UTS 2019

MACHINE LEARNING – LEARNING PROBLEMS

1. LEARNING PROBLEM

THE LEARNING PROBLEM

- Your imagination about what we will learn in the subject of “machine learning”. What machine learning can do for us or what kind of difficult challenges we will face in its lack of?

- **Q:** Decide whether the following are learning problems we will study

A-D. watch video

E. Chinese-Go program

Explanation video

INTRODUCTION TO LEARNING PROBLEM @ UDACITY

- Auto Car as a Learning Problem
- Observation: Car Sensor Data
- Concept:
  - Sensor readings that require turning the wheel by 45deg to the left.
  - Sensor readings that require applying break by 50%.
  - …

- Supervised Learning
- Features and Labels

BINARY CLASSIFICATION PROBLEM

- When discussing the basics of machine learning theory, we will focus on building models that produce a binary prediction from observed data. I.e. we are considering the following problem:

$$X \xrightarrow{f} Y, Y \in \{0,1\}$$

Let us get well motivated at the very beginning of our journey in the machine learning area:

- Q: How do you think about binary prediction, as a class of problems?

  A. In theory, it covers all problems that we can possibly expect a machine (or any intelligent agent) could provide information with. In practice, it does represent a large class of machine learning problems. Basic theories developed on the binary classification is almost complete.

  B. Binary classification can only provide answers to yes/no questions, which is very limited. We need significantly more development in the theory.

BINARY CLASSIFICATION PROBLEM

- When discussing the basics of machine learning theory, we will focus on building models that produce a binary prediction from observed data. I.e. we are considering the following problem:

$$X \xrightarrow{f} Y, Y \in \{0,1\}$$

Let us get well motivated at the very beginning of our journey in the machine learning area:

- Q: How do you think about binary prediction, as a class of problems?

A. In theory, it covers all problems that we can possibly expect a machine (or any intelligent agent) could provide information with. In practice, it does represent a large class of machine learning problems. Basic theories developed on the binary classification is almost complete.

B. Binary classification can only provide answers to yes/no questions, which is very limited. We need significantly more development in the theory.

BINARY CLASSIFICATION PROBLEM

EXPLANATION

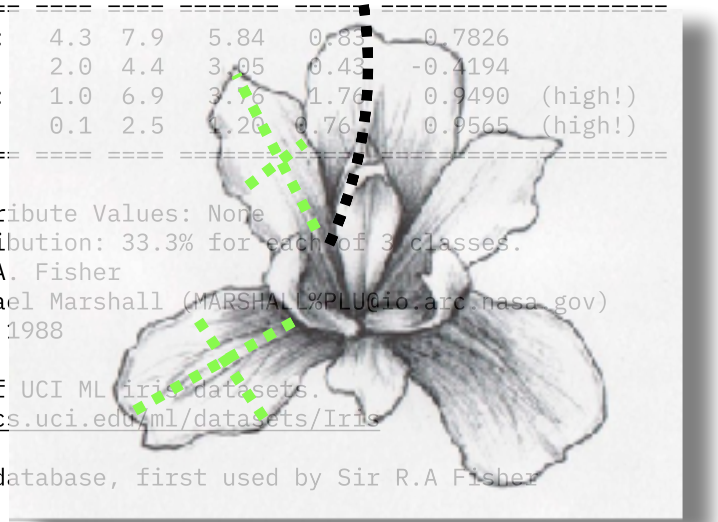
- You may feel surprised how much information we could gain from just answering yes/no questions. All problems that demand discrete answers can be conveniently converted to classic problems. That is, whenever you can count out all the possibilities (maybe infinitely many), binary classification can be applied.

CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1

Consider the IRIS flower classification problem.

- X1 (sepal-L)
- X2 (sepal-W)
- X3 (pedal-L)
- X4 (pedal-H)

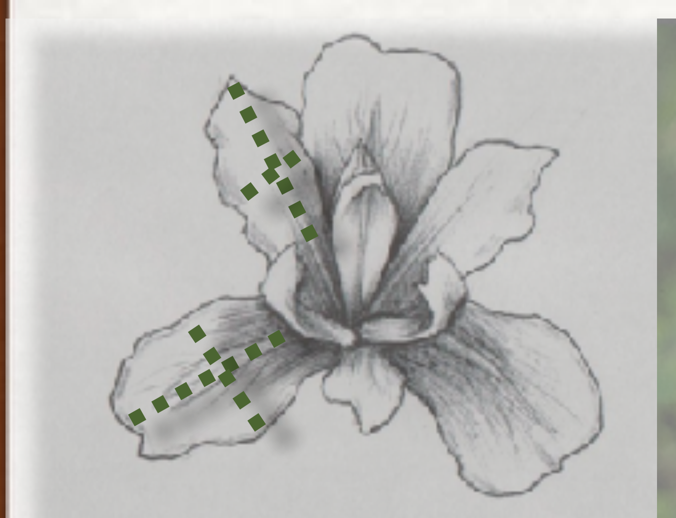
```
Iris Plants Database
====================

Notes
-----
Data Set Characteristics:
    :Number of Instances: 150 (50 in each of three classes)
    :Number of Attributes: 4 numeric, predictive attributes and the
class
    :Attribute Information:
        - sepal length in cm
        - sepal width in cm
        - petal length in cm
        - petal width in cm
        - class:
                - Iris-Setosa
                - Iris-Versicolour
                - Iris-Virginica
    :Summary Statistics:

    ============== ==== ==== ======= ===== ====================
                    Min  Max   Mean    SD   Class Correlation
    ============== ==== ==== ======= ===== ====================
    sepal length:   4.3  7.9   5.84   0.83    0.7826
    sepal width:    2.0  4.4   3.05   0.43   -0.4194
    petal length:   1.0  6.9   3.76   1.76    0.9490  (high!)
    petal width:    0.1  2.5   1.20   0.76    0.9565  (high!)
    ============== ==== ==== ======= ===== ====================

    :Missing Attribute Values: None
    :Class Distribution: 33.3% for each of 3 classes.
    :Creator: R.A. Fisher
    :Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
    :Date: July, 1988

This is a copy of UCI ML iris datasets.
http://archive.ics.uci.edu/ml/datasets/Iris

The famous Iris database, first used by Sir R.A Fisher

This is perhaps the best known database to be found in the
pattern recognition literature.  Fisher's paper is a classic in the
field and
is referenced frequently to this day.  (See Duda & Hart, for
example.)  The
```



CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1

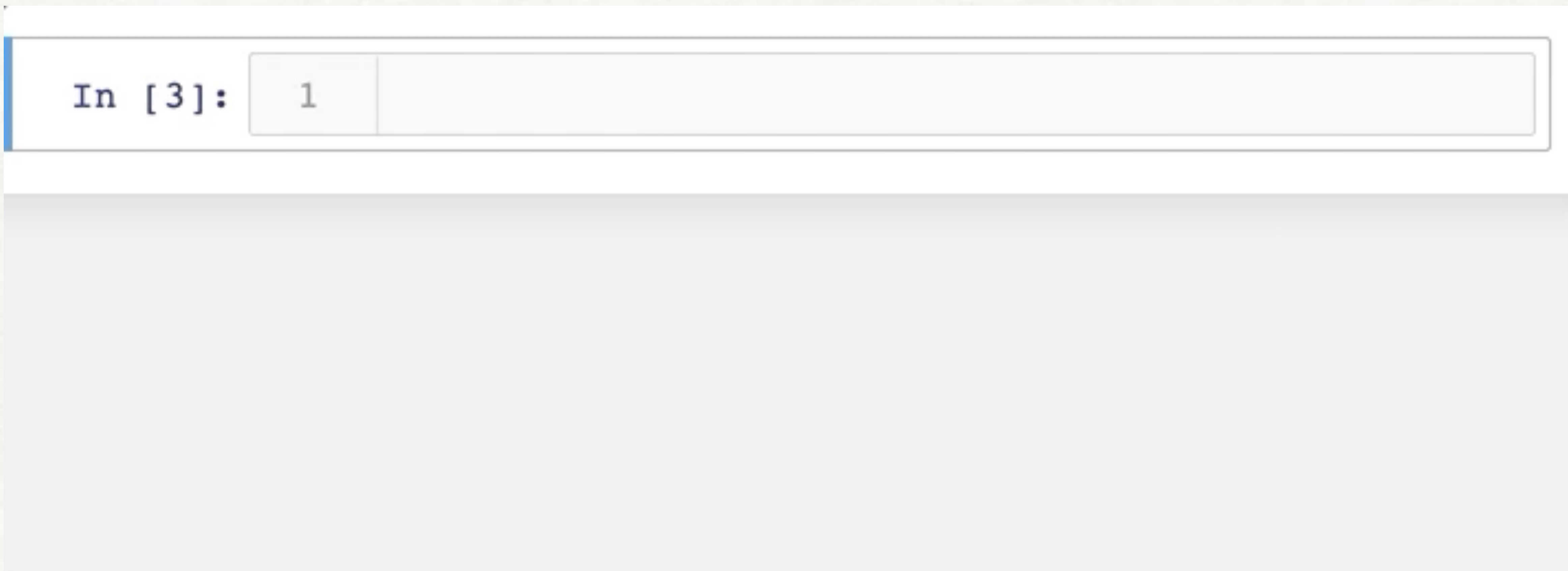
Data and Task

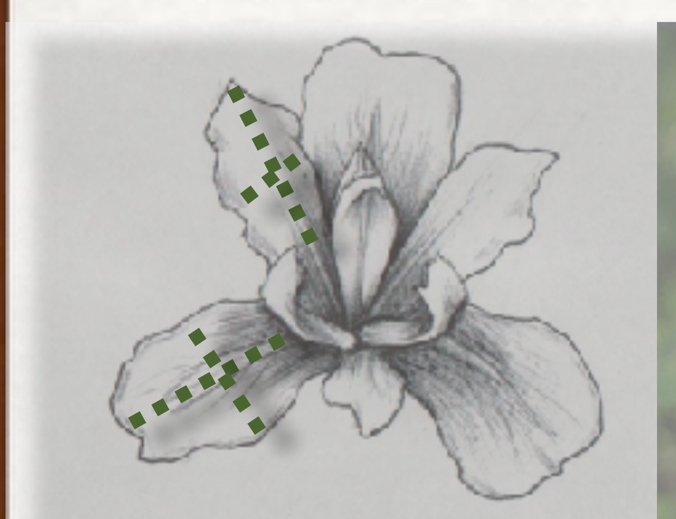
```
In [3]:  1
```



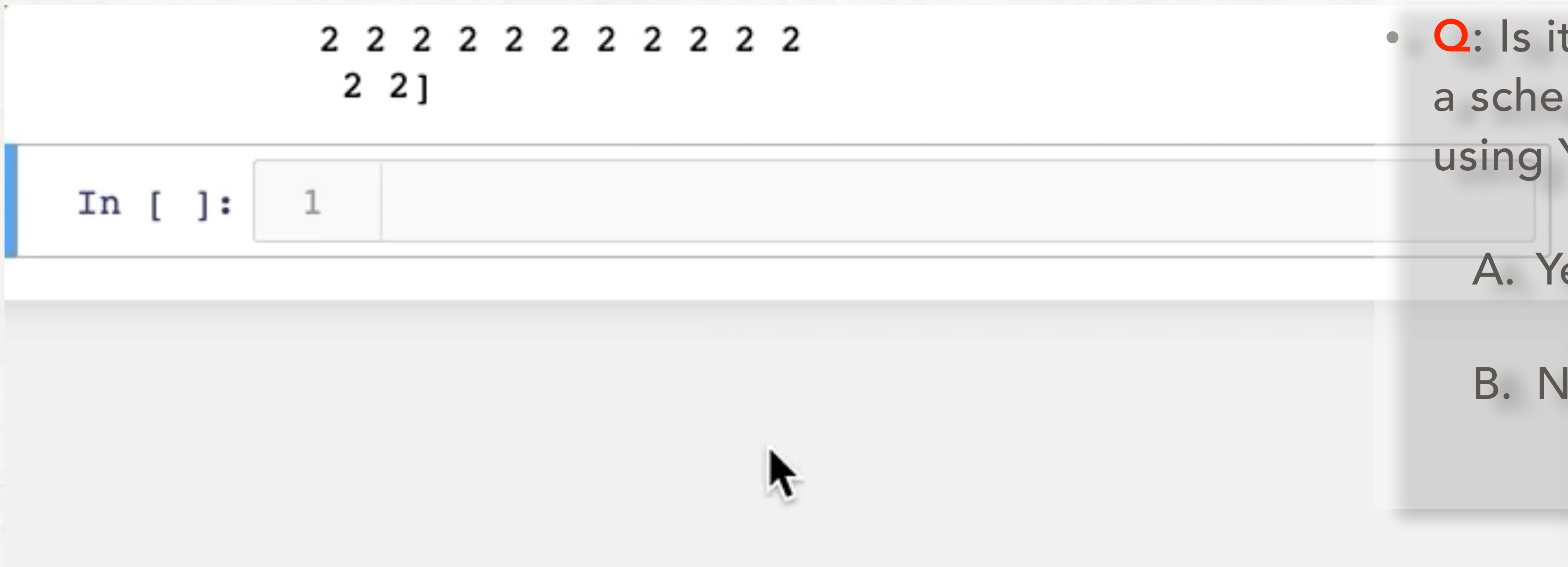

CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1

Data and Task






CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1



