label

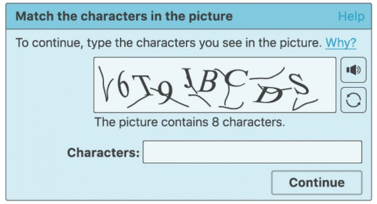
3) Correction :

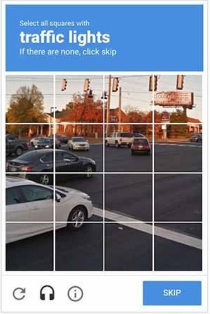
- Identify prediction errors

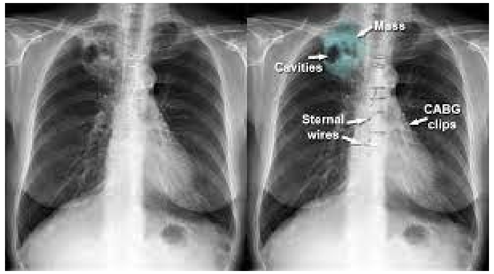
4) User feedback :

- Google translate
- LLMs

The expert in (semi-)supervised learning







1) Labeling

- Not always an « expert » - labeling images (captchas)
- Its own service sector : Kenya, Amérique latine, Asie (badly paid)
- Often still expert knowledge necessary

2) Active learning :

- Semi-supervised setting : identify data for which prediction is not very confident (knowing the true label would add a lot of information)
- Ask the expert to supply the label

3) Correction :

- Identify prediction errors

4) User feedback :

- Google translate
- LLMs

The expert in unsupervised learning

1) Constraints

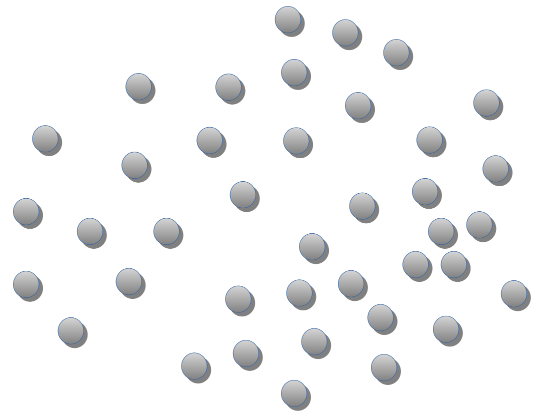
- Minimal/maximal group size
- Constraints saying two points « have to », « mustn't » be in the same group

2) Assign points by hand

3) Manually merge/split groups

4) Give more importance to certain descripteurs

- Of data points



The expert in unsupervised learning

1) Constraints

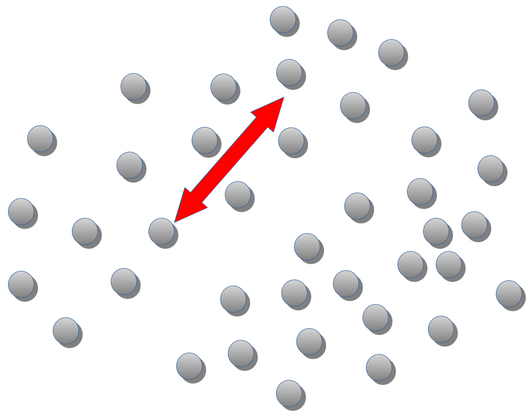
- Minimal/maximal group size
- Constraints saying two points « have to », « mustn’t » be in the same group

2) Assign points by hand

3) Manually merge/split groups

4) Give more importance to certain descripteurs

- Of data points



The expert in unsupervised learning

1) Constraints

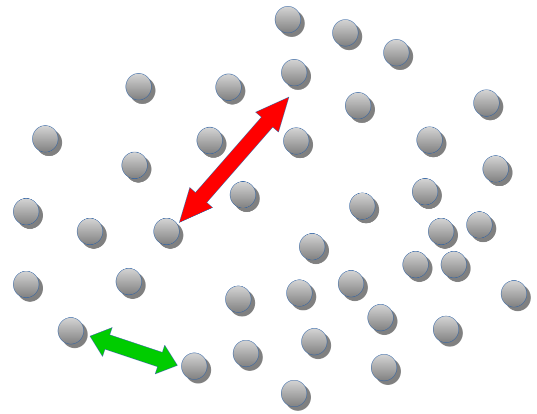
- Minimal/maximal group size
- Constraints saying two points « have to », « mustn't » be in the same group

2) Assign points by hand

3) Manually merge/split groups

4) Give more importance to certain descripteurs

- Of data points

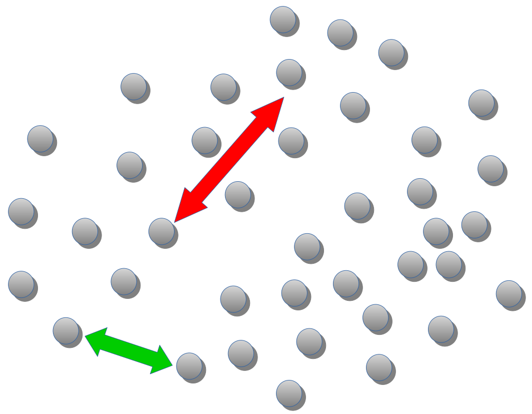


The expert in unsupervised learning

1) Constraints

- Minimal/maximal group size
- Constraints saying two points « have to », « mustn't » be in the same group

Can be done as active learning



2) Assign points by hand

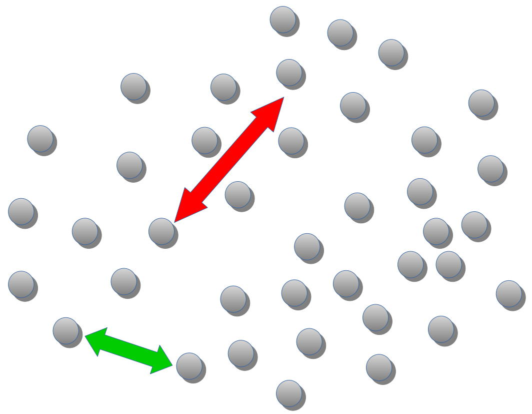
3) Manually merge/split groups

4) Give more importance to certain descripteurs

- Of data points

The expert in unsupervised learning

1) Constraints
- Minimal/maximal group size
- Constraints saying two points « have to », « mustn't » be in the same group

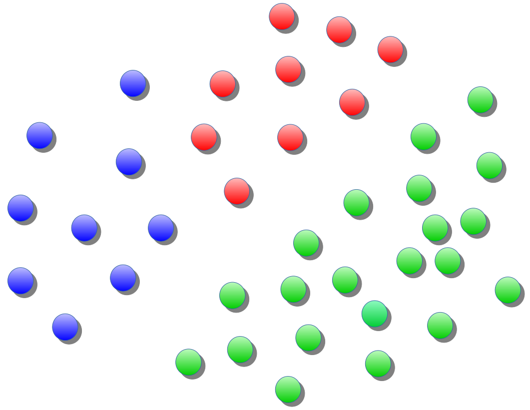


Can be done as active learning

2) Assign points by hand

3) Manually merge/split groups

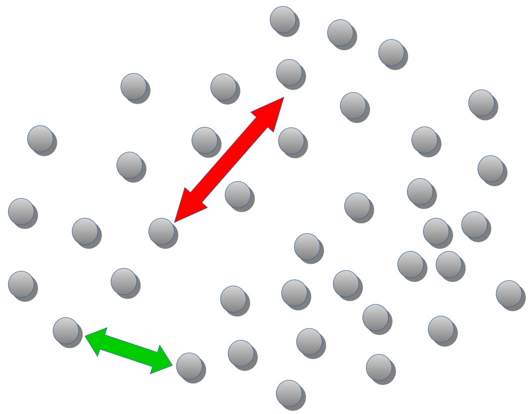
4) Give more importance to certain descripteurs
- Of data points



The expert in unsupervised learning

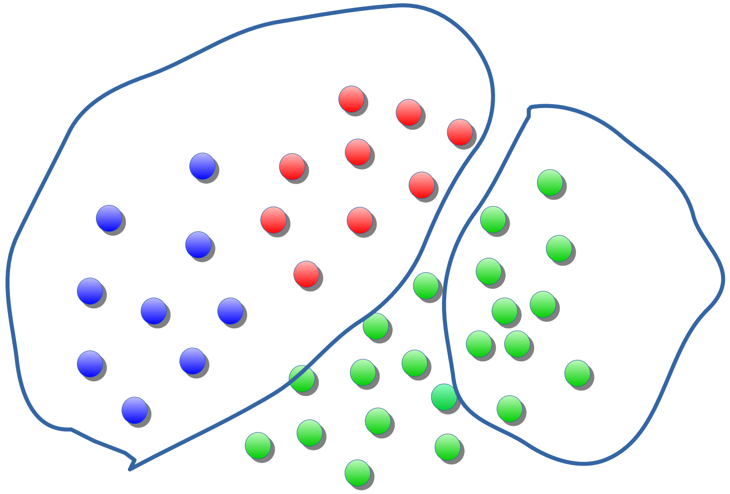
1) Constraints

- Minimal/maximal group size
- Constraints saying two points « have to », « mustn't » be in the same group



Can be done as active learning

2) Assign points by hand



3) Manually merge/split groups

4) Give more importance to certain descripteurs

- Of data points

The expert in reinforcement learning

1) Behavioral cloning
   - Learning to control a drone
   - Learning to drive a car

2)Modify/add rewards
   - Identify sub-goals

3) Action advice
   - Help with selecting an action (instead of completely random)

4) Modify the importance of descripteurs
   - Of states

The expert in reinforcement learning

1) Behavioral cloning

- Learning to control a drone
- Learning to drive a car

A human performs the actions, the recording's given to the algorithm

2)Modify/add rewards

- Identify sub-goals

3) Action advice

- Help with selecting an action (instead of completely random)

4) Modify the importance of descripteurs

- Of states

The expert in reinforcement learning

1) Behavioral cloning

- Learning to control a drone
- Learning to drive a car

A human performs the actions, the recording's given to the algorithm

2)Modify/add rewards

- Identify sub-goals

Example : a robot is supposed to fill a glass, we add a reward for grabbing the bottle

3) Action advice

- Help with selecting an action (instead of completely random)

4) Modify the importance of descripteurs

- Of states

(Interactive) data mining

What’s a pattern?

