[AI Theory&App] Introduction to Artificial Intelligence

Instructor: Hao-Shang Ma

Department Of Computer Science And Information Engineering, NTUST

Artificial Intelligence?

- Artificial intelligence is the simulation of human intelligence processes by machines
- Specific applications of AI include expert systems, natural language processing, speech recognition and computer vision

Artificial Intelligence?

- Artificial Intelligence
- Big Data
- Statistics
- Data Mining
- Machine Learning
- Deep Learning

Big Data

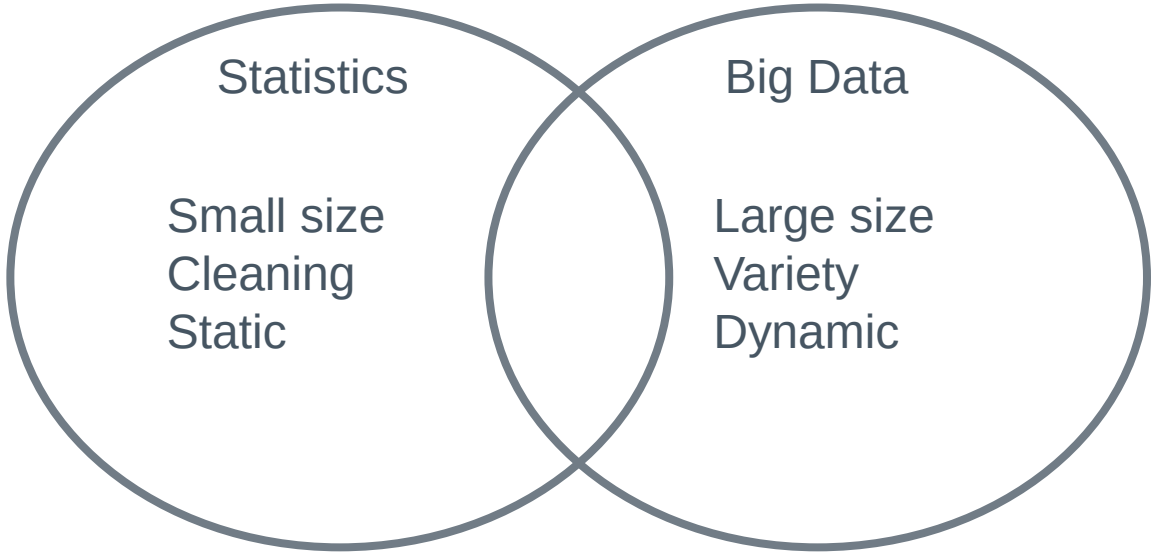
- Big dataset property
  - Volume
  - Variety
  - Velocity
  - Veracity

- Data Warehouse vs Data Lake

Statistics

- A branch of mathematics and a field of study that deals with collecting, analyzing, interpreting, presenting, and organizing data.

- Statistics vs Big Data



Introduction to Data Mining

What is Data Mining

- **Knowledge Discovery from Data**

- Extraction of interesting, non-trivial, implicit, previously unknown and potentially useful patterns or knowledge from huge amount of data

Data Mining Processes

- Data collection
- Data cleaning
- Data integration

Preprocessing

- Data warehousing
- Data selection

Exploration

- Pattern evaluation
- Knowledge discovery

Data Mining

- Information presentation
- Decision making

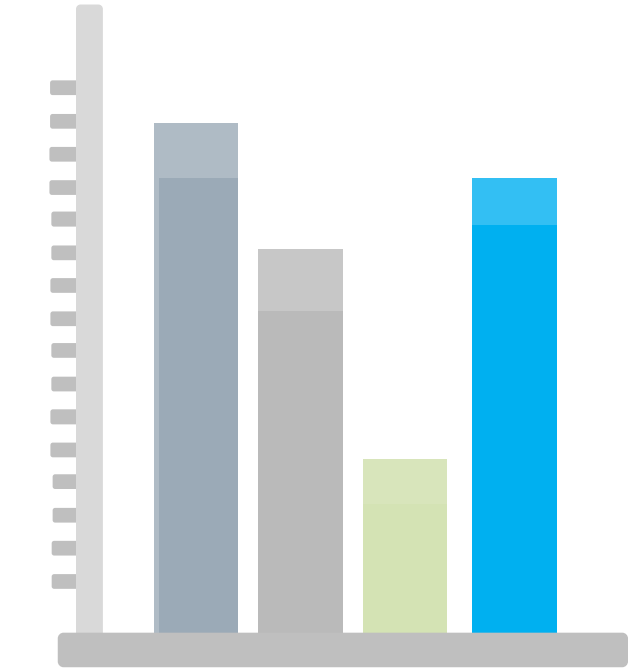
Post processing

Information Presentation

› Data Visualization

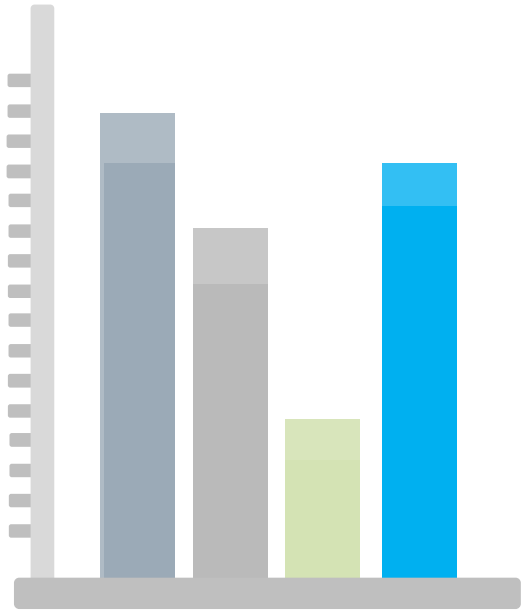
› Make users understand the knowledge and make decision easily

```
C:\Users\Tsai>cat sports_statistics.txt
badminton 230 536
basketball 753 621
beach volleyball 536 123
billiards 125 597
bowling 23 10
boxing 123 46
broad jump 95 146
bullfight 875 14
cricket 45 12
croquet 12 564
cudgel play 217
cycling 125 34
discus throwing 489 12
dodgeball 147 681
exercise bike 13 41
fencing  135 513
figure skating 179 16
football 14 397
German handball 345 135
golf 148 567
gymnastics 169 147
hammer throwing 146 45
handball 132 87
hard-ball tennis 148 56
heel-and-toe walking race 461
high jump
hockey
```



Decision Making

› Based on the evaluation results to decide the policy



General Types of Data Mining

- Association rule mining
- Sequential pattern mining
- Classification
- Clustering
- etc.

Association Rules Mining

› Mining implications between items

Tid	Items
100	A,B,D
200	B,C,E
300	A,D
400	A,B,C,D

min_support = 2
min_conf = 2/3

› Frequent item sets
- {A}, {B}, {C}, {D}, {B,C}, {A,B}, {A,D}, {B,D}, {A,B,D}

› Strong rules
- {A,D} → B(2/3)
- B → C(2/3)
- C → B(2/2)

Apriori Algorithm

Tid	Items
100	A,B,D
200	B,C,E
300	A,D
400	A,B,C,D

Database

C_1

Itemsets
{A}
{B}
{C}
{D}
{E}

$1^{st}\ scan$

itemsets	Sup
{A}	3
{B}	3
{C}	2
{D}	3
{E}	1

L_1

Itemsets	Sup
{A}	3
{B}	3
{C}	2
{D}	3

C_2

Itemsets
{A,B}
{A,C}
{A,D}
{B,C}
{B,D}
{C,D}

$2^{nd}\ scan$

Itemsets	Sup
{A,B}	2
{A,C}	1
{A,D}	3
{B,C}	2
{B,D}	2
{C,D}	1

Infrequent Itemsets pruning

L_2

Itemsets	Sup
{A,B}	2
{A,D}	2
{B,C}	3
{B,D}	2

C_3

Itemsets
{A,B,D}

$3^{rd}\ scan$

Itemsets	Sup
{A,B,D}	2

L_3

Itemsets	Sup
{A,B,D}	2

Candidate Generation

› Step 1: self-joining L_k

Itemsets	Sup
{A,B}	2
{A,D}	2
{B,C}	3
{B,D}	2

$\{A, B\} \cup \{A, D\} = \{A, B, D\}$

$\{B, C\} \cup \{B, D\} = \{B, C, D\}$

Itemsets
{A,B,D}
{B,C,D}

› Step 2: pruning

Itemsets
{A,B,D}
{B,C,D}

{C,D} is infrequent

Itemsets	Sup
{A,B,D}	2

Downward Closure Property

› Any subset of a frequent itemset must be frequent

Itemsets	Sup
{A,B,D}	2

Itemsets	Sup
{A,B}	2
{A,D}	2
{B,C}	3
{B,D}	2

Itemsets	Sup
{A}	3
{B}	3
{C}	2
{D}	3

› If there is any itemset which is infrequent, its superset should not be frequent

Sequential Pattern Mining

› Mining sequential implications between items

SID	Sequences
100	<(1) (2,5) (3) (4)>
200	<(1) (3) (4) (2),(3,5)>
300	<(1,3,5) (4) (2) (3)>
400	<(1) (3) (5)>
500	<(4) (3,5,6)>

min_support = 2

- Some frequent sequential patterns
  - {1,4,2,3},{1,3,5},{4,5}

The Steps of Apriori-like Algorithm

- 1. Sort phase
  - Sort the database
  - Customer id as the primary key and time as the second key
- 2. Litemset phase
  - Count the frequency of the itemset
  - The fraction of customers who bought the itemset

The Steps of Apriori-like Algorithm

- 3. Transformation phase
  - Transform each tx to all litemsets in the form of

C01: <(1,5) (2) (3) (4)>

C02: <(1) (3) (4) (3,5)>

C03: <(1) (2) (3) (4)>

C04: <(1) (3) (5)>

C05: <(4) (5)>

SID	Sequences
100	<(1) (2,5) (3) (4)>
200	<(1) (3) (4) (2),(3,5)>
300	<(1,3,5) (4) (2) (3)>
400	<(1) (3) (5)>
500	<(4) (3,5,6)>

itemset	Count
1	4
2	3
3	4
4	4
5	4
6	1
{2,5}	1
{1,3}	1
{1,5}	1
{3,5}	3
{1,3,5}	1
{5,6}	1
{3,6}	1
{3,5,6}	1

itemset	Count	New ID
1	4	A
2	3	B
3	4	C
4	4	D
5	4	E
{3,5}	3	F

SID	Sequences
100	<(1) (2,5) (3) (4)>
200	<(1) (3) (4) (2),(3,5)>
300	<(1,3,5) (4) (2) (3)>
400	<(1) (3) (5)>
500	<(4) (3,5,6)>

SID	Sequences
100	<(A) (B,E) (C) (D)>
200	<(A) (C) (D) (B),(C,E,F)>
300	<(A,C,E,F) (D) (B) (C)>
400	<(A) (C) (E)>
500	<(D) (B,E,F)>

The Steps of Apriori-like Algorithm

- 4. Mining phase
  - Apriori-like algorithm
- 5. Maximal phase
  - Find the maximum patterns

SID	Sequences
100	<(A) (B,E) (C) (D)>
200	<(A) (C) (D) (B),(C,E,F)>
300	<(A,C,E,F) (D) (B) (C)>
400	<(A) (C) (E)>
500	<(D) (B,E,F)>