- **Q**: Is it possible to design a scheme to implement "f" using Yes/No predictions?

  A. Yes.

  B. No.

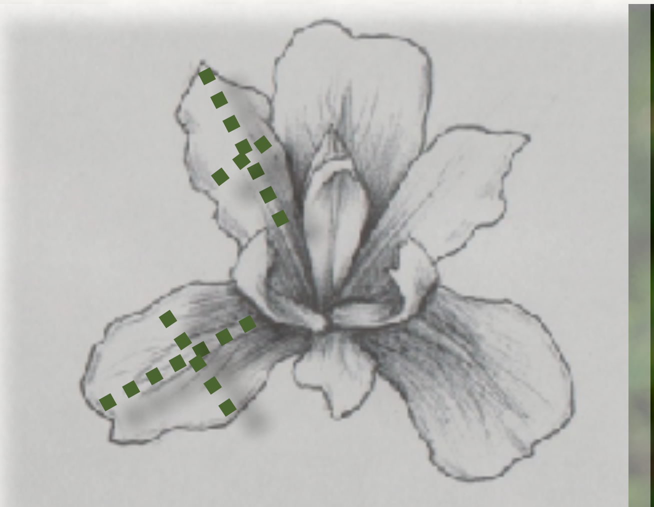
CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1

```
 2  2  2  2  2  2  2  2  2  2  2
  2  2]

In [ ]:  1
```

- **Q**: Is it possible to design a scheme to implement “f” using Yes/No predictions?

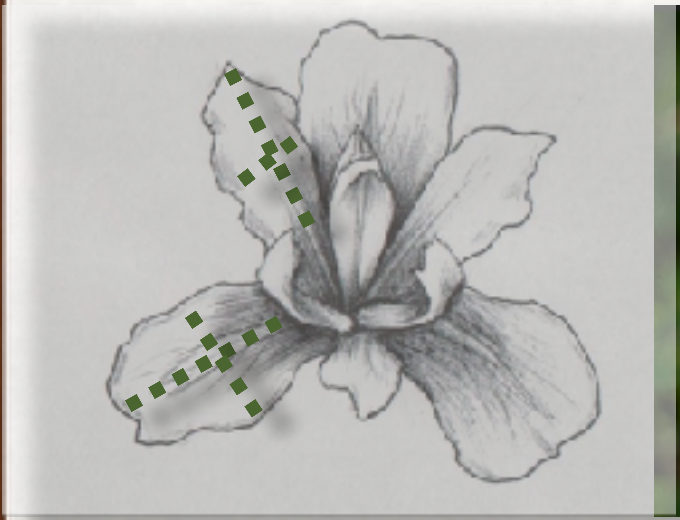
  A. Yes.