Can’t be done (well) with NNs

- Can be expressed in a symbolic language

What's a pattern?

Can't be done (well) with NNs

- Can be expressed in a symbolic language

What’s a pattern?

Can’t be done (well) with NNs

- Can be expressed in a symbolic language

What’s a pattern?

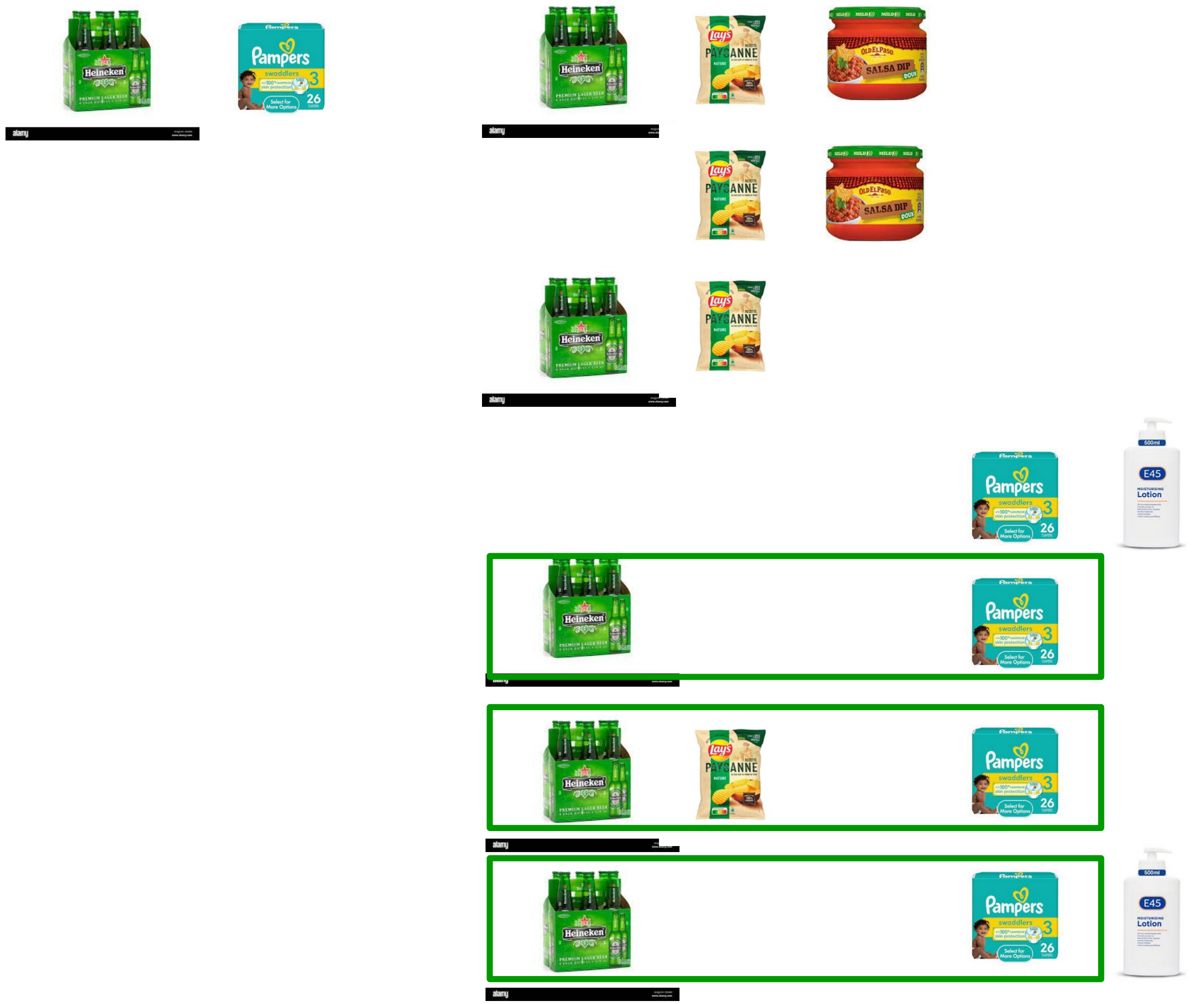
Can’t be done (well) with NNs

- Can be expressed in a symbolic language
- One can find occurrences in the data

What’s a pattern?

Can’t be done (well) with NNs

- Can be expressed in a symbolic language
- One can find occurrences in the data



What's a pattern?

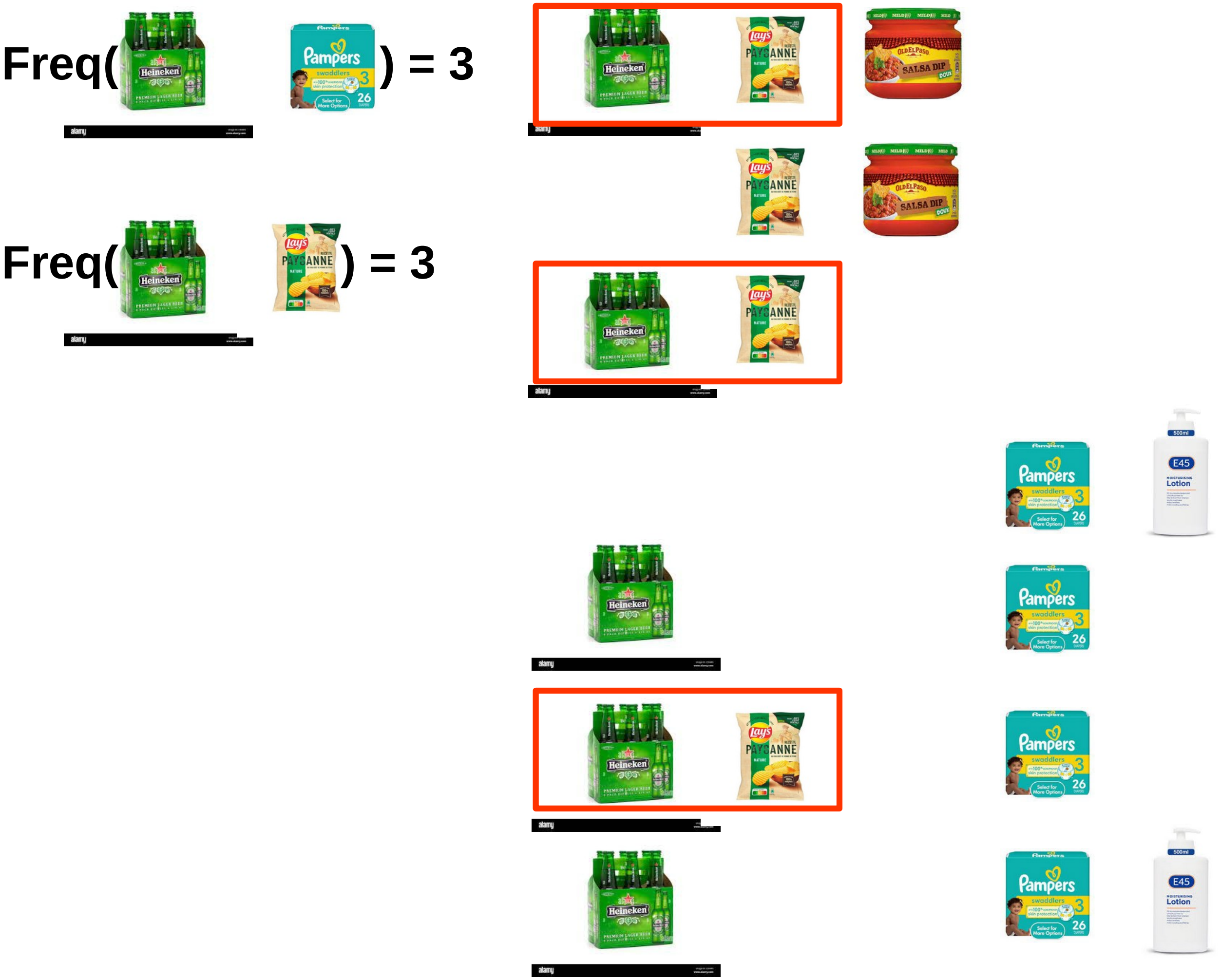
Can't be done (well) with NNs

- Can be expressed in a symbolic language
- One can find occurrences in the data

What’s a pattern?