3-itemset	Count
A,B,C	3
A,C,D	2
D,B,C	1
A,D,B	2

2-Itemset	count
A,B	3
A,C	4
A,D	3
A,E	3
A,F	2
B,A	0
B,C	3
B,D	1
B,E	1
B,F	1
C,A	0
C,B	2
C,D	3
C,E	2
C,F	0

2-Itemset	count
D,A	0
D,B	3
D,C	2
D,E	2
D,F	2
E,A	0
E,B	1
E,C	2
E,D	2
E,F	0
F,A	0
F,B	1
F,C	1
F,D	1
F,E	0

SID	Sequences
100	<(A) (B,E) (C) (D)>
200	<(A) (C) (D) (B),(C,E,F)>
300	<(A,C,E,F) (D) (B) (C)>
400	<(A) (C) (E)>
500	<(D) (B,E,F)>

2-Itemset	count
A,B	3
A,C	4
A,D	3
A,E	3
B,C	3
C,D	3
D,B	3

3-itemset	Count
A,B,C	3
A,C,D	2
D,B,C	1
A,D,B	2

Itemset	count
A,B	3
A,C	4
A,D	3
A,E	3
B,C	3
C,D	3
D,B	3
A,B,C	3

itemset	Count	New ID
1	4	A
2	3	B
3	4	C
4	4	D
5	4	E
{3,5}	3	F

Frequent patterns: {1,2}, {1,3}, {1,4}, {1,5}, {2,3}, {3,4}, {4,2}, {1,2,3}

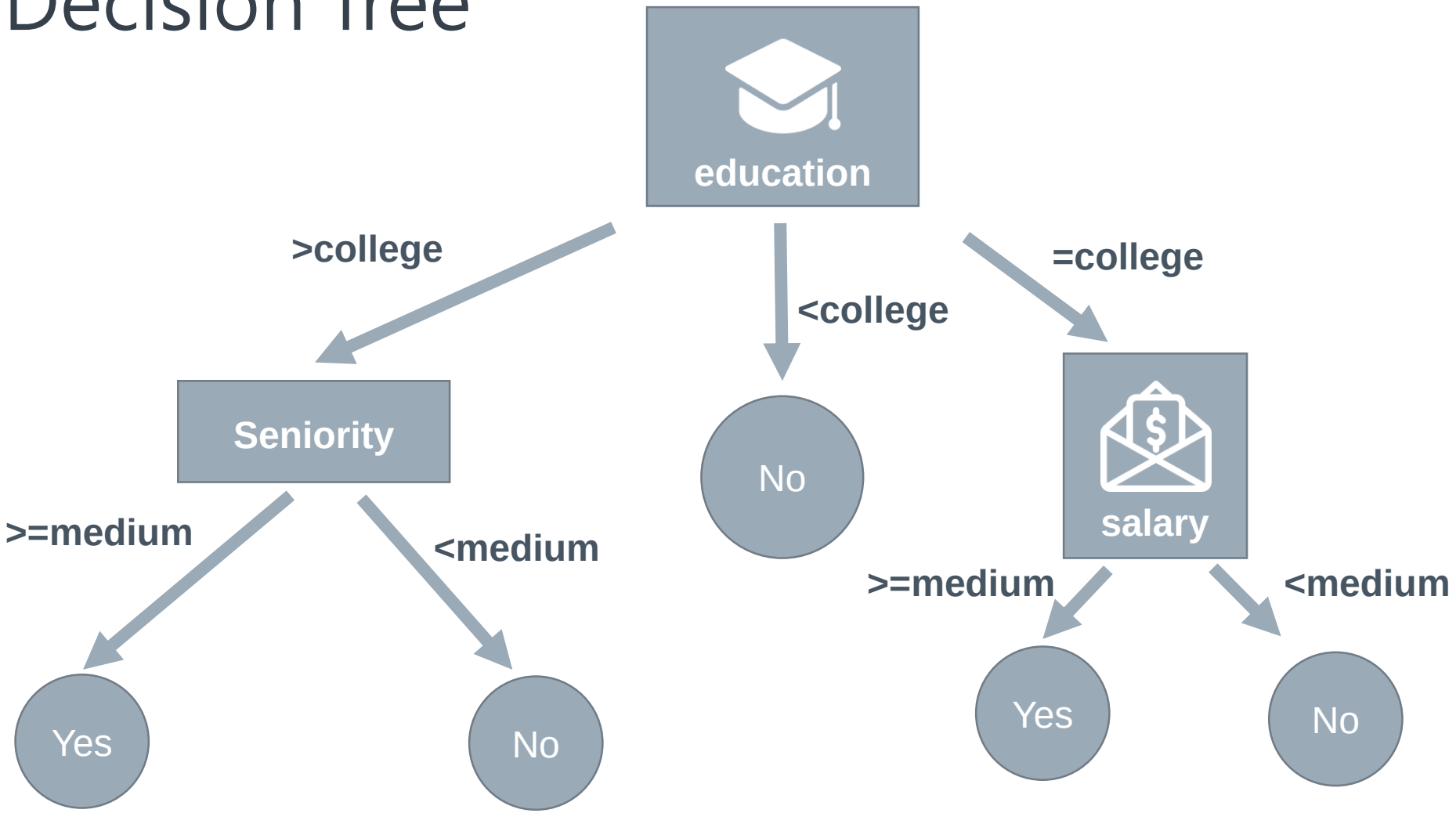
Classification

› Determine the label of data given some observations

Name	Education	Salary	Job	Seniority	Buy car
Allen	Ph.D	High	Engineer	Medium	Yes
Bubble	College	Low	Worker	Medium	No
Chris	Senior high	Medium	Worker	Medium	No
Dustin	Master	Medium	Engineer	Long	Yes
Elan	Ph.D	High	Professor	Long	Yes
Frank	College	High	Sport player	Long	Yes
George	Master	Medium	Engineer	Short	No

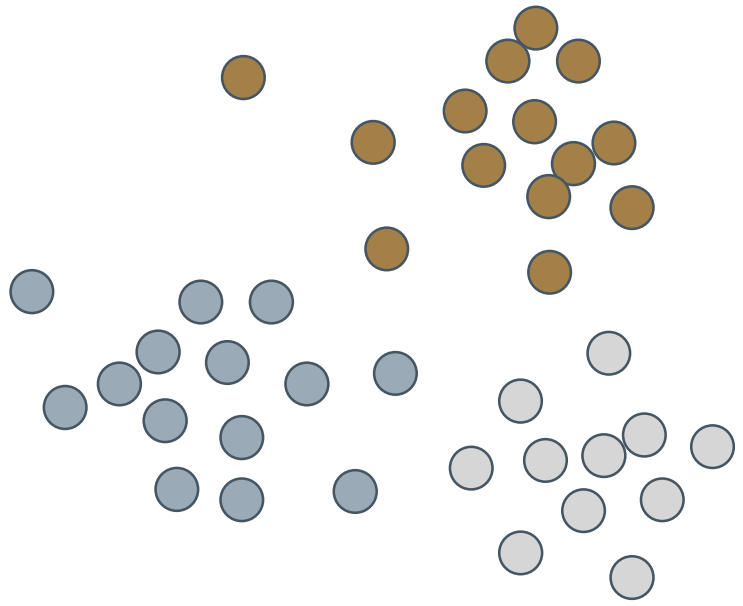
(Education >= college) & (salary >= Medium) => buy car

Decision Tree

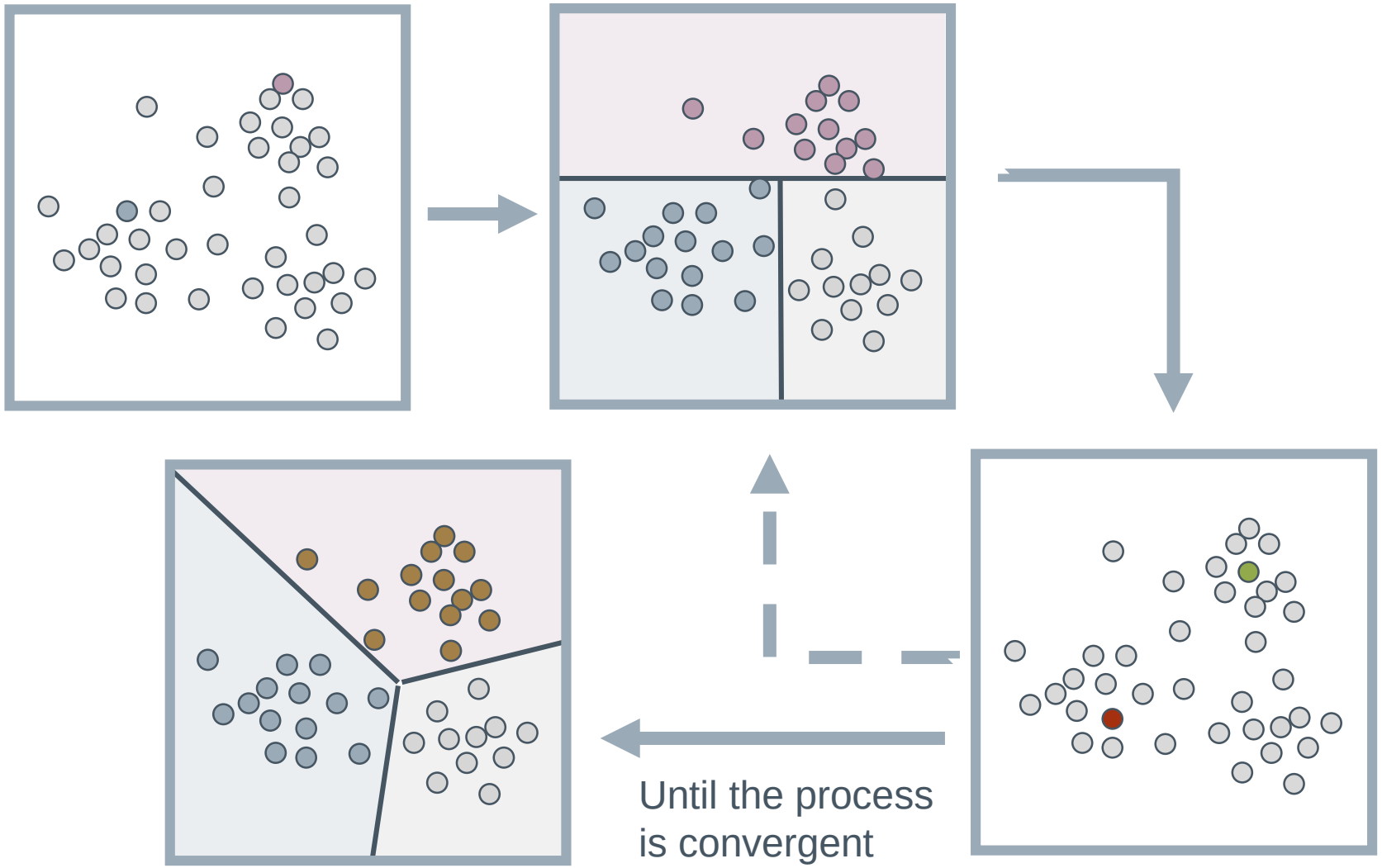


Clustering

› Determine which items should be similar to one another within a group



K-means Algorithm



Introduction to Machine Learning

The Origins of Machine Learning

› How can computers learn to solve problems without being explicitly programmed?