  B. No.



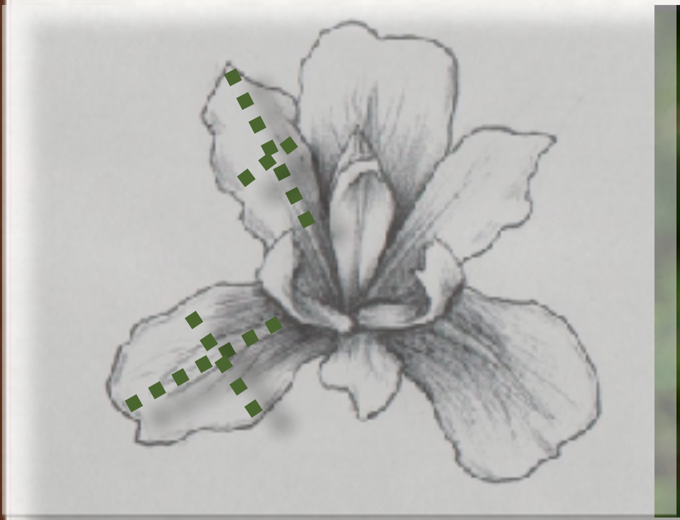

CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1

```python

def f(x):
    # x: x[0], x[1], x[2], x[3]


    return prediction # 0, 1, 2





```



CLASSIFICATION BY BINARY PREDICTIONS: EXAMPLE1

```
def f(x):
    # x: x[0], x[1], x[2], x[3]


    return prediction # 0, 1, 2
```




CLASSIFICATION BY BINARY PREDICTIONS

Example 2: If you consider ranking all the students in this classroom by their performance in an assignment, The ordering operation can be considered as answering questions like

- whether student1's performance is greater than student2's;
- whether student1's performance is greater than student3's;
- …
- whether student30' performance is greater than student40's;
- …

CLASSIFICATION BY BINARY PREDICTIONS

Example 3: Consider the task of translating from English to French,

Nice to meet you -> Ravi de vous rencontrer

The questions can be: when to translate "nice to meet you", let exhaustively go through a French dictionary:

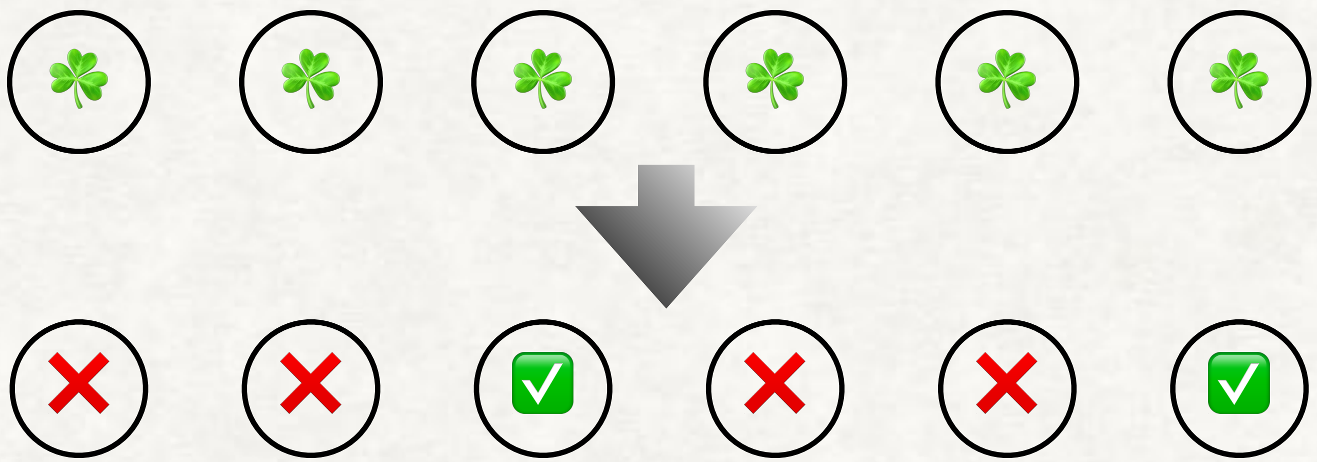
- whether the first word of the output sentence should be "merci";
- whether the first word of the output sentence should be "de";
- …
- whether the first word of the output sentence should be "ravi";

BINARY FORMAT DELIVERS INFORMATION

- Put it in a simple way, consider the information you could possibly get from your mobile or your computer while connecting the device to the Internet, it is literally the entire knowledge base developed by humanity. While down to the physical layer of the connection, all the information has to be communicated to you through a channel that can only carry 0's and 1's. The point is when putting the questions right, the answers 0/1's can carry a lot of information.

A TARGET CONCEPT IN THE DATA SPACE

- Target (Label / Y-Variable) Whether Versicolour?



"Final_Y" ↦ The customers whose Final_Y tend to be True.

- Describe a subset of samples / a sub-population of a distribution / build a discriminative model.

- **Q:** Having finished learning, Alice got a predictor $X \xrightarrow{f} y$. What kinds of data (the X attributes) can f accept as input?