Can’t be done (well) with NNs

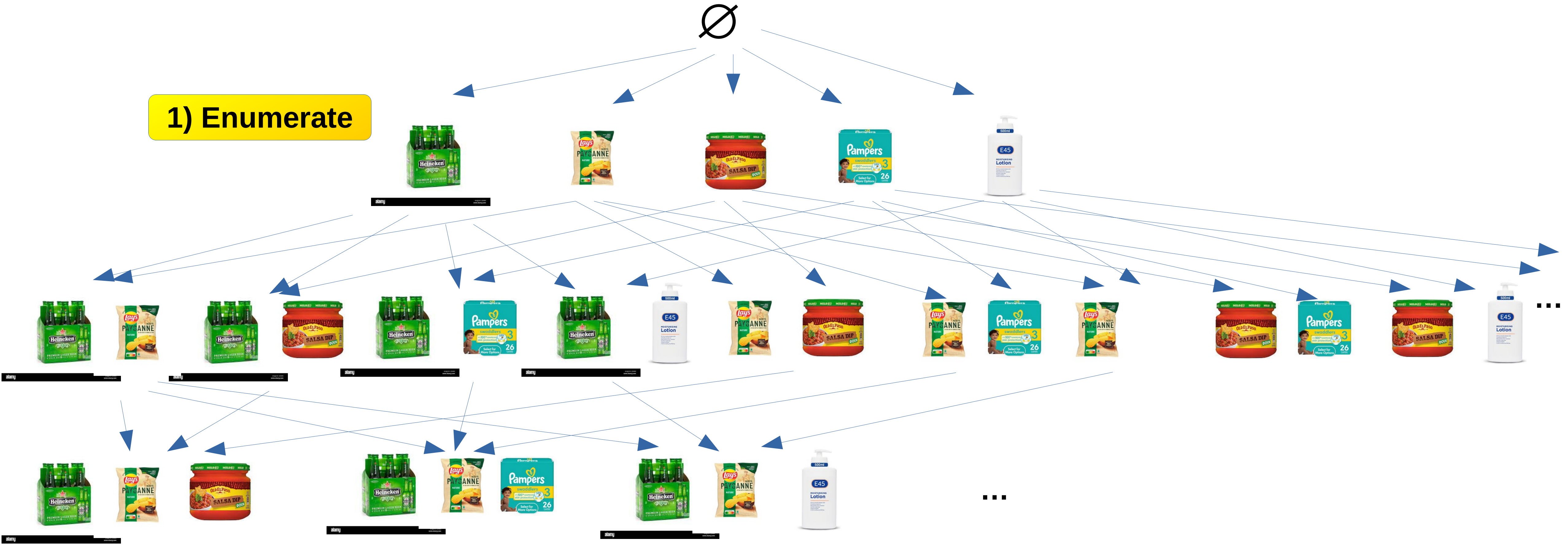
- Can be expressed in a symbolic language
- One can find occurrences in the data
- One can count the number of occurrences



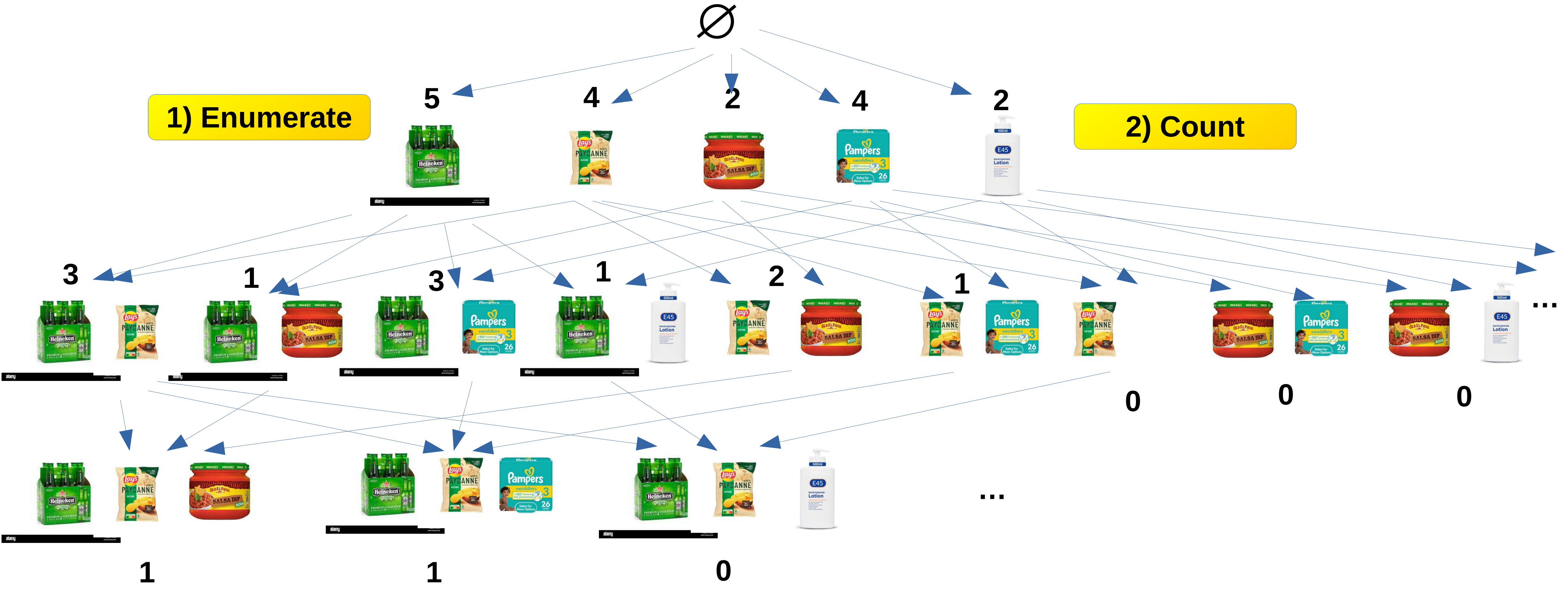
...and how to find them?

GREYC
Laboratoire de recherche en sciences du numérique

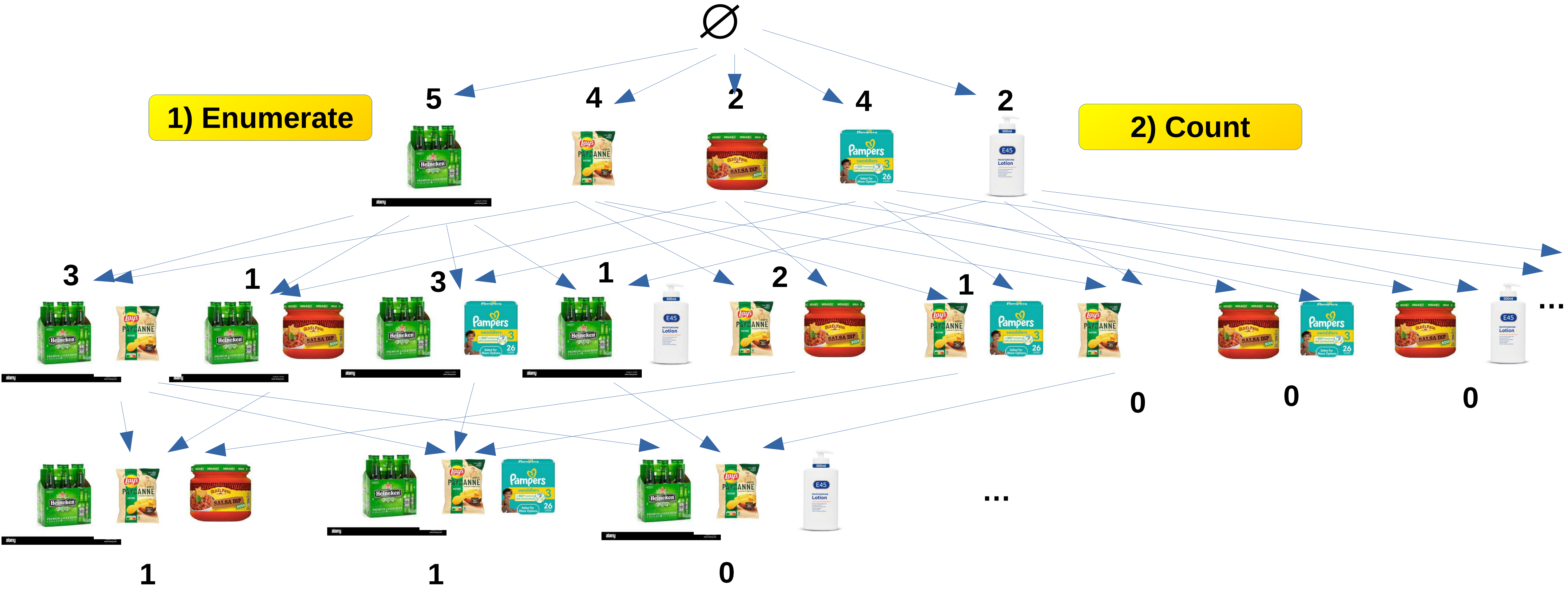
1) Enumerate



…and how to find them?