-----Arthur Lee Samuel, 1959

› The answer is: Machine Learning

Arthur Lee Samuel, **Some Studies in Machine Learning Using the Game of Checkers,** *IBM Journal of Research and Development*. 3, 1959

What is Machine Learning

- **Machine learning** is a field of computer science that gives computer systems the ability to "learn," i.e., progressively improve performance on a specific task, with data but without being explicitly programmed.

- Learn knowledges from data

General Types of Machine Learning

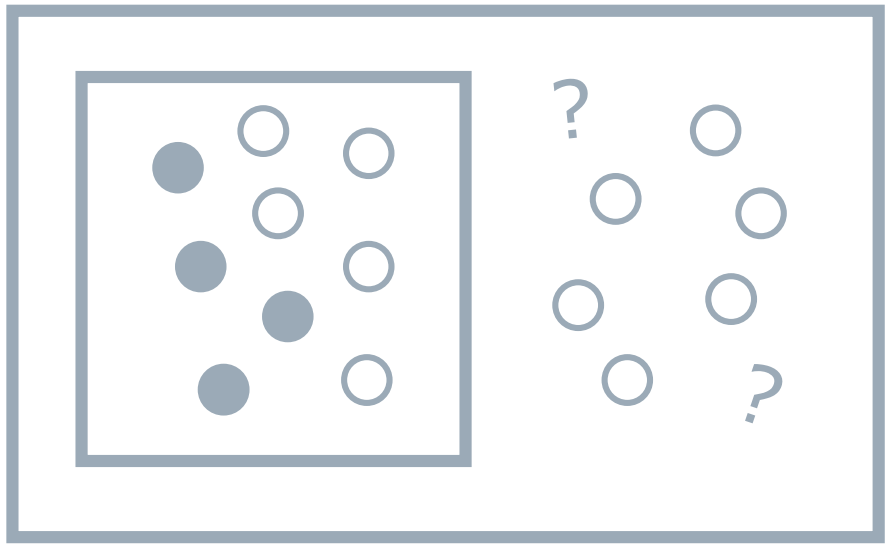
- Supervise Learning
  - Classification problems
- Unsupervised Learning
  - Clustering problems
  - Text mining
- Semi-supervised learning
  - The classification problem with some missing labels
  - The clustering problem with some labels

Some Special Types of Machine Learning

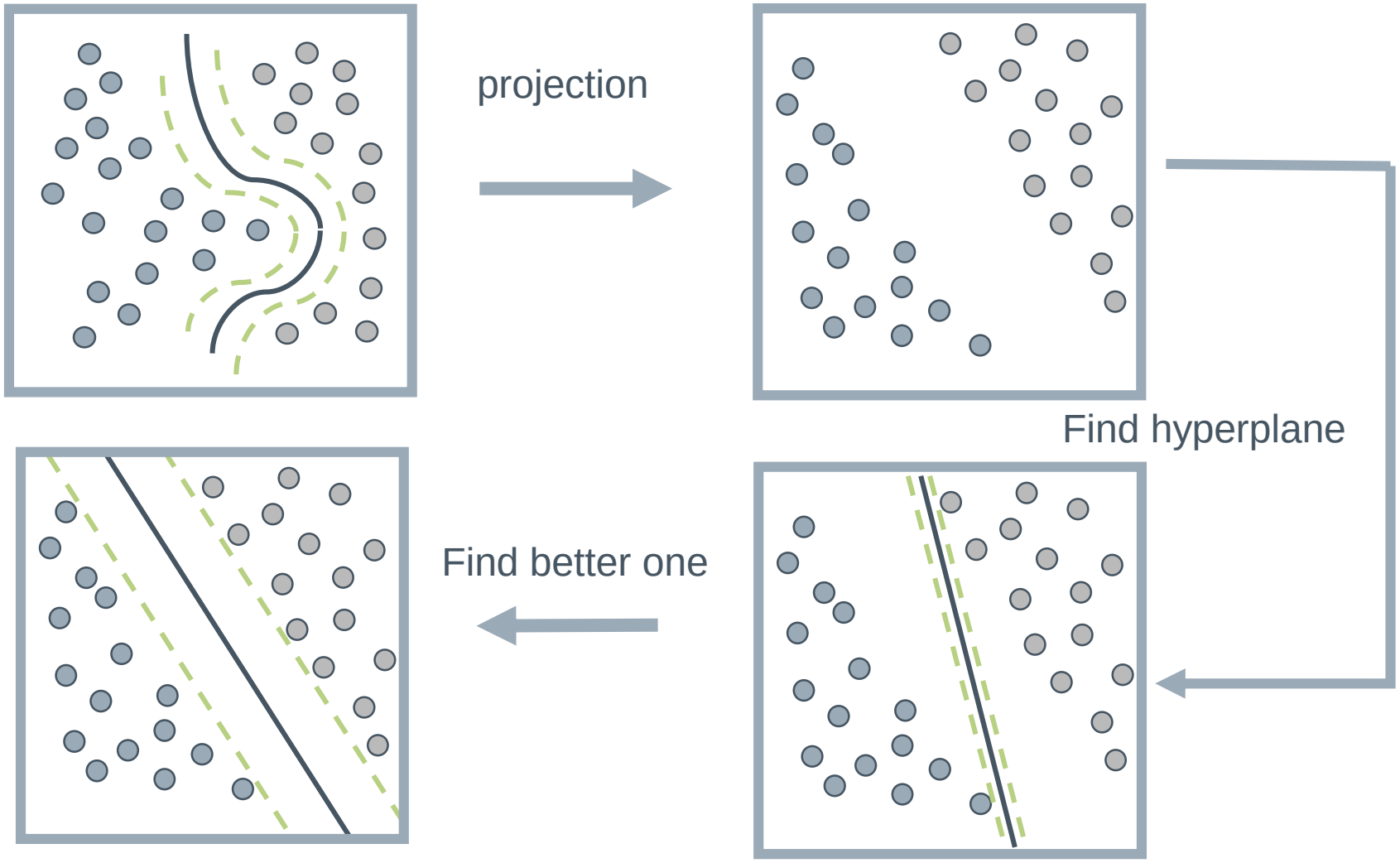
- Reinforcement Learning
  - alphaGo
  - alphaGoZero
- Interactive learning
- Transfer learning

Supervised Learning

› Learning from some labeled data
  –We have the answers of training data

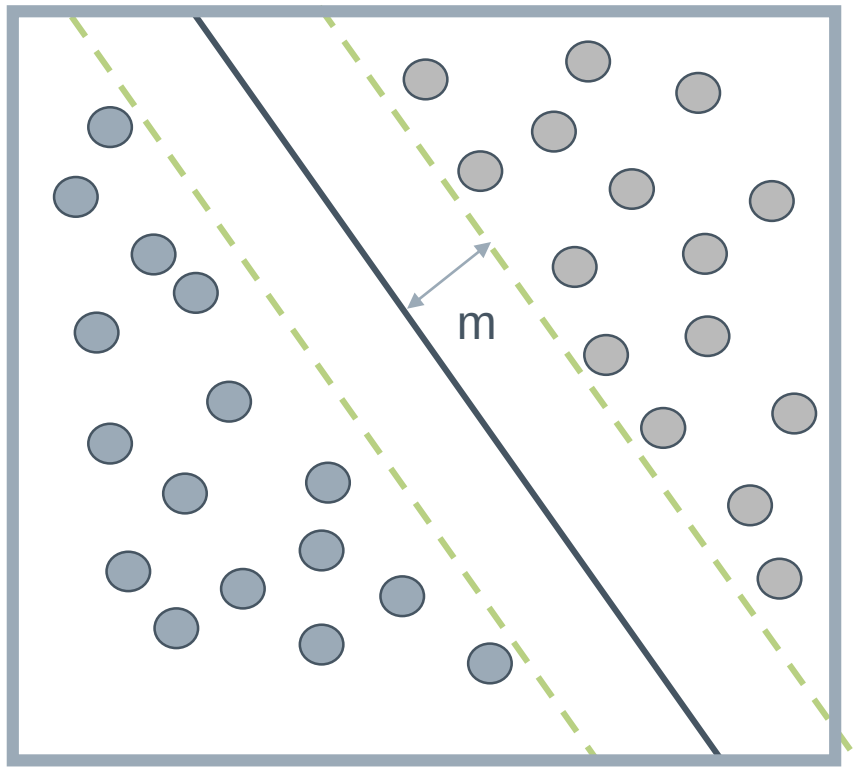


Support Vector Machine



Find The Best Hyperplane

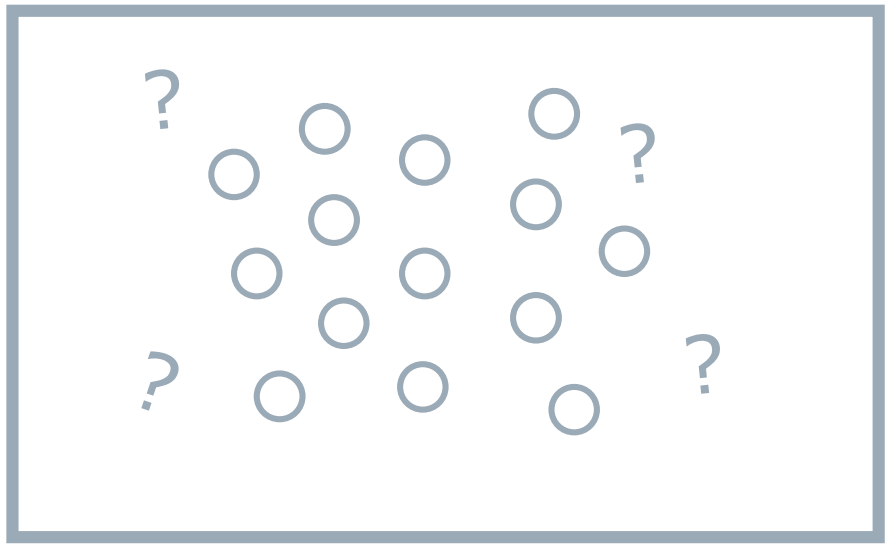
› Find the hyperplane with the largest margin