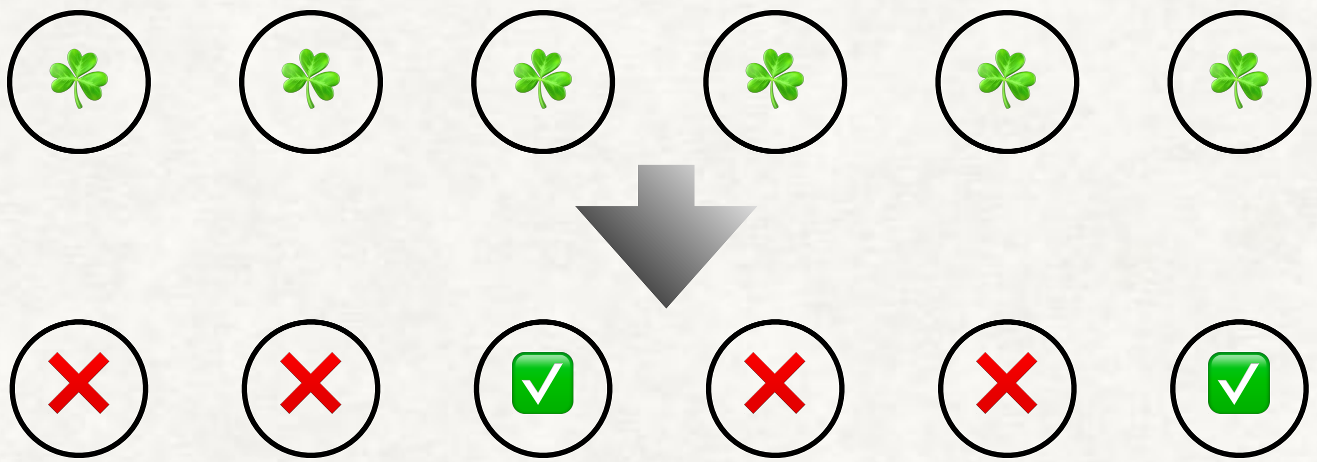
A. X-samples that Alice were **given** when training f.

B. The attributes in X can take **any** value and be combined arbitrarily.

C. The attributes in X can take values **unseen** during training, but from the **same distribution** from which the training samples were drawn.

A TARGET CONCEPT IN THE DATA SPACE

- Target (Label / Y-Variable) Whether Versicolour?



“Final_Y” ↦ The customers whose Final_Y tend to be True.

- Describe a subset of samples / a sub-population of a distribution / build a discriminative model.

- **Q:** Having finished learning, Alice got a predictor $X \xrightarrow{f} y$. What kinds of data (the X attributes) can f accept as input?

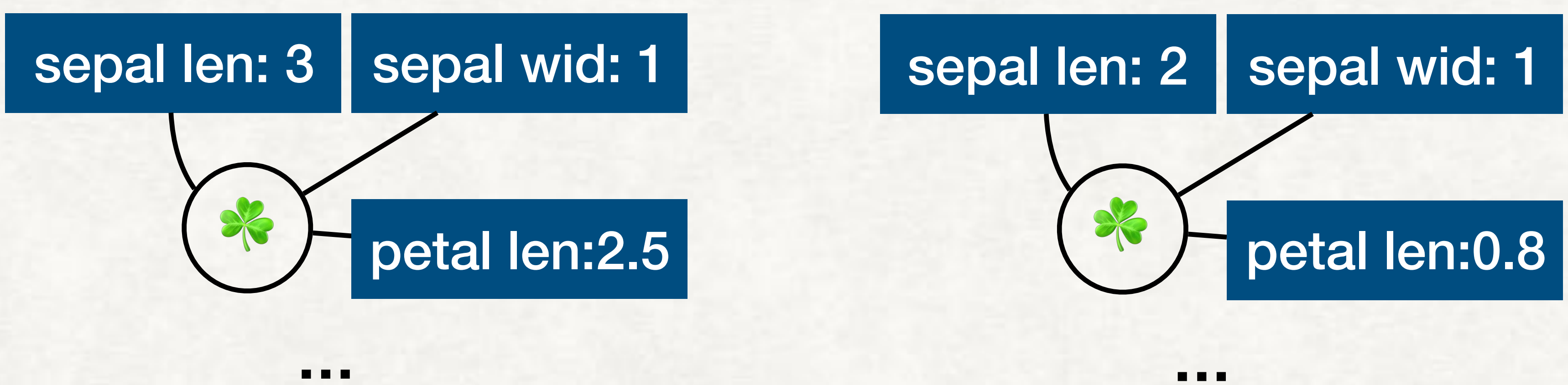
A. X-samples that Alice were **given** when training f.

B. The attributes in X can take **any** value and be combined arbitrarily.

C. The attributes in X can take values **unseen** during training, but from the same **distribution** from which the training samples were drawn.

THE DATA SPACE

- X / Observations / Input
  - suffixes: “Variables” “Space” “Attributes”