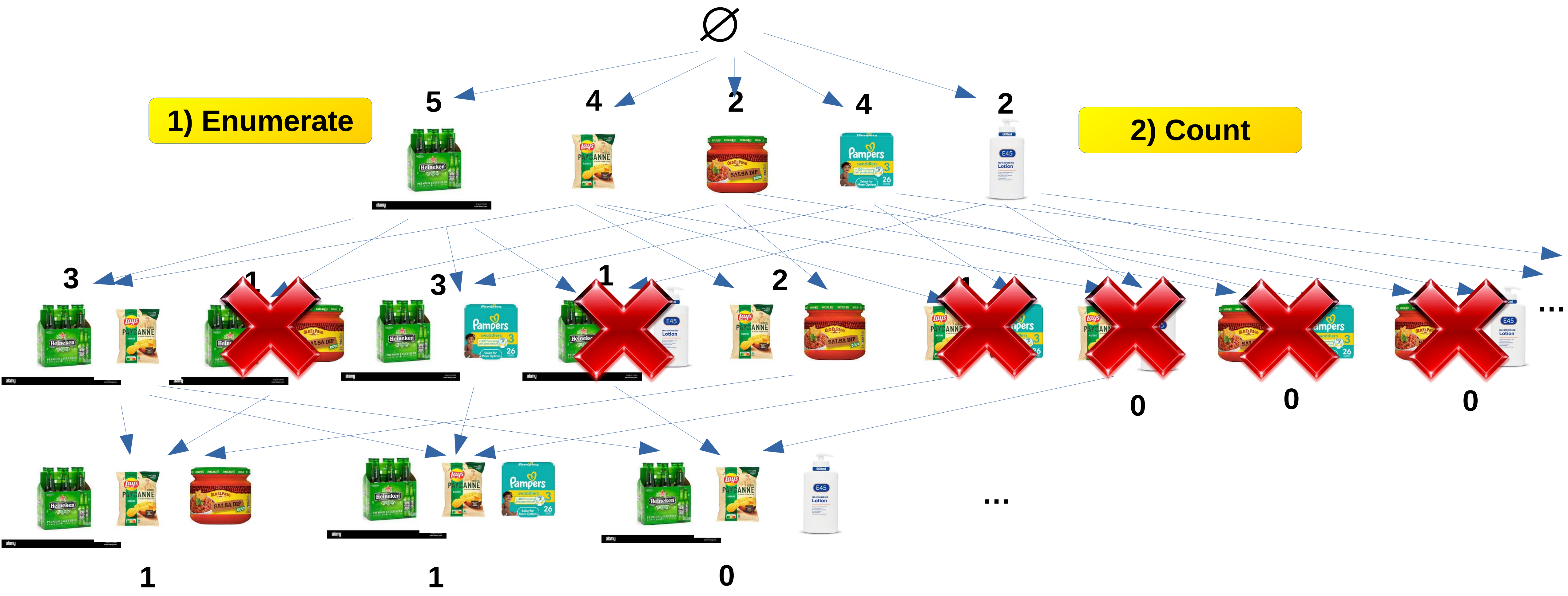


…and how to find them?



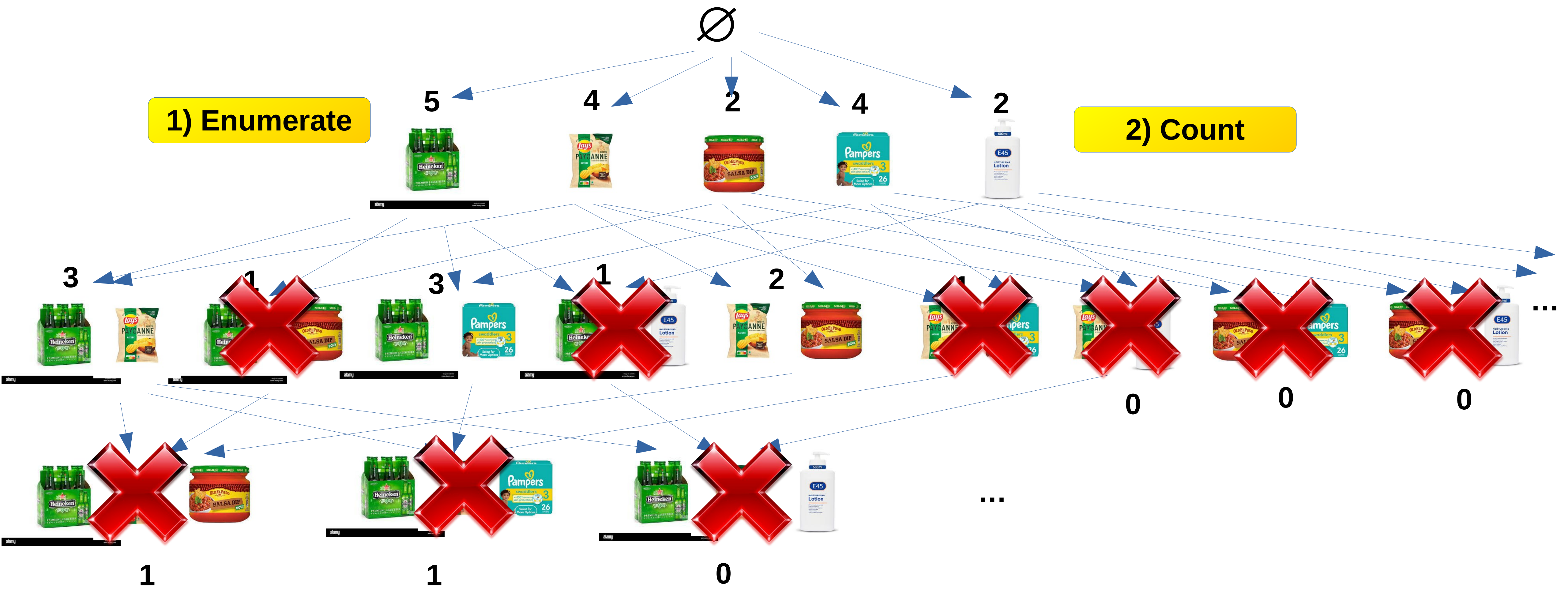
3) Constraint: only interested in patterns with freq ≥ 2

…and how to find them?



3) Constraint: only interested in patterns with freq ≥ 2

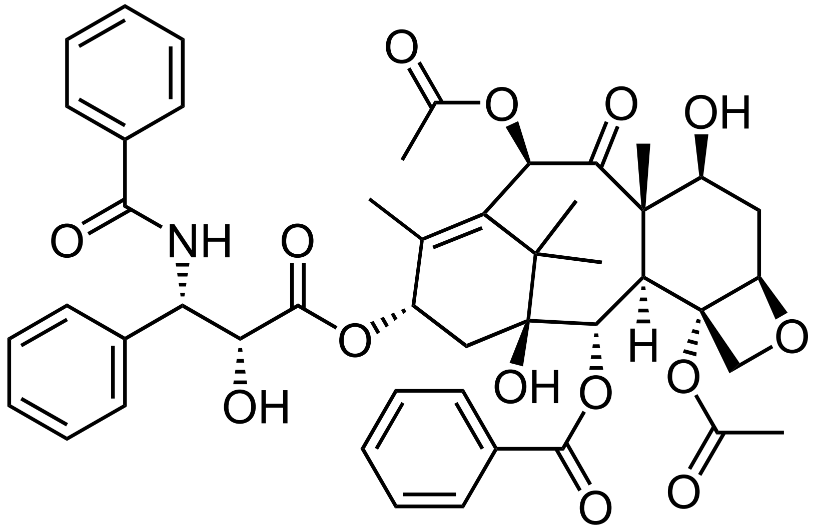
…and how to find them?



3) Constraint: only interested in patterns with freq ≥ 2

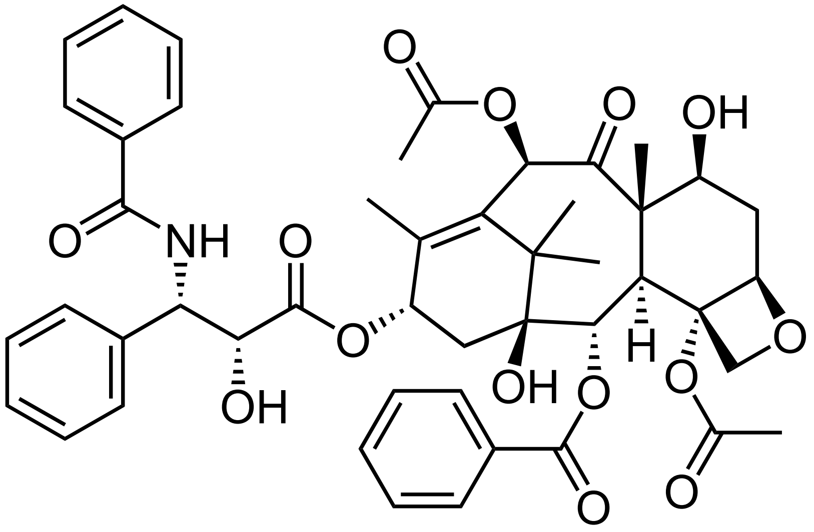
Another pattern language – graphs

- Graph :
  - Vertices ↔ entities
  - Edges ↔ relations
- Molecules can be represented as graphs:
  - Vertices : atoms
  - Edges : chemical bonds
- Interested in finding precursors for drugs
  - Collaboration with CERMN