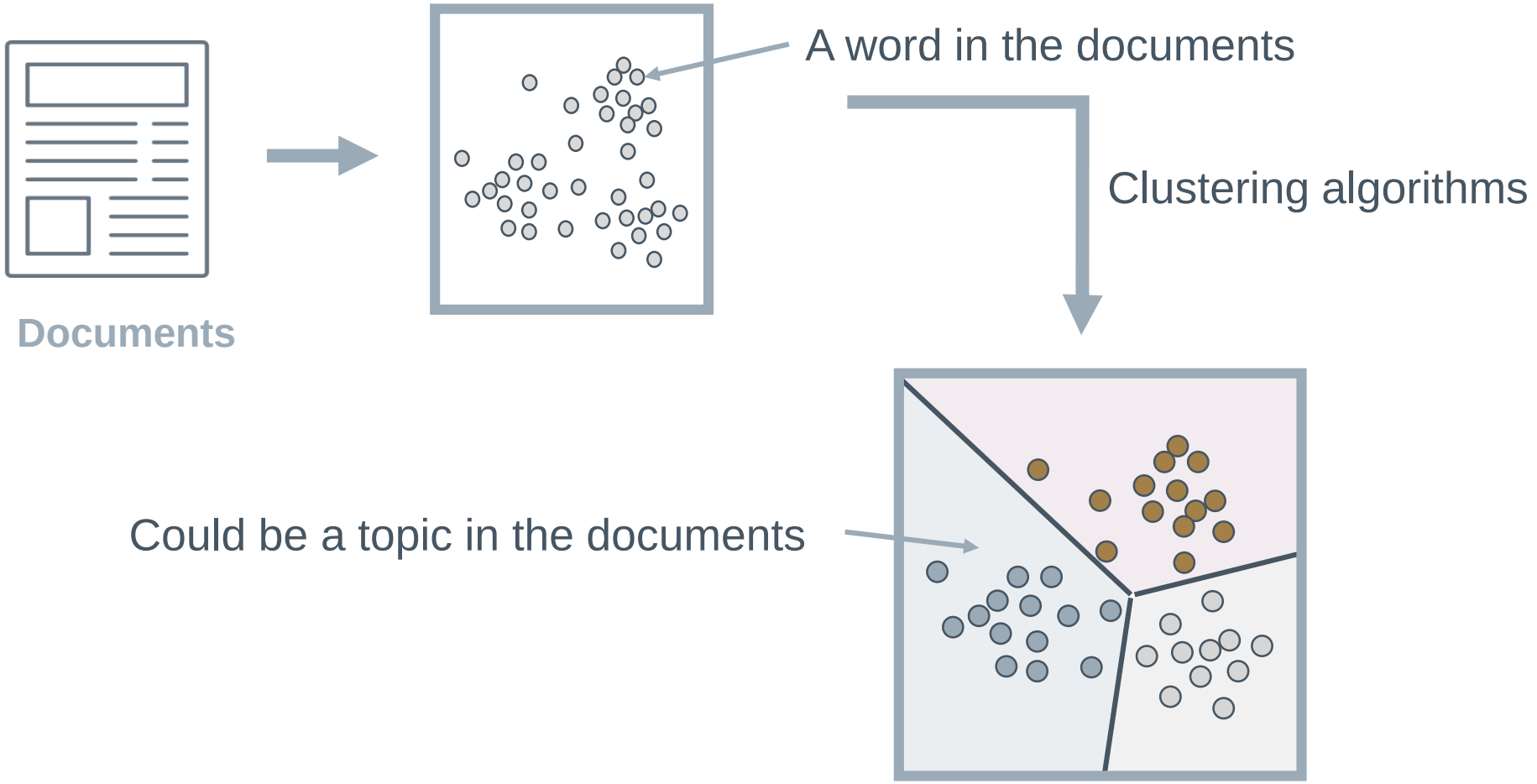
m: margin

Unsupervised Learning

- Learning on the un-labeled data
  - We do not have the answer of training data

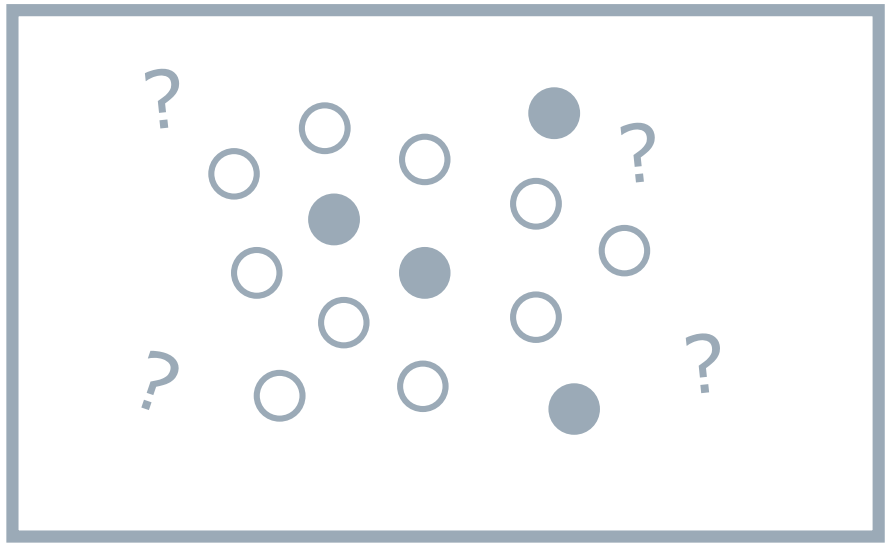


Similar to Clustering



Semi-supervised Learning

› Learning on partly-labeled data
  –We have only some answers for training data

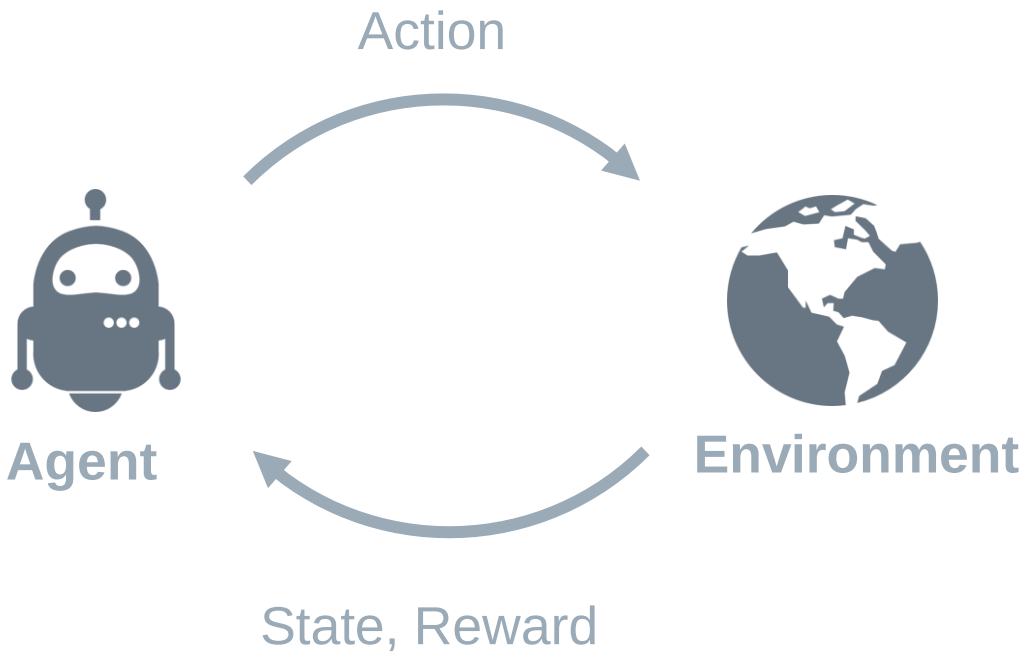


How to Add Labels for Un-labeled Data

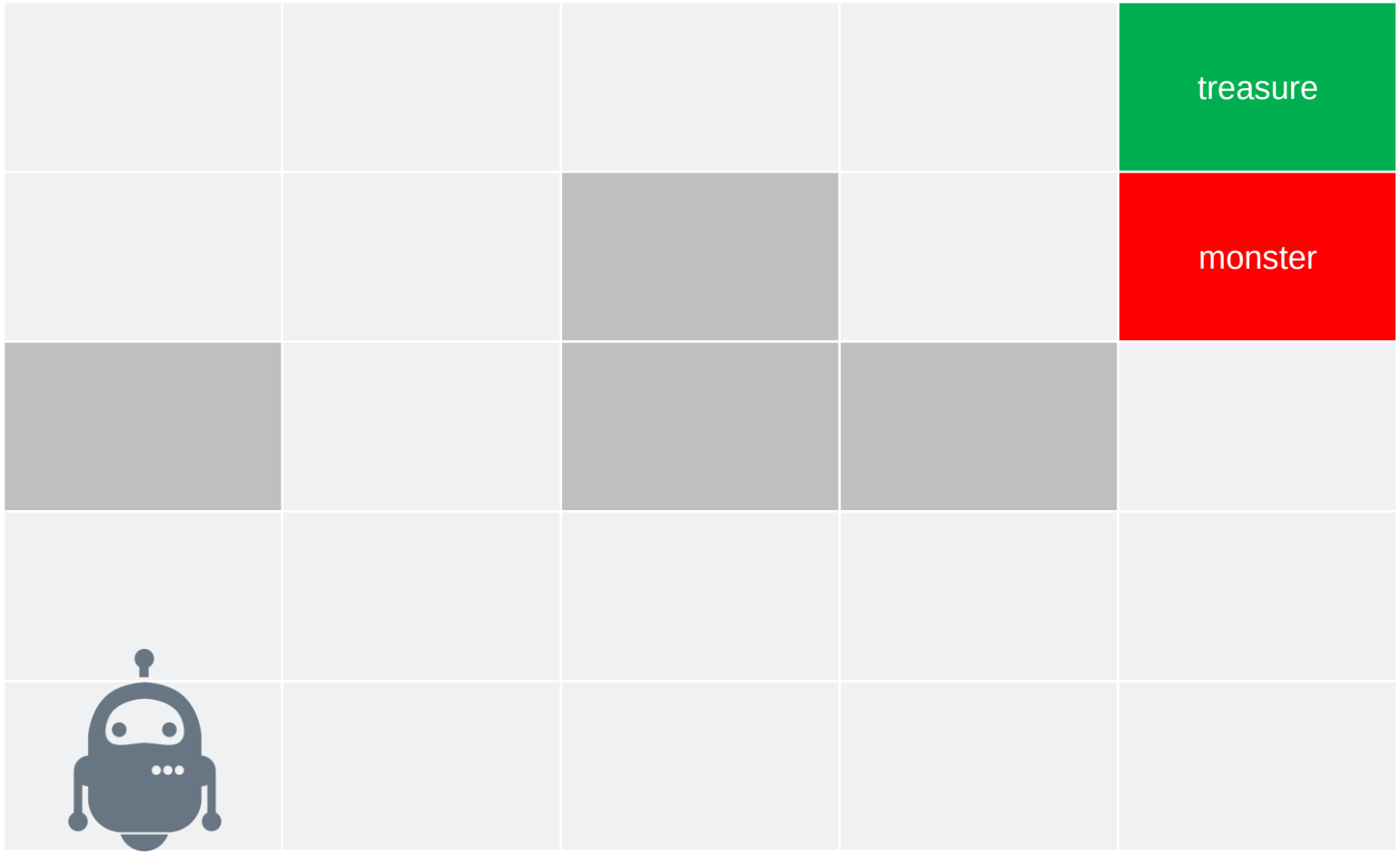
- Continuity assumption
  - The nearby data might have the same labels
- Cluster assumption
  - The data in the same cluster might have the same labels
- Manifold assumption
  - The data lie approximately on a manifold of much lower dimension than the input space