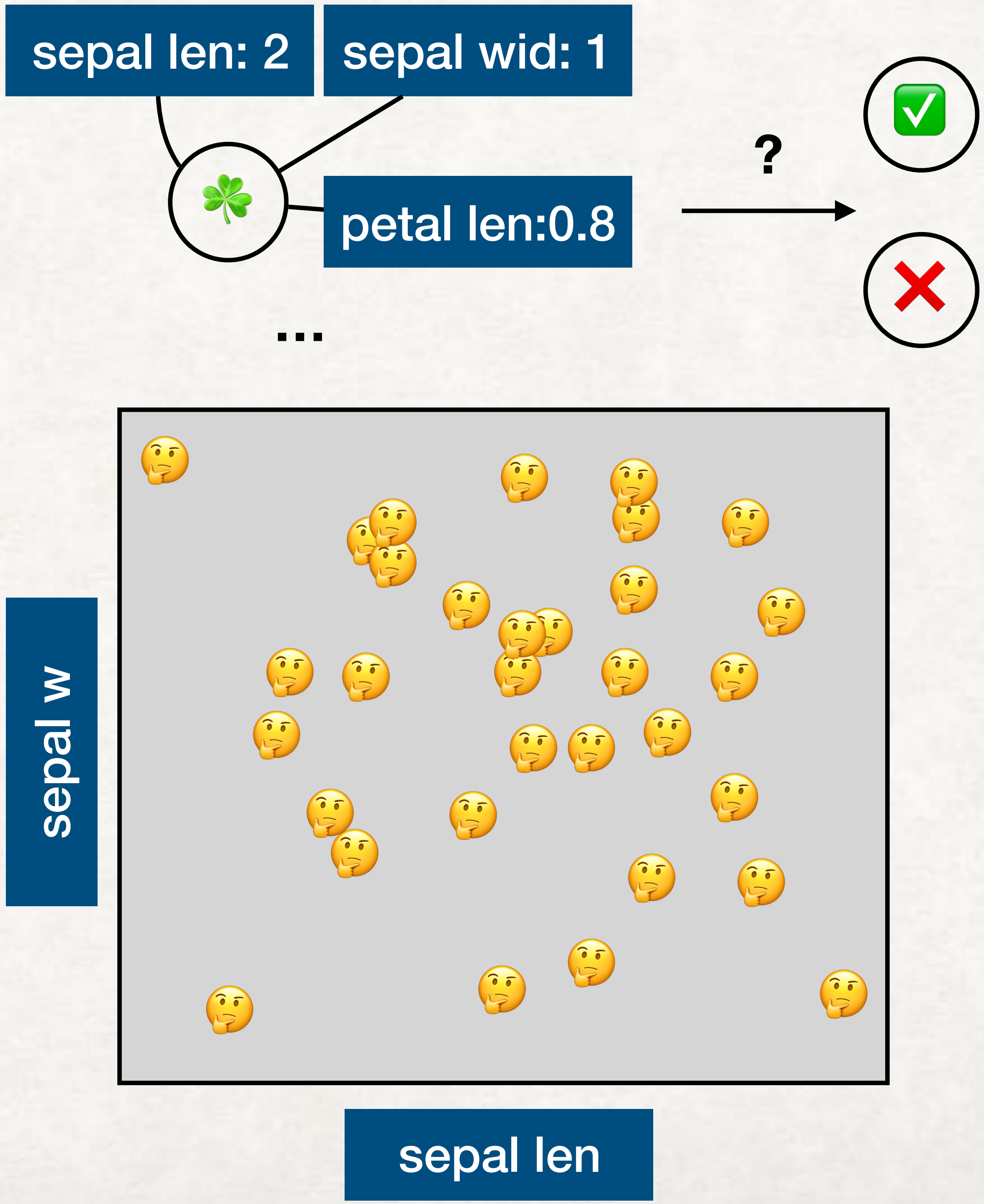


Information transforms the samples into objects of particular form.

- Describe a subset of samples in an X-space.

A SIMPLIFIED EXAMPLE

- For example, we may have different types of information about the objects. On the other hand, to learn some concepts, we may choose to use some subset of the information, or the information is some extracted format. Such as shown in the right figure.