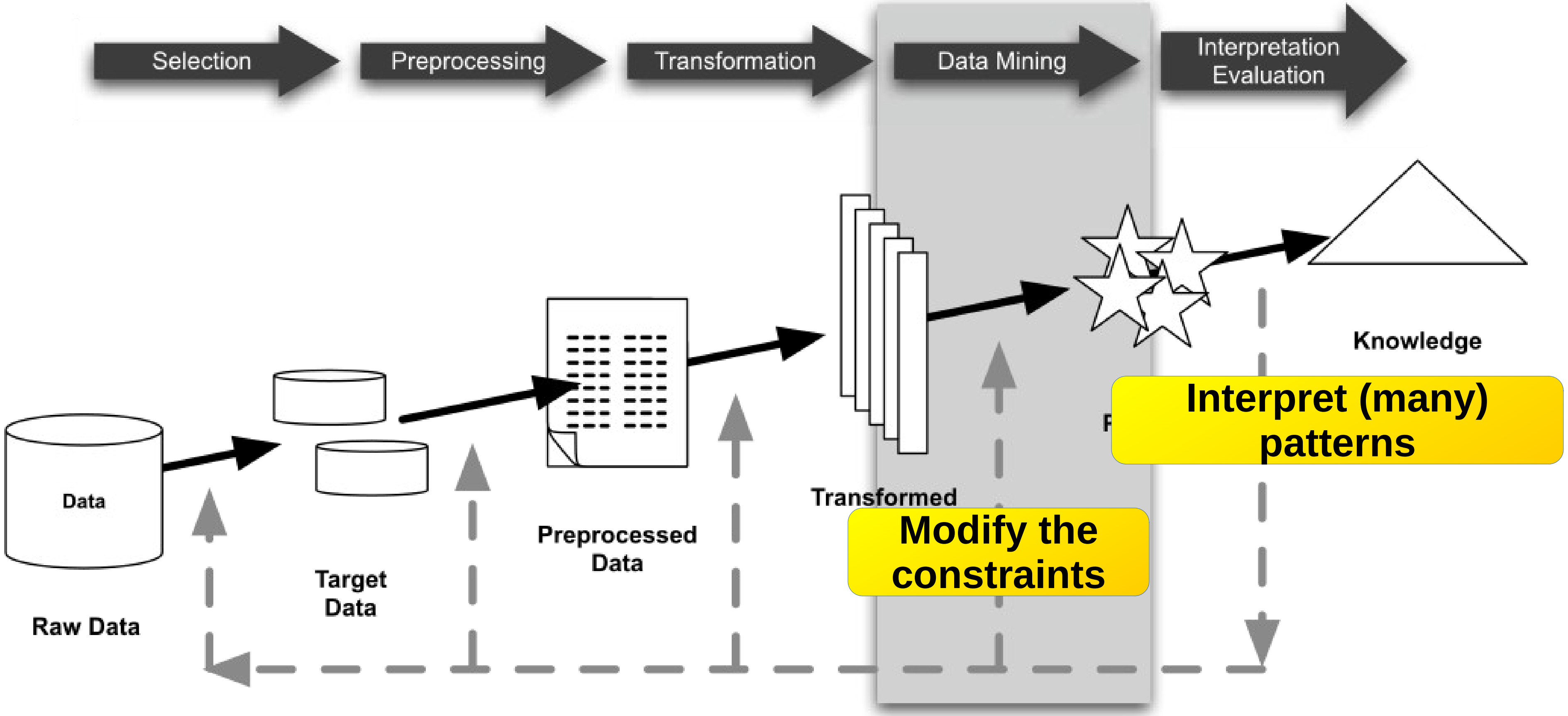
Another pattern language – graphs

- Graph :
  - Vertices ↔ entities
  - Edges ↔ relations
- Molecules can be represented as graphs:
  - Vertices : atoms
  - Edges : chemical bonds
- Interested in finding precursors for drugs
  - Collaboration with CERMN

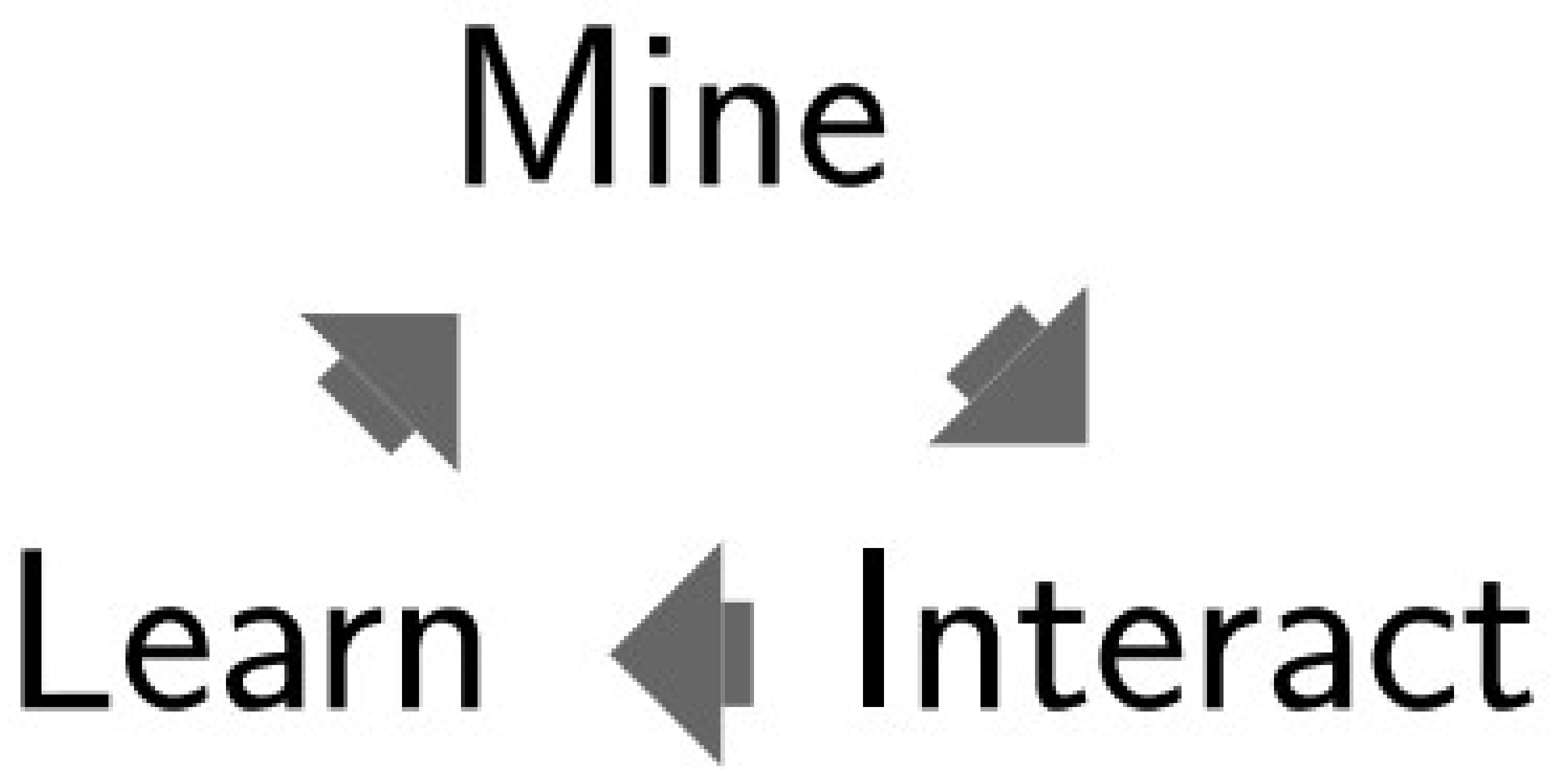


Structured patterns → many more possible patterns

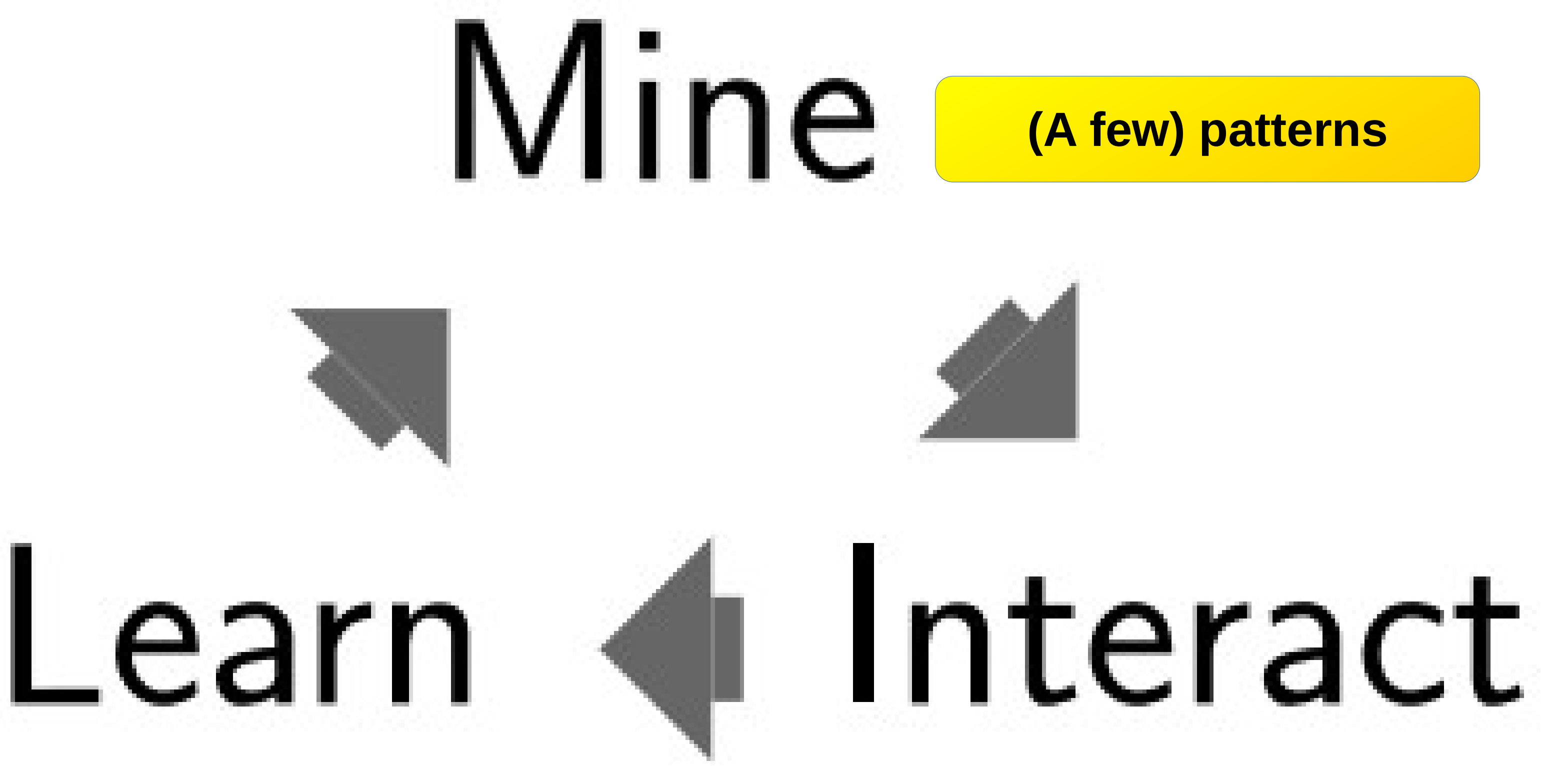
The Knowledge Discovery from Databases process