Reinforcement Learning

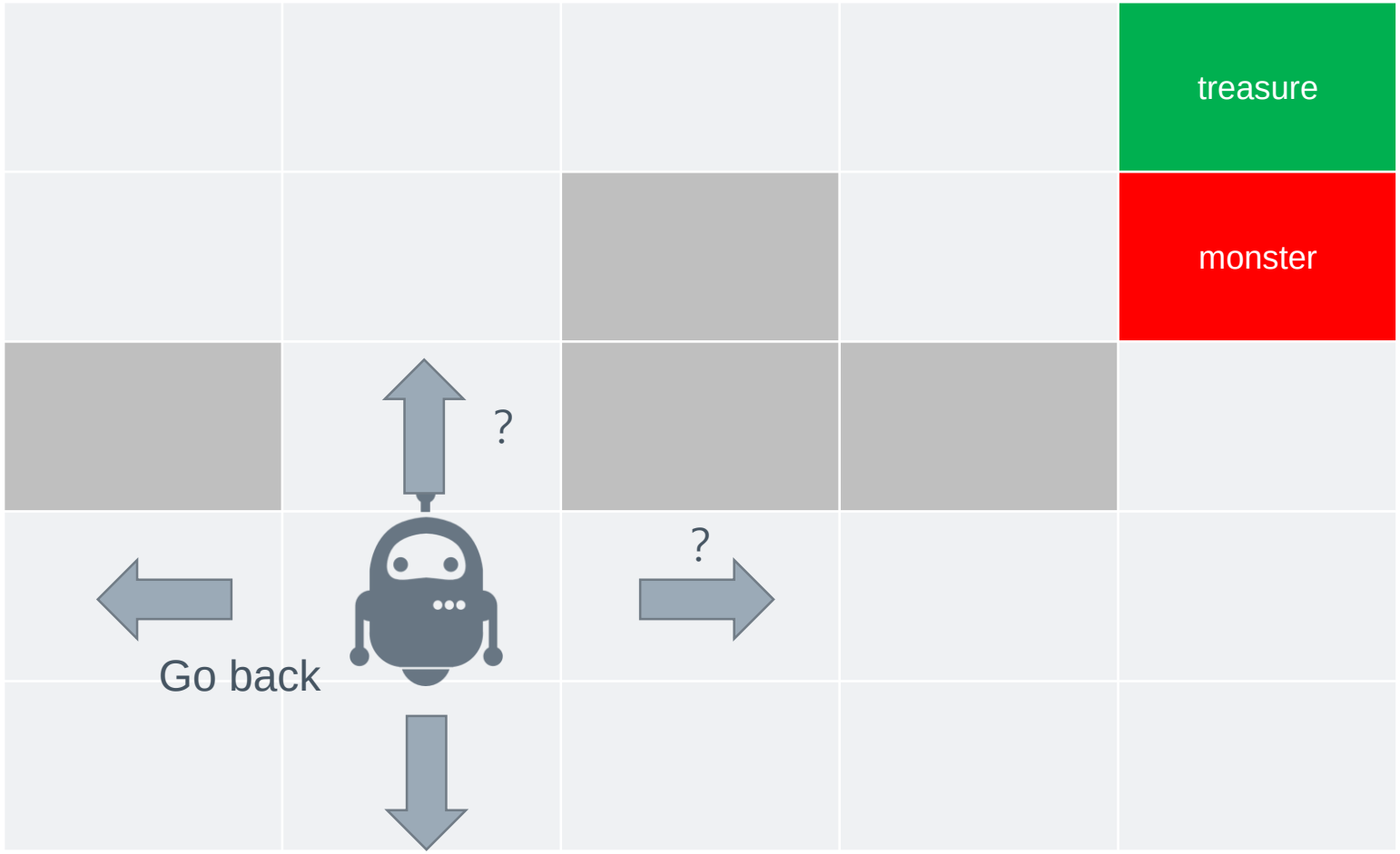
› Learning from rewards
  –Training our kids