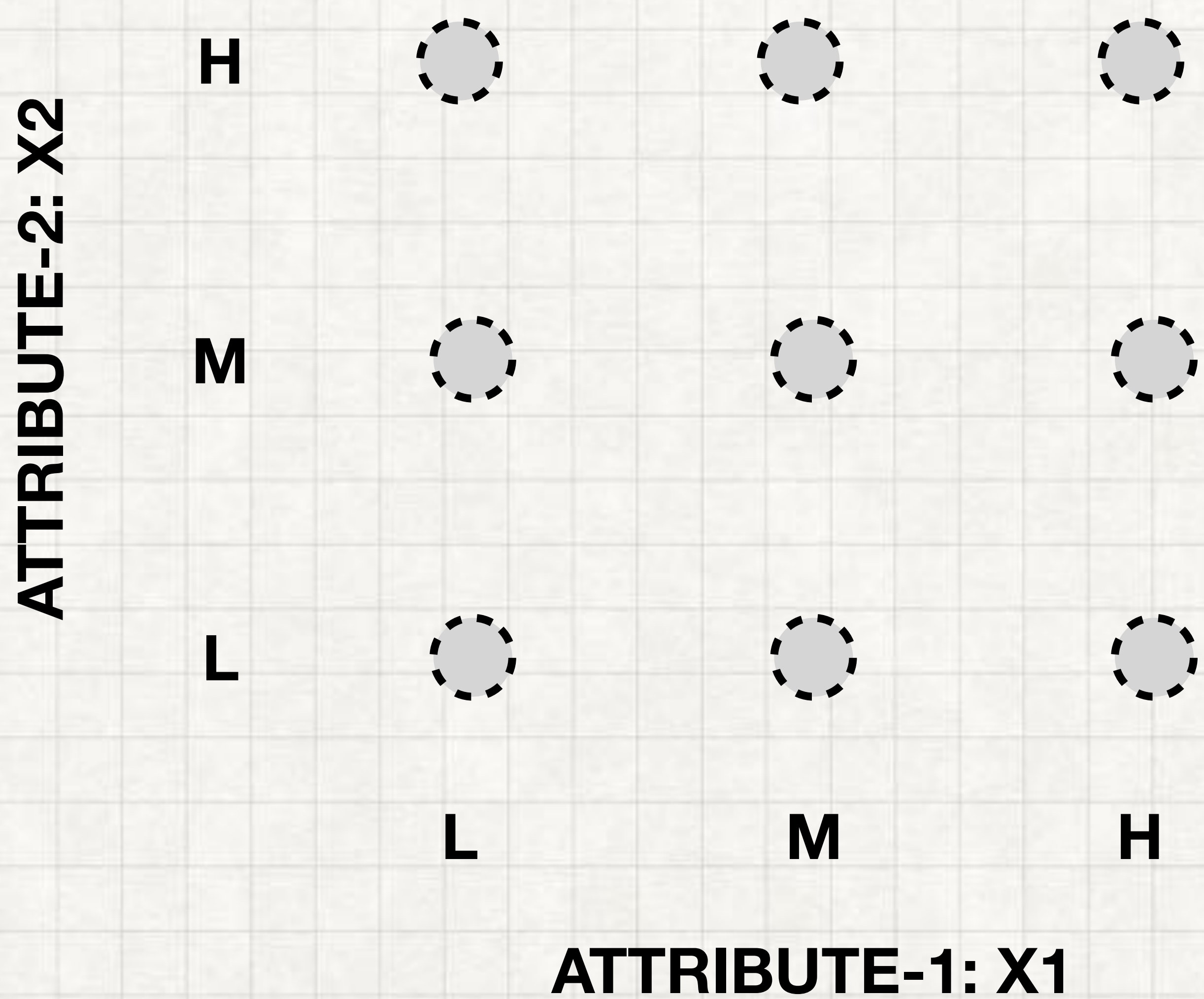


An armchair data scientist

- Try to:
  - Picture money — **data**: historical trade transactions and price of a share, **concept**: is it now profitable to buy?
  - Picture people — **data**: social media posts, **concept**: going to vote for Trump in 2020?
  - Picture pictures — **data**: all pixel RGB tuples, **concept**: does the image contain dogs?

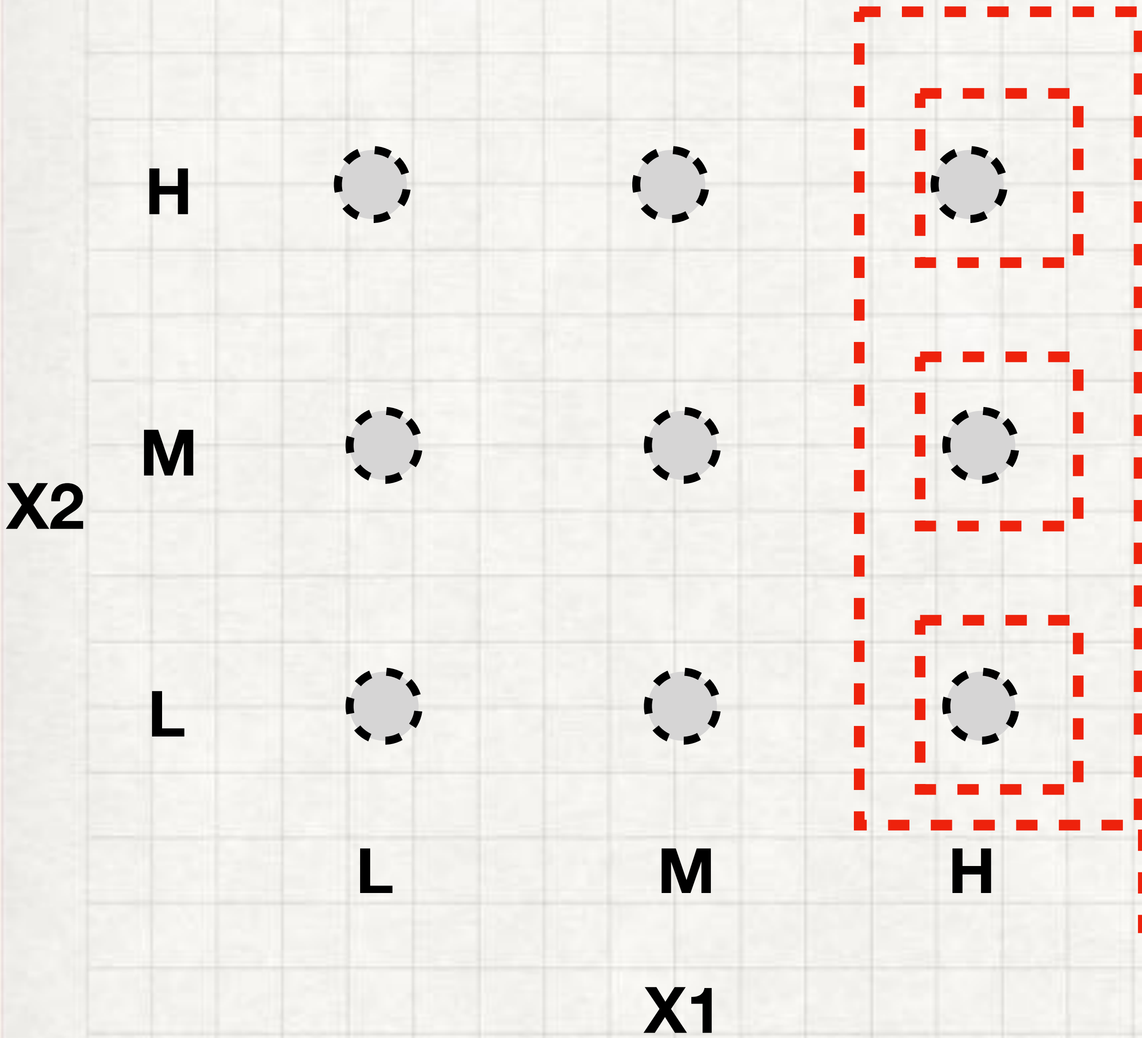
2. LEARNING CHALLENGES

SIMPLE-IRIS: CLASSIFICATION ON A 2D GRID



Three-step simplification, we consider the binary classification "belonging to class 1", we use only two attributes, we discretised the values of the age appears into "high" "medium" and "low".