Interactive data mining

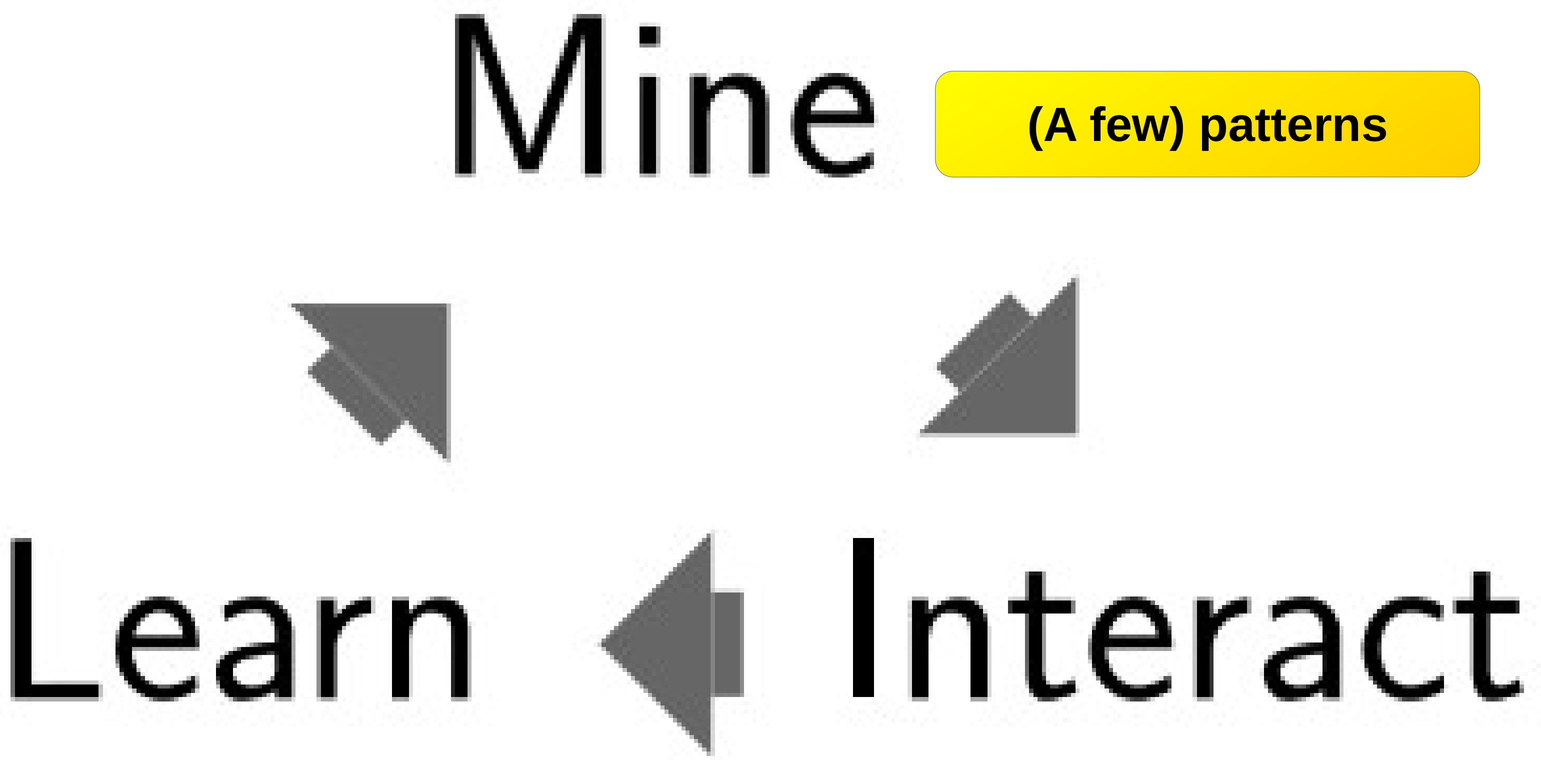


Interactive data mining



(A few) patterns

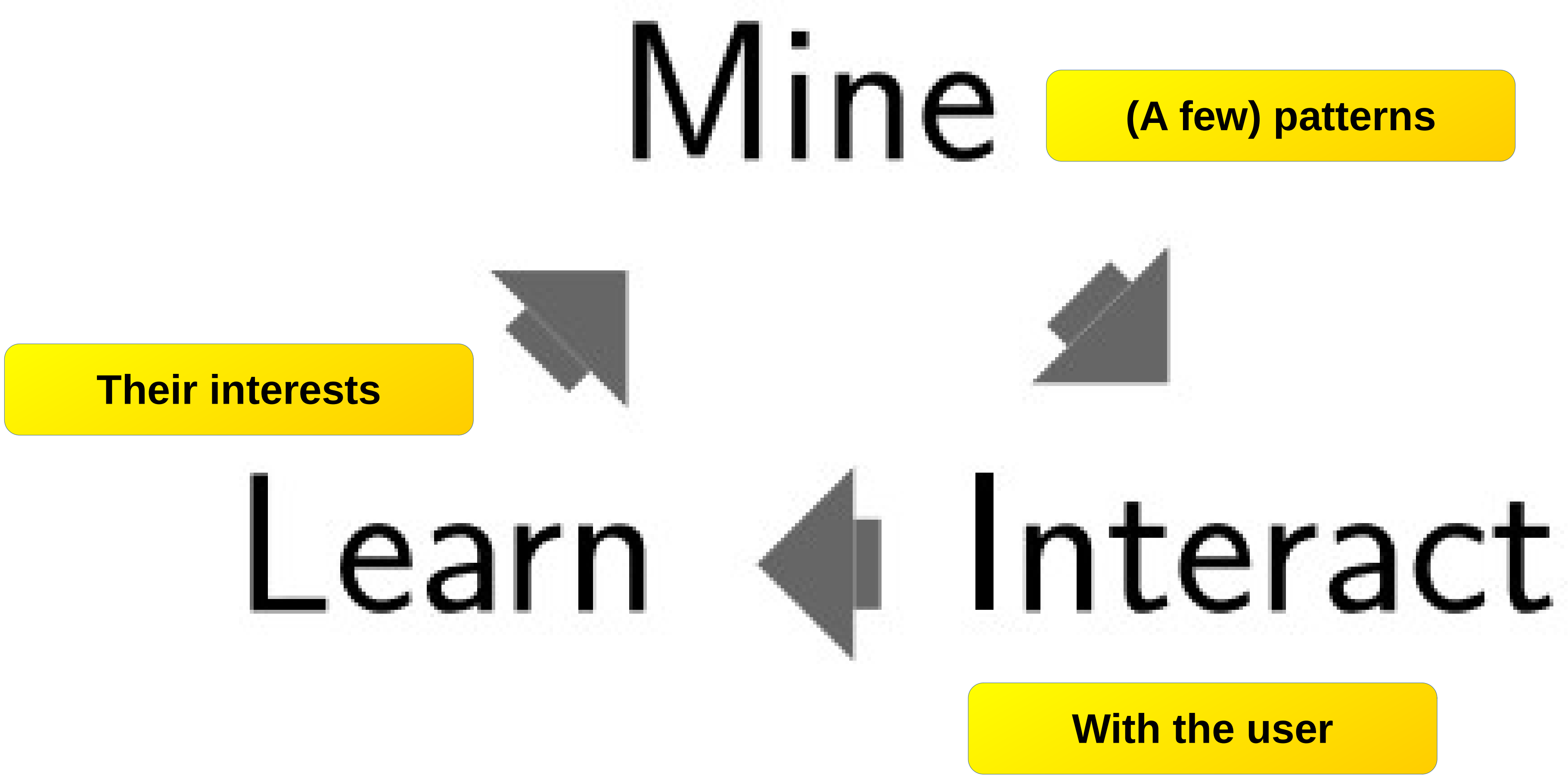
Interactive data mining



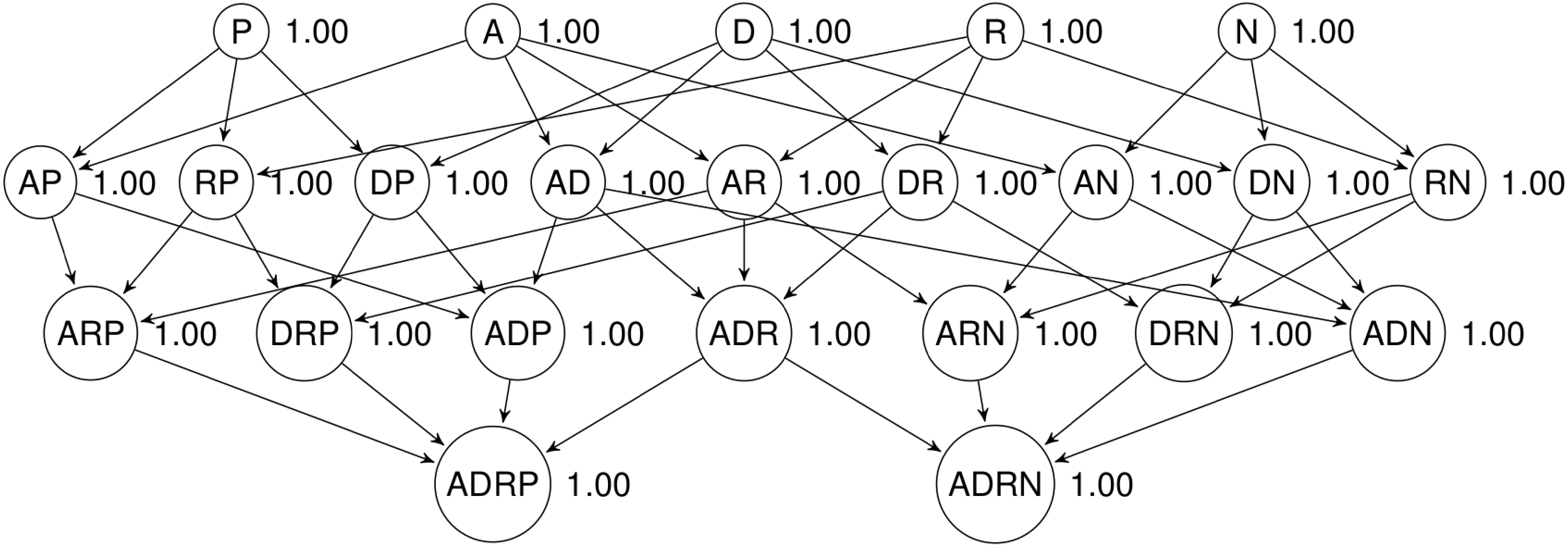
(A few) patterns

With the user

Interactive data mining

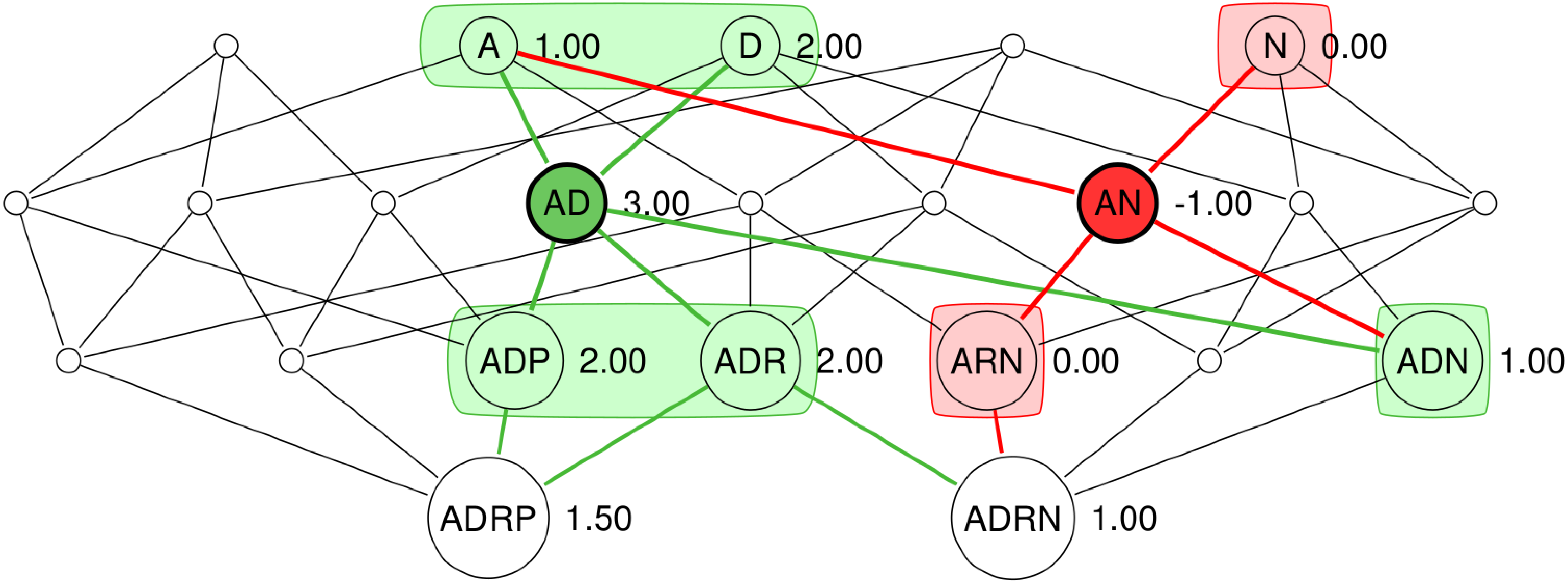


Interactive mining – guide the expert during pattern exploration/interpretation



Before interaction with the expert : all patterns have the same weight