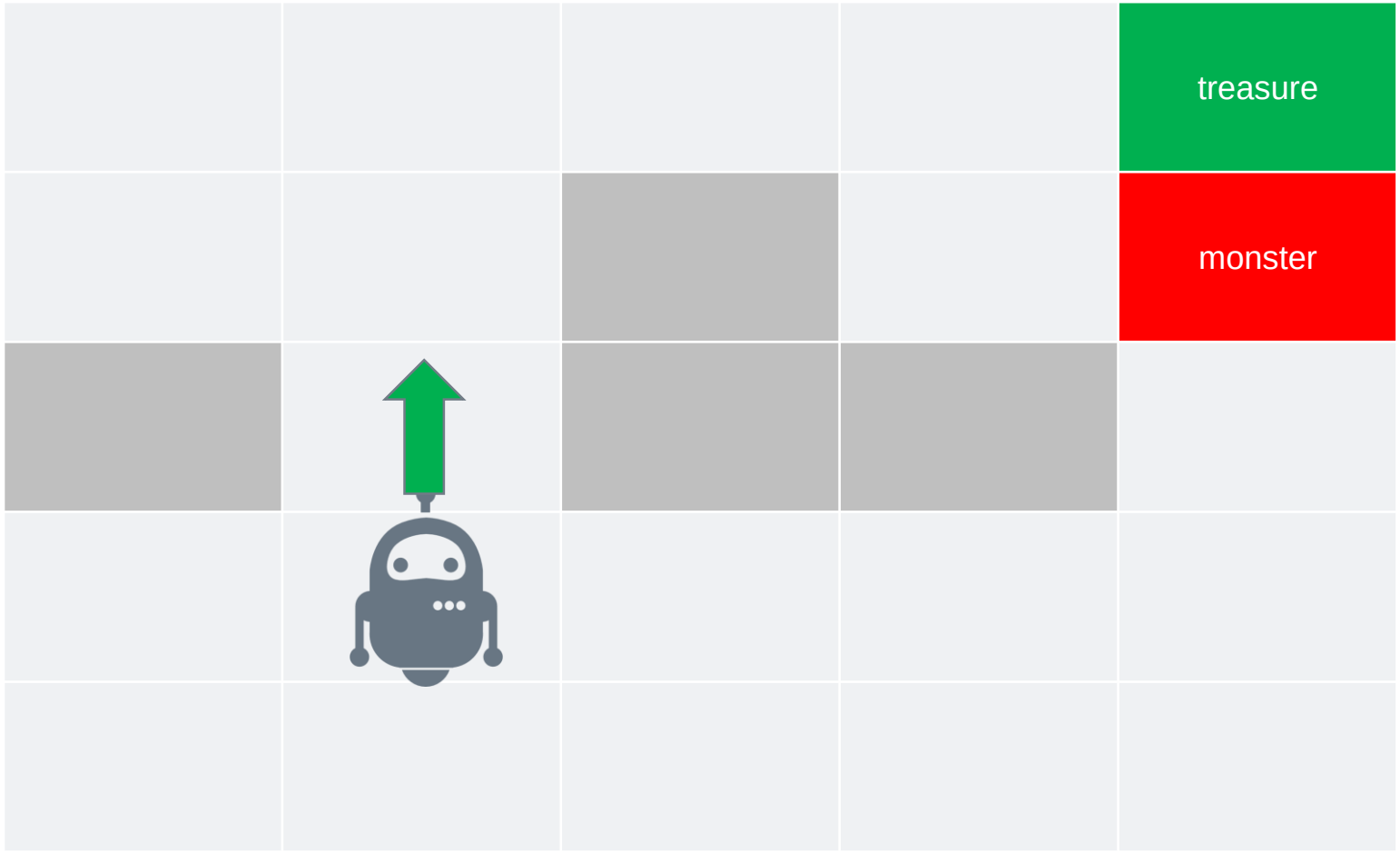
The Example

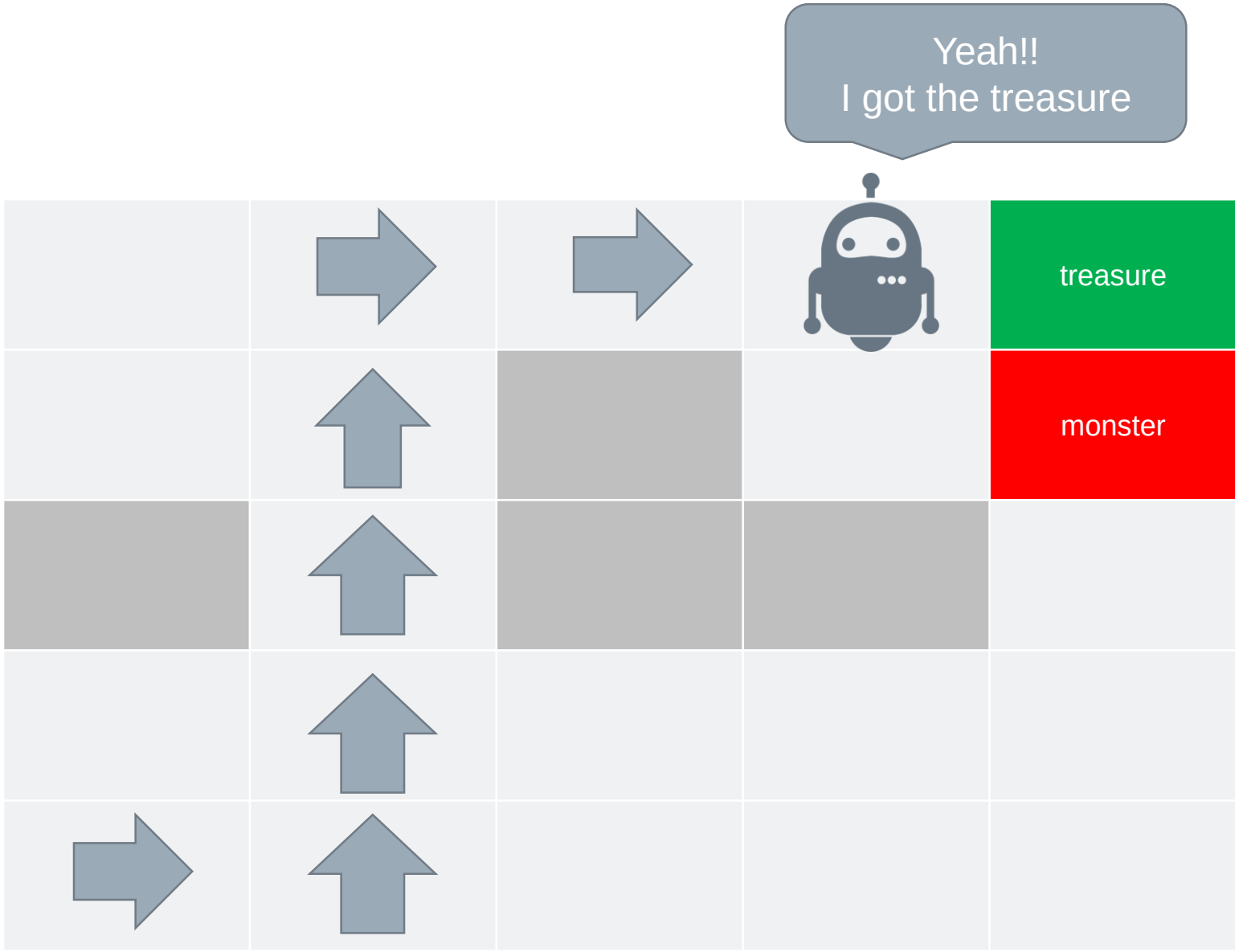


How Do I Decide Next Step

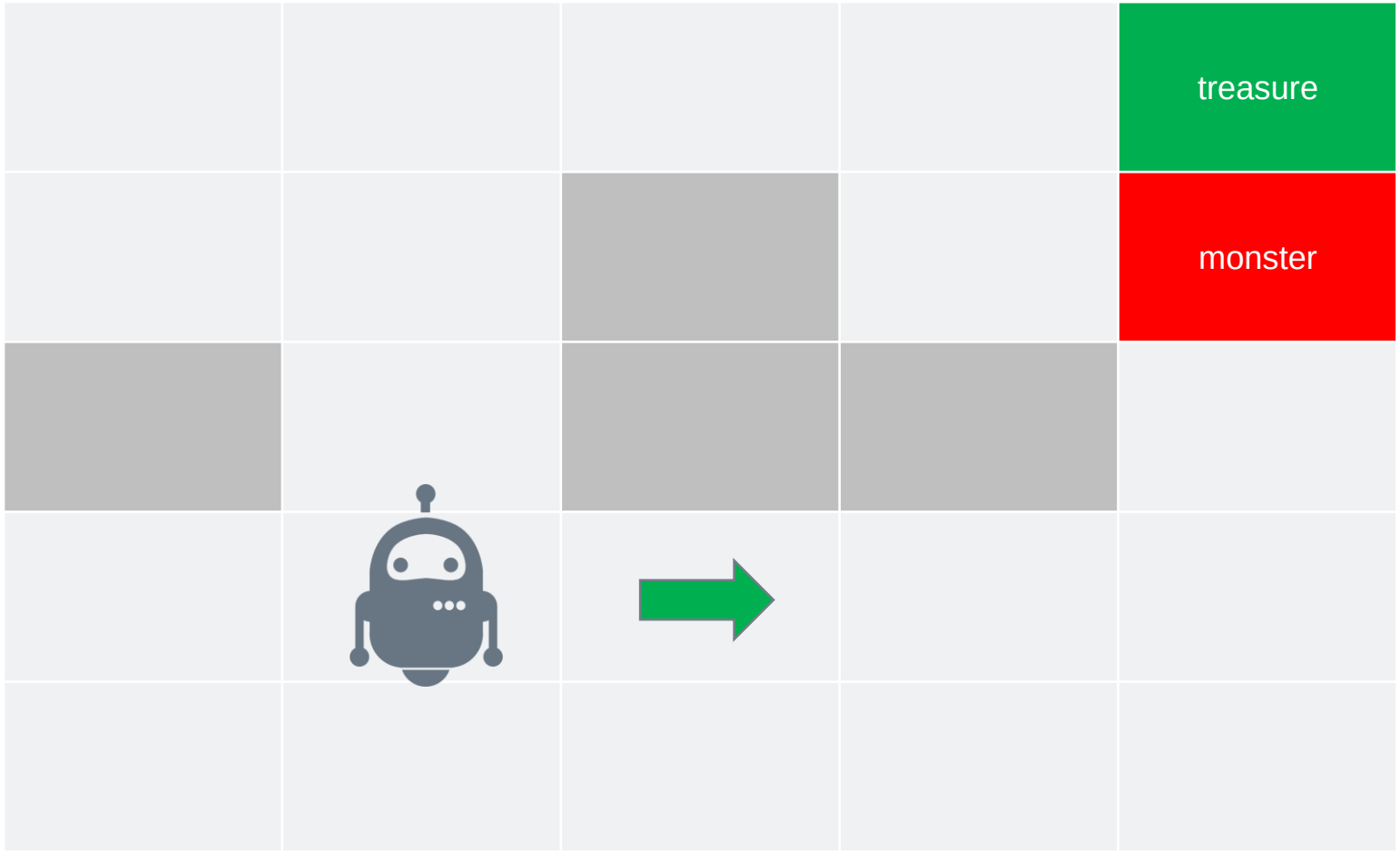


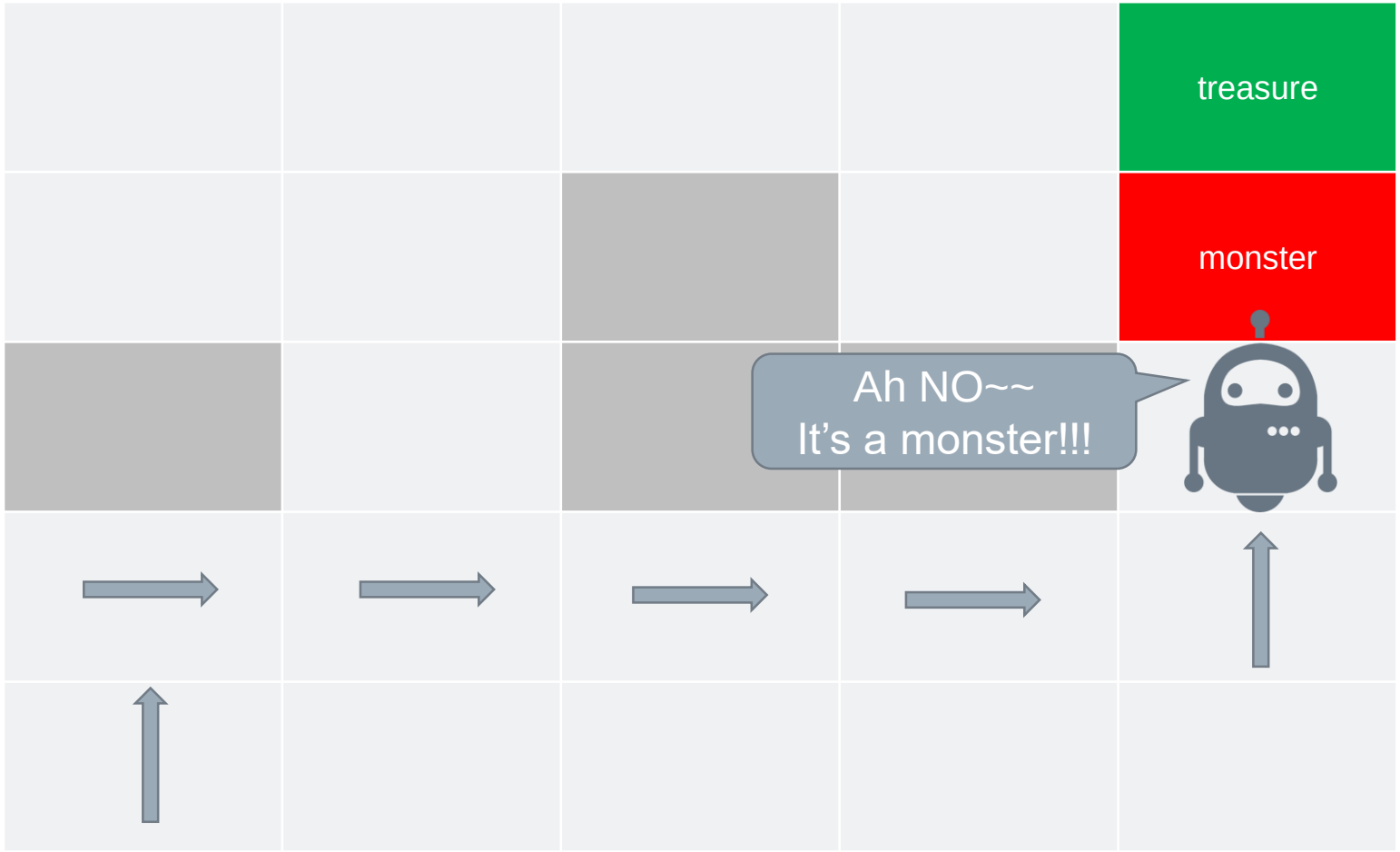
1st Choice



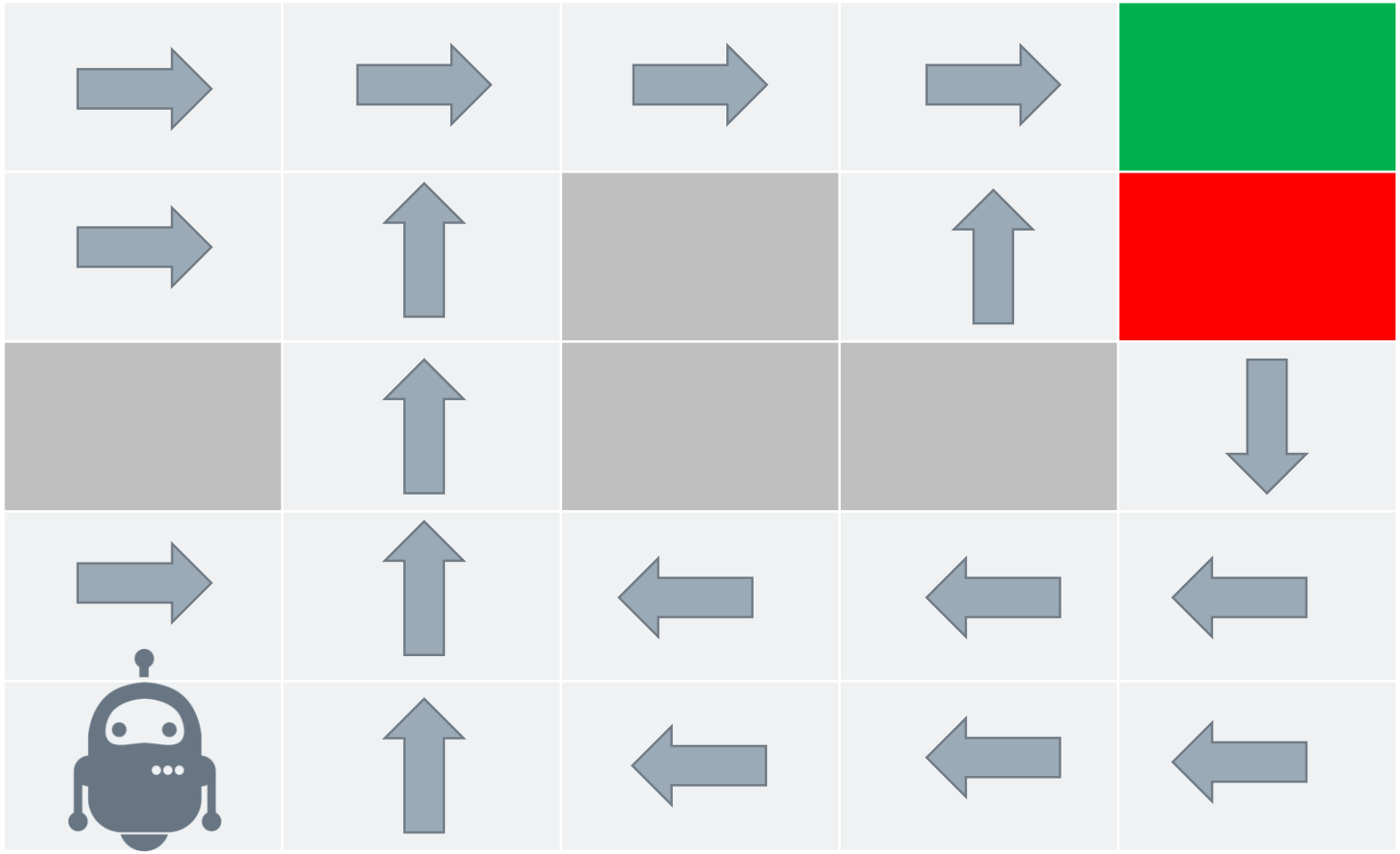

Yeah!!
I got the treasure
treasure
monster

2nd Choice



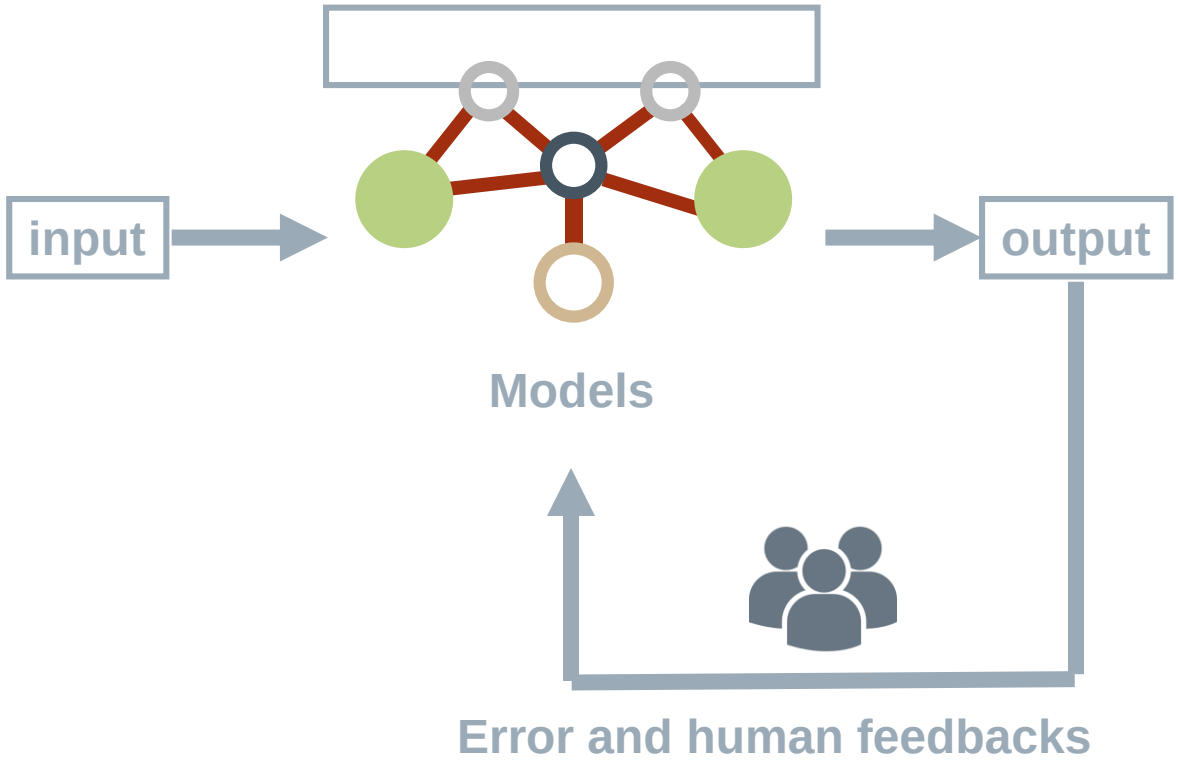

treasure
monster
Ah NO~~
It's a monster!!!

The Best Action in Each Step



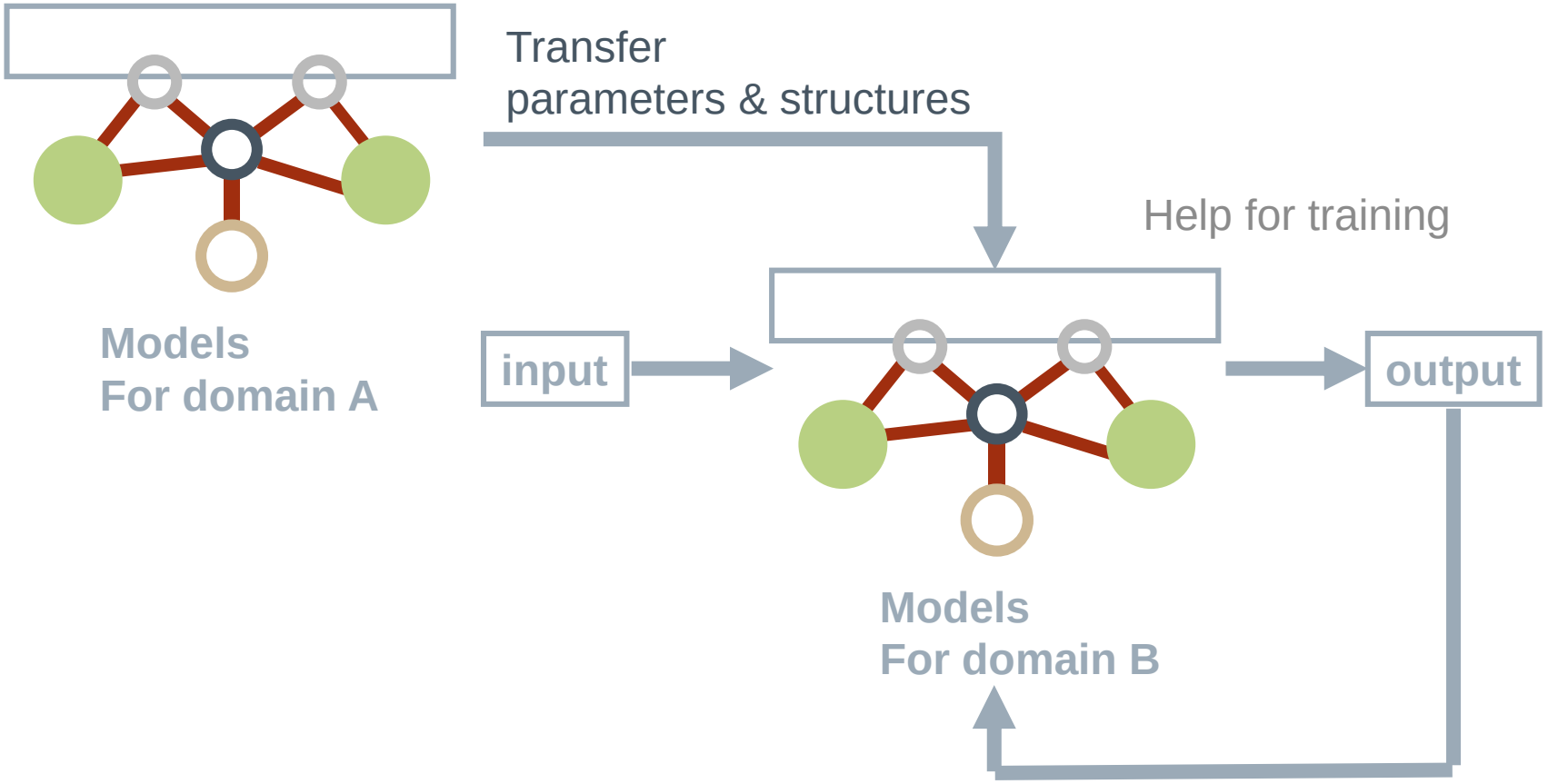
Interactive Learning

› Human can give feedbacks to machine to help machine train a better model



Transfer Learning

› Transfer some knowledge from trained models to new models



General Machine Learning Steps

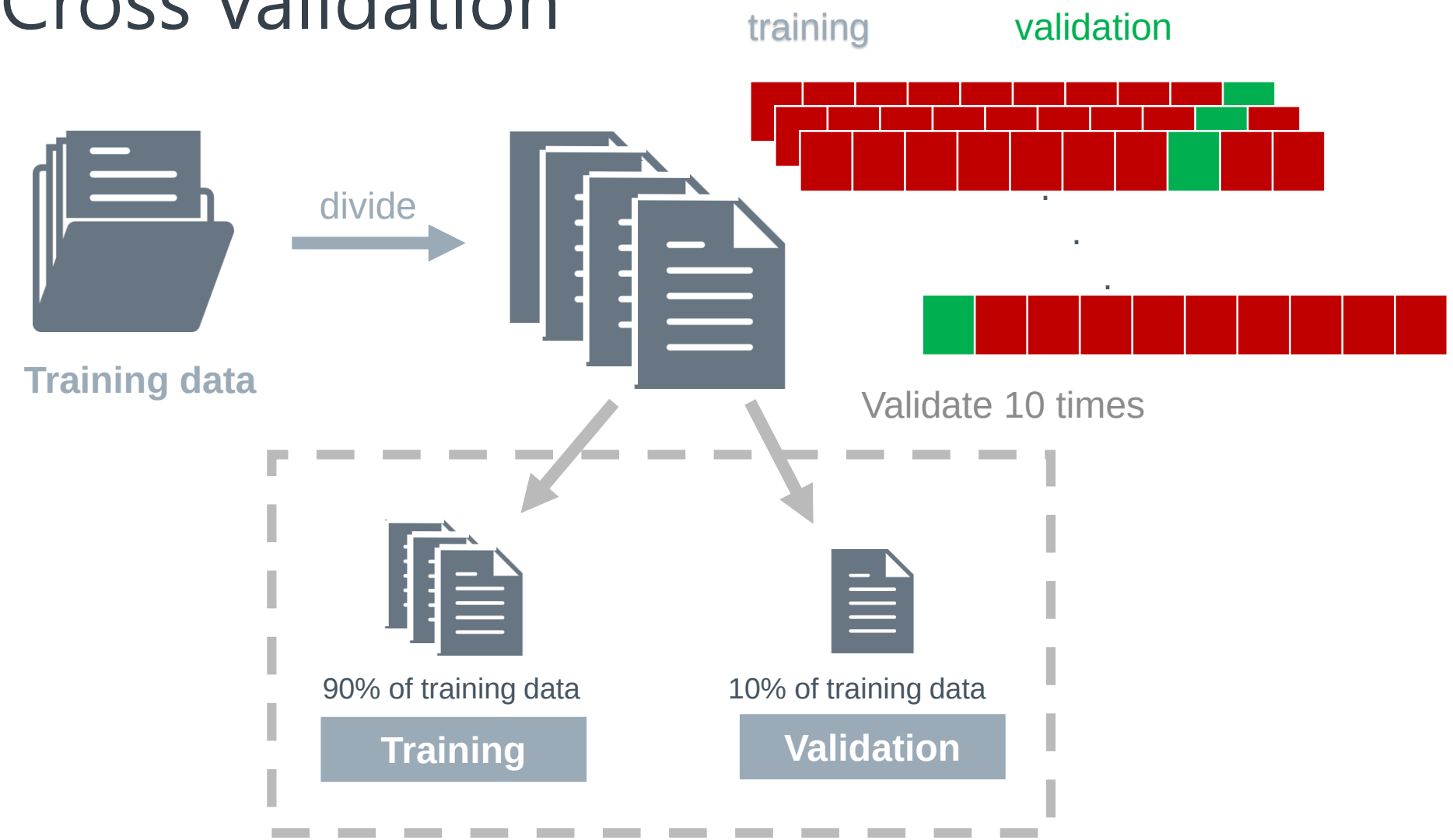
Data preparation
- Preprocessing
- Exploration
- Split data into training set and testing set

Model development
- Determine your objective functions
- Determine the learning types

Training
- Solve the optimization problem you designed or fit your model with data

Testing
- Test the model performance on the testing data

Cross Validation



Quality Measurement

- Accuracy
  - $ACC = \frac{TP+TN}{P+N}$
- Precision
  - $Pre = \frac{TP}{PP}$
- Recall
  - Rec= $\frac{TP}{P}$
- F-measure
  - $F1 = 2 \times \frac{Pre \times Rec}{Pre+Rec}$

prediction \ condition	Positive (P)	Negative (N)
Predict Positive (PP)	True Positive (TP)	False Positive (FP)
Predict Negative (PN)	False Negative (FN)	True Negative (FN)

Compare with Data Mining

Data Mining

- Data collection
- Data cleaning
- Data integration
- Data warehousing
- Data selection
- Pattern evaluation
- Knowledge discovery
- Information presentation
- Decision making

Machine Learning

- Data collection
- Data cleaning
- Data integration
- Data warehousing
- Data selection
- Model design
- Model training & testing
- Information presentation
- Decision making