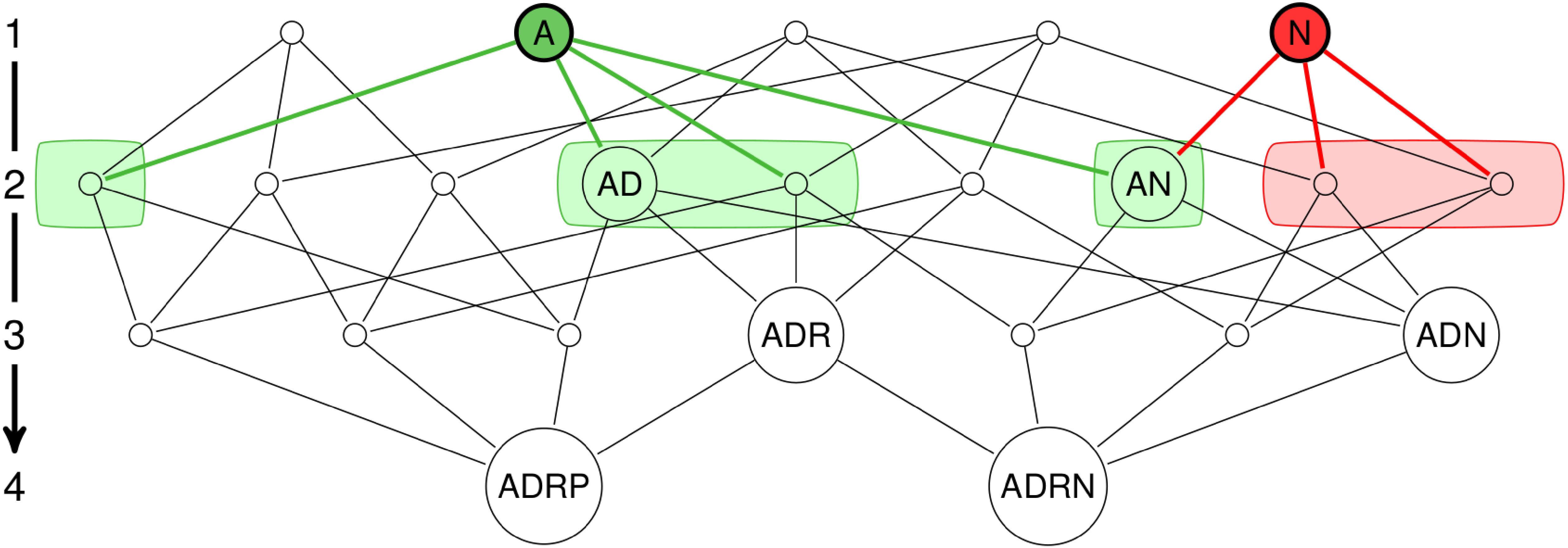
Interactive mining – guide the expert during pattern exploration/interpretation



Positive interaction (the expert finds the pattern interesting) : weight increases
Negative interaction (the expert doesn't find the pattern interesting or rejects it) : weight decreases
→ effect is spread through the network

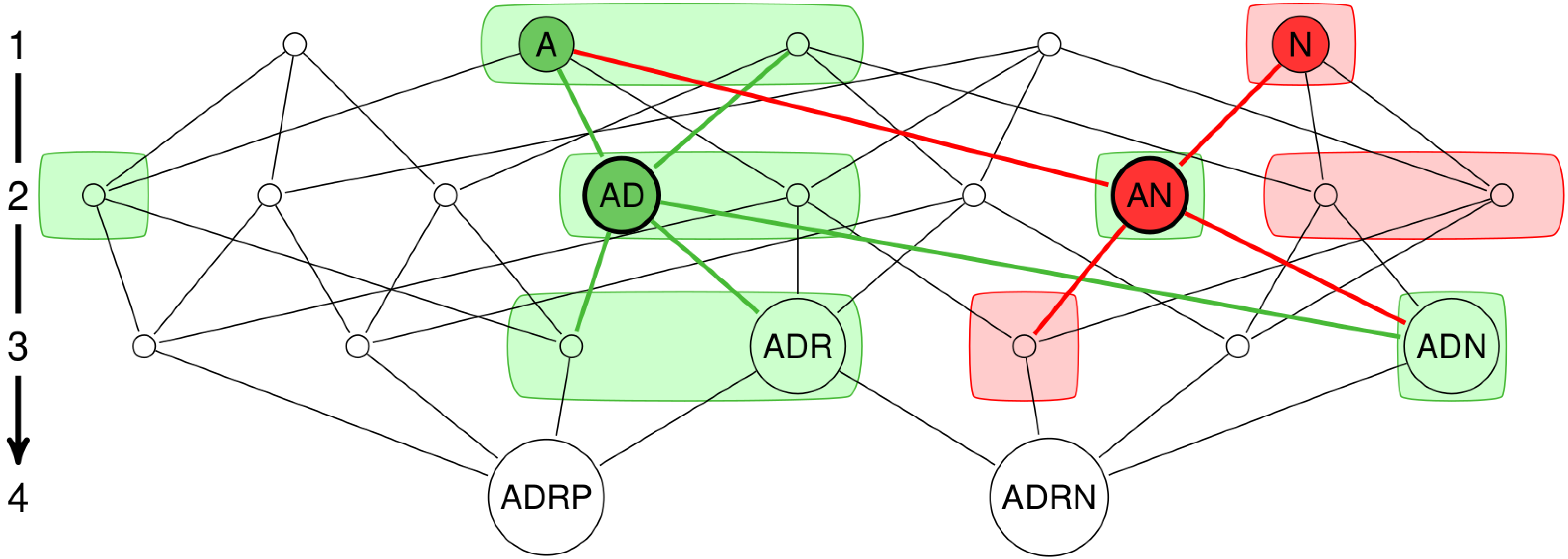
Interactive mining – guide the expert during pattern exploration/interpretation