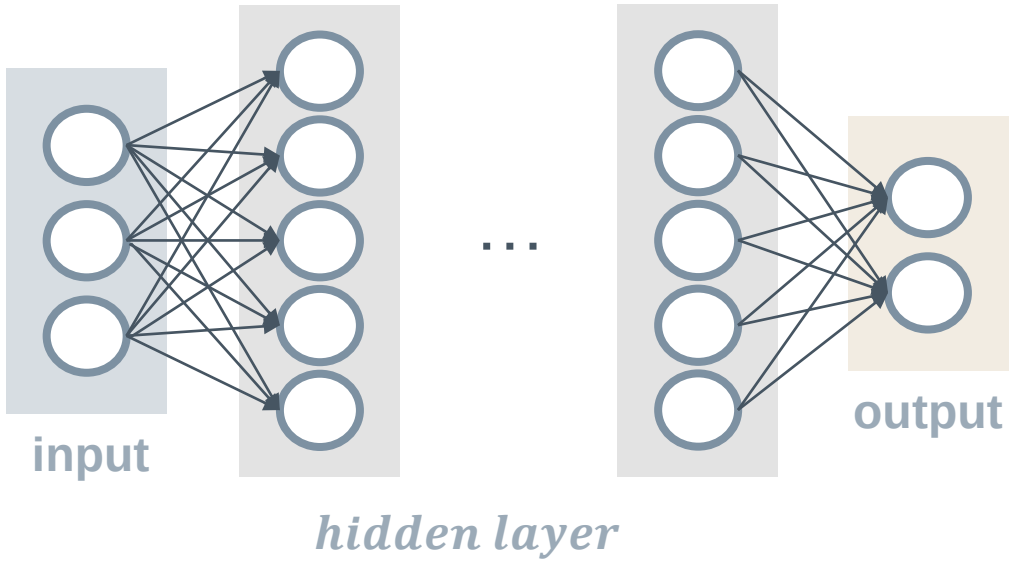
Introduction to Deep Learning

Deep Learning

› Using some deep (with many layers) models to solve the problem

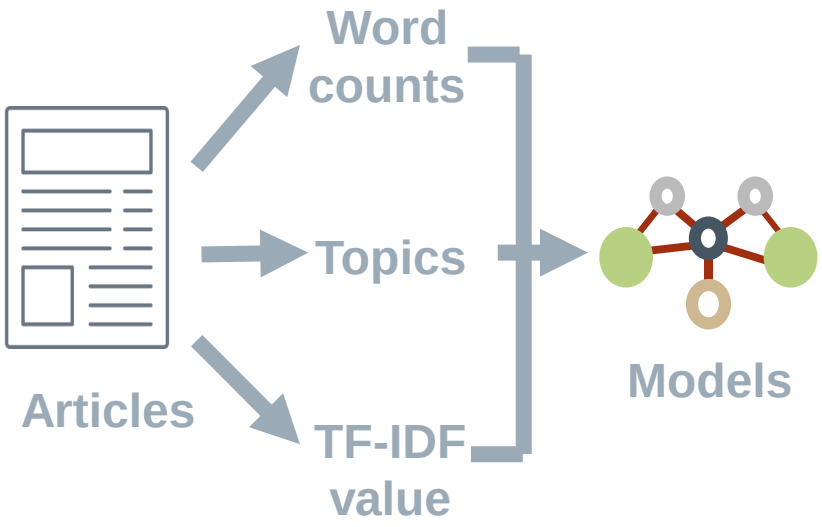
› Each layer can represent different features

– We can learn more knowledge from input data

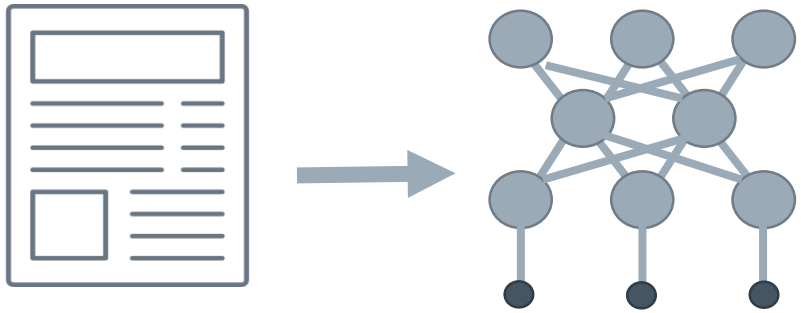


The Strong Points of Deep Learning

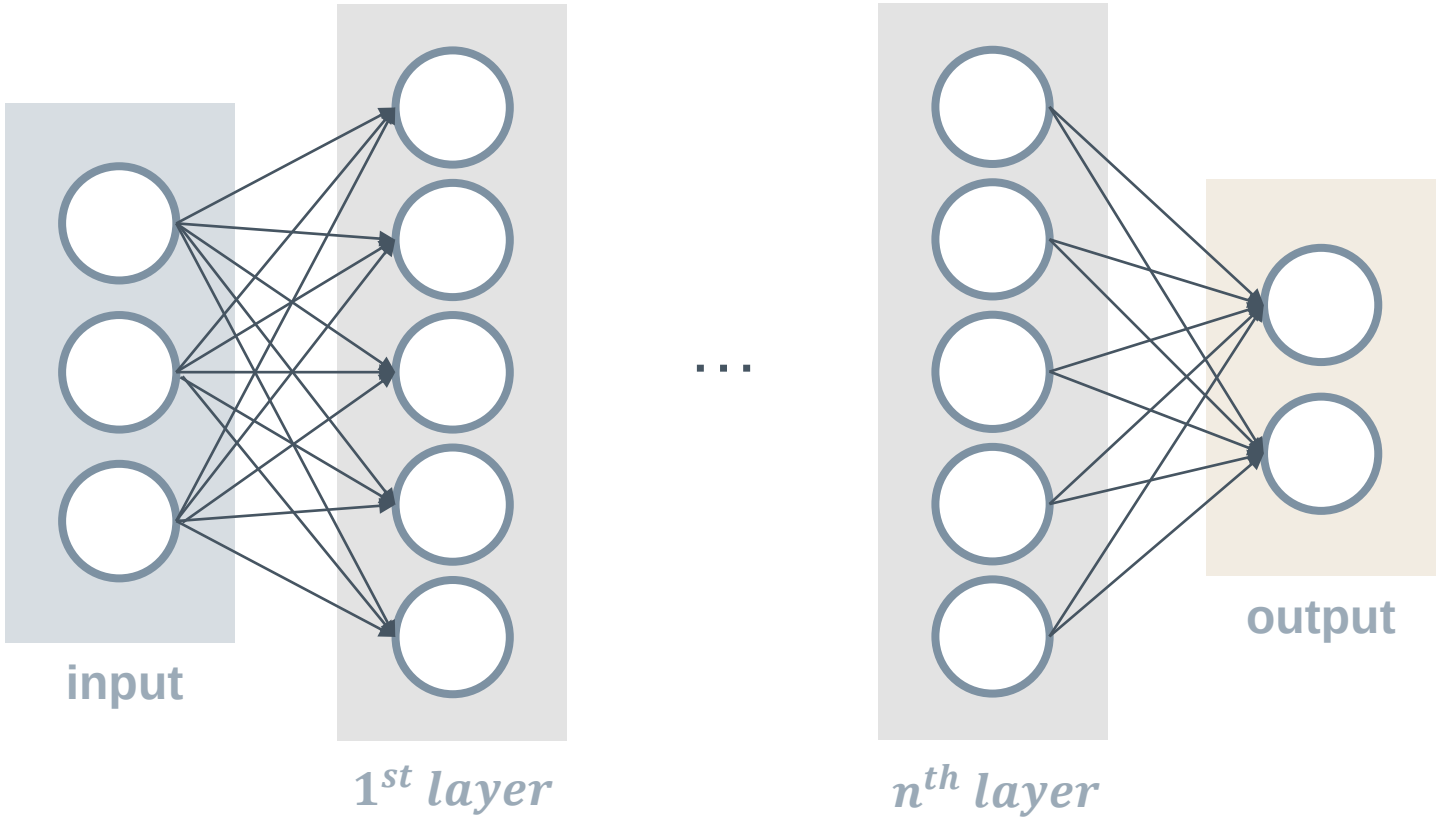
Without deep learning



With deep learning

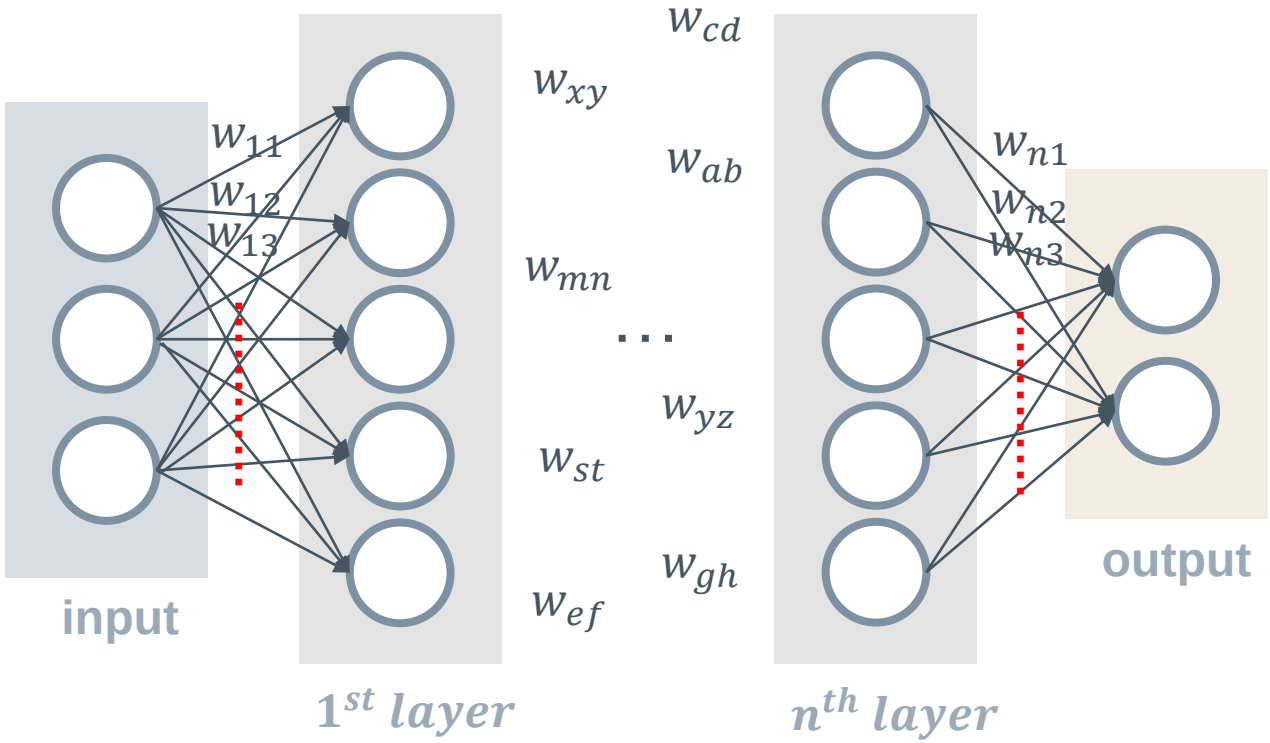


The Strong Points of Deep Learning



Represent different features

The Weak Points of Deep Learning



Lots of parameters to be trained
Need lots of data to train a good model