We start with the least complex patterns

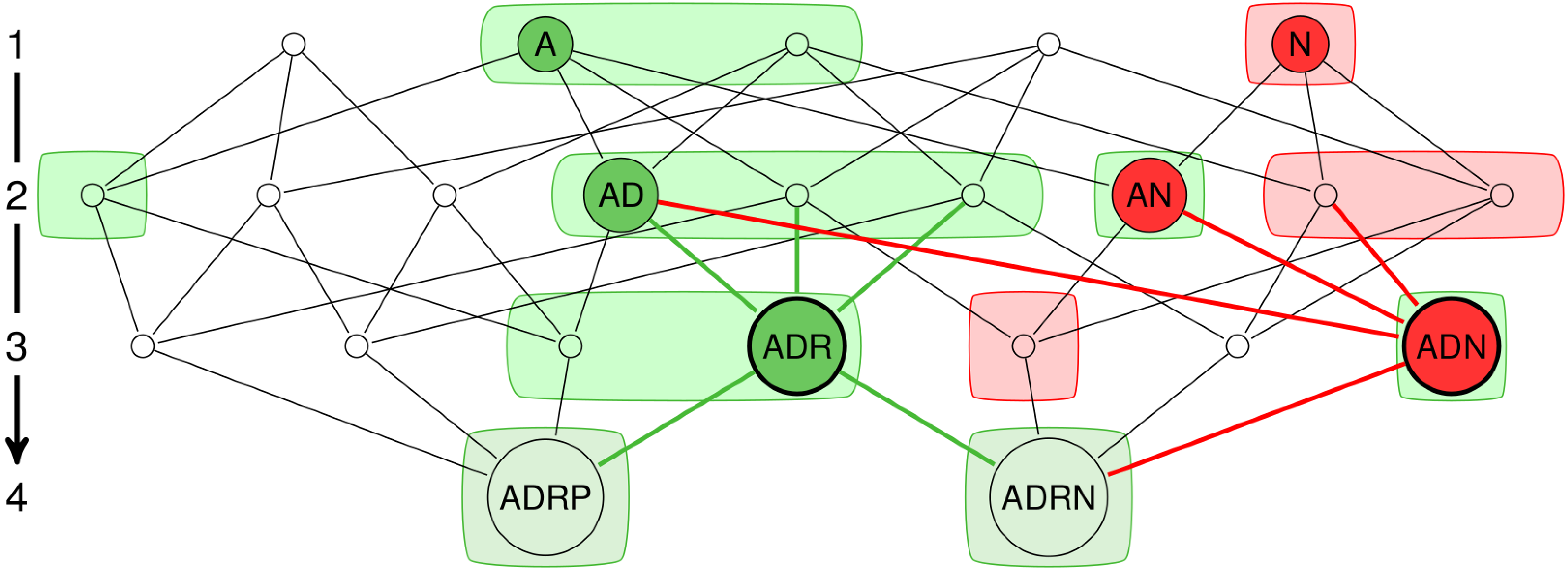


Interactive mining – guide the expert during pattern exploration/interpretation

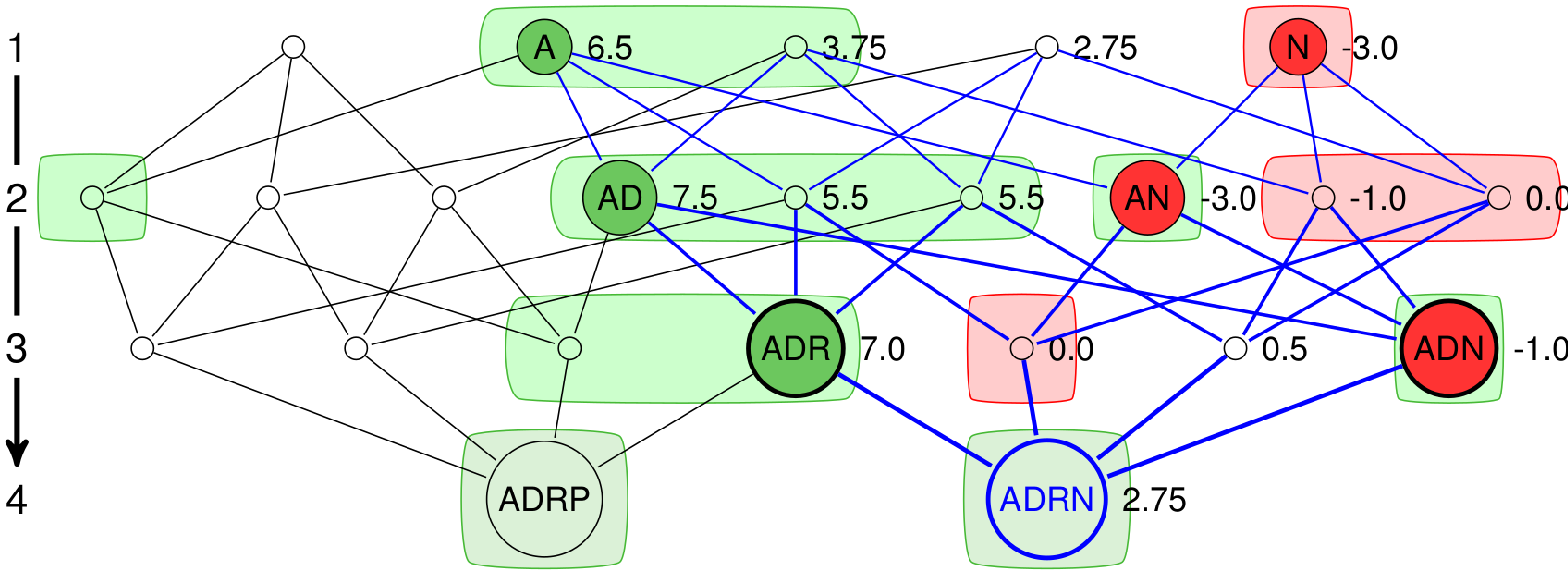
We don’t propose patterns from excluded regions



Interactive mining – guide the expert during pattern exploration/interpretation

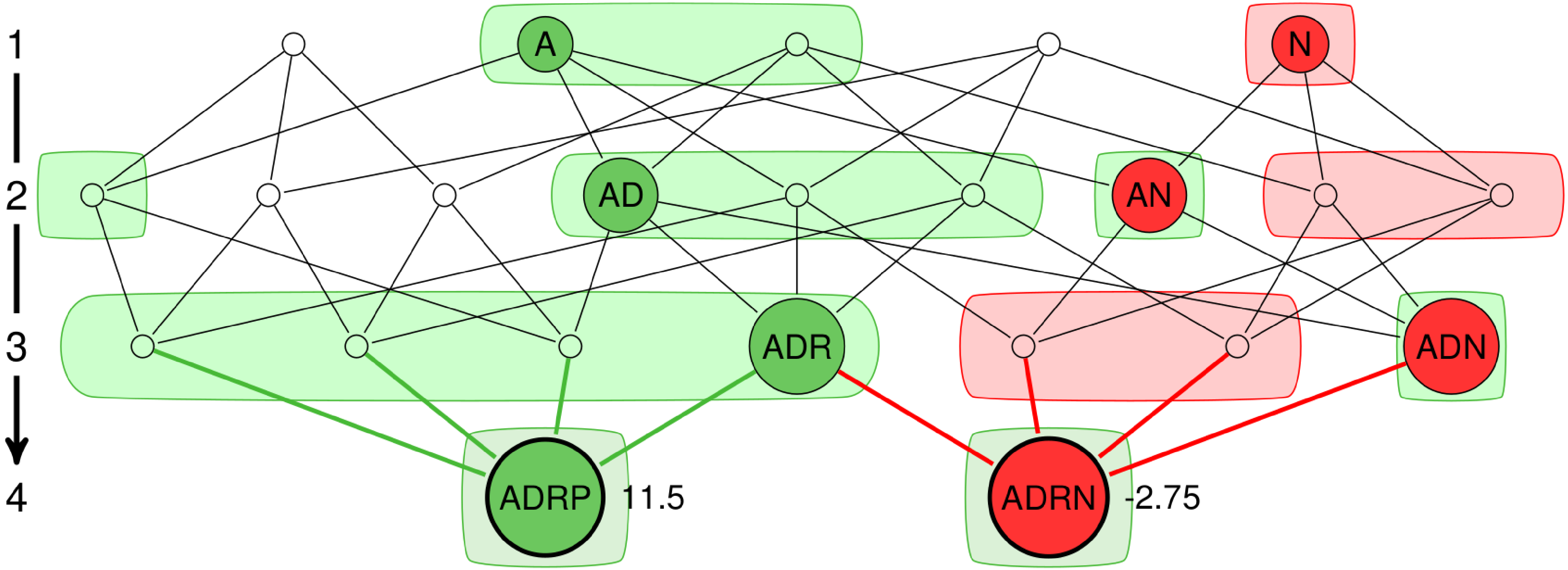


Interactive mining – guide the expert during pattern exploration/interpretation



A pattern whose relatives include both positives and negatives should be explored further

Fouille interactive – guider l'expert pendant l'interprétation



When we're « at the bottom » (the most specific patterns), we roll back up

Questions ?