COMPLETE X-Y COMBINATIONS



	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
q0		0	0	0	0	0	0	0	0	0
q1		0	0	0	0	0	0	0	0	1
q2		0	0	0	0	0	0	0	1	0
q3		0	0	0	0	0	0	0	1	1
q4		0	0	0	0	0	0	1	0	0
q5		0	0	0	0	0	0	1	0	1
q6		0	0	0	0	0	0	1	1	0
q7		0	0	0	0	0	0	1	1	1
...										

SIZE OF A COMPLETE TABLE

	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
q0		0	0	0	0	0	0	0	0	0
q1		0	0	0	0	0	0	0	0	1
q2		0	0	0	0	0	0	0	1	0
q3		0	0	0	0	0	0	0	1	1
q4		0	0	0	0	0	0	1	0	0
q5		0	0	0	0	0	0	1	0	1
q6		0	0	0	0	0	0	1	1	0
q7		0	0	0	0	0	0	1	1	1
…										

- Q: How many rows in this table?

  A. 9

  B. 18

  C. 512

  D. ∞

SIZE OF A COMPLETE TABLE

	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
q0		0	0	0	0	0	0	0	0	0
q1		0	0	0	0	0	0	0	0	1
q2		0	0	0	0	0	0	0	1	0
q3		0	0	0	0	0	0	0	1	1
q4		0	0	0	0	0	0	1	0	0
q5		0	0	0	0	0	0	1	0	1
q6		0	0	0	0	0	0	1	1	0
q7		0	0	0	0	0	0	1	1	1
...										

- Q: How many rows in this table?

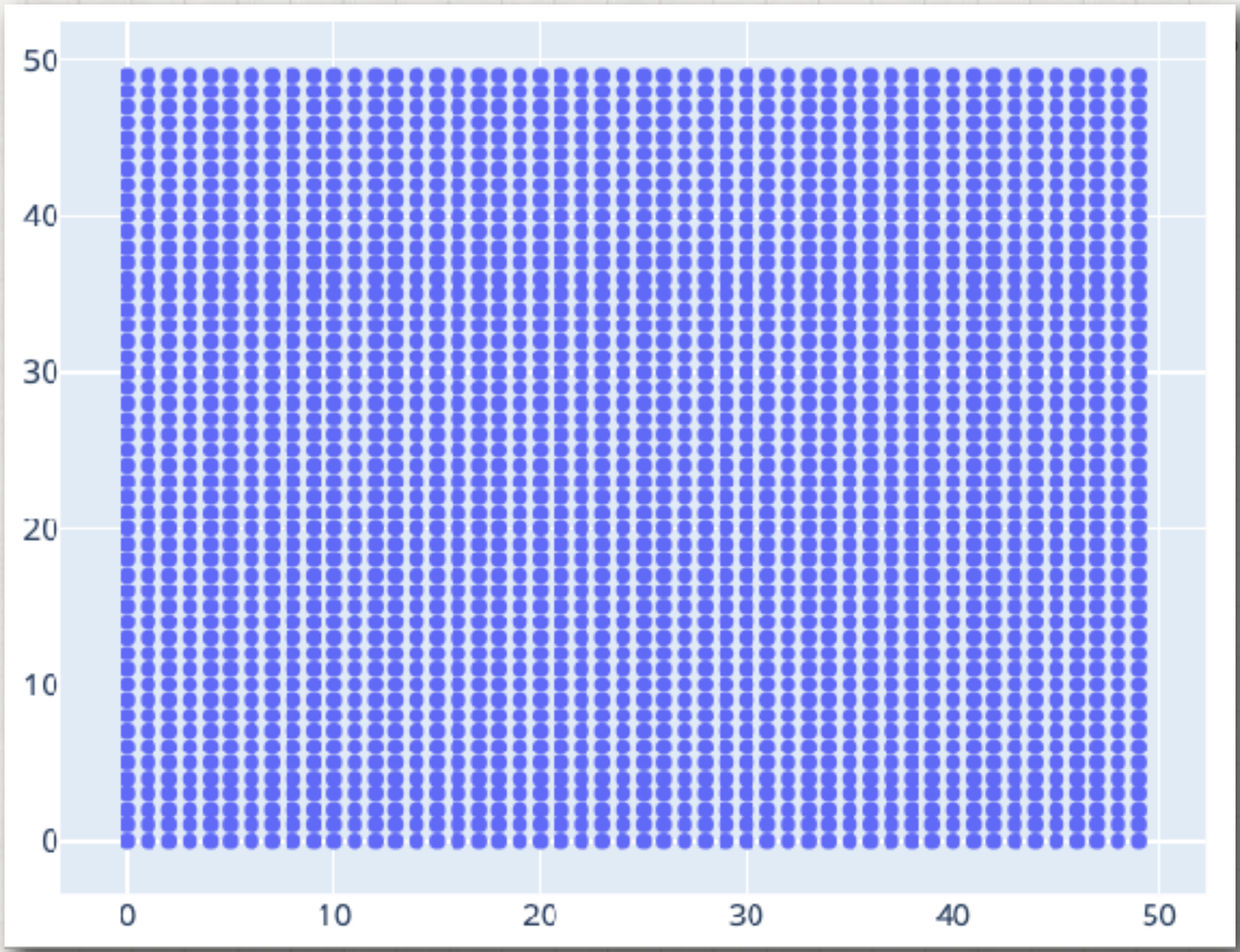
A. 9

B. 18

C. 512

D. ∞

A (SLIGHTLY) MORE COMPLEX CASE



Q: How many rows in the corresponding table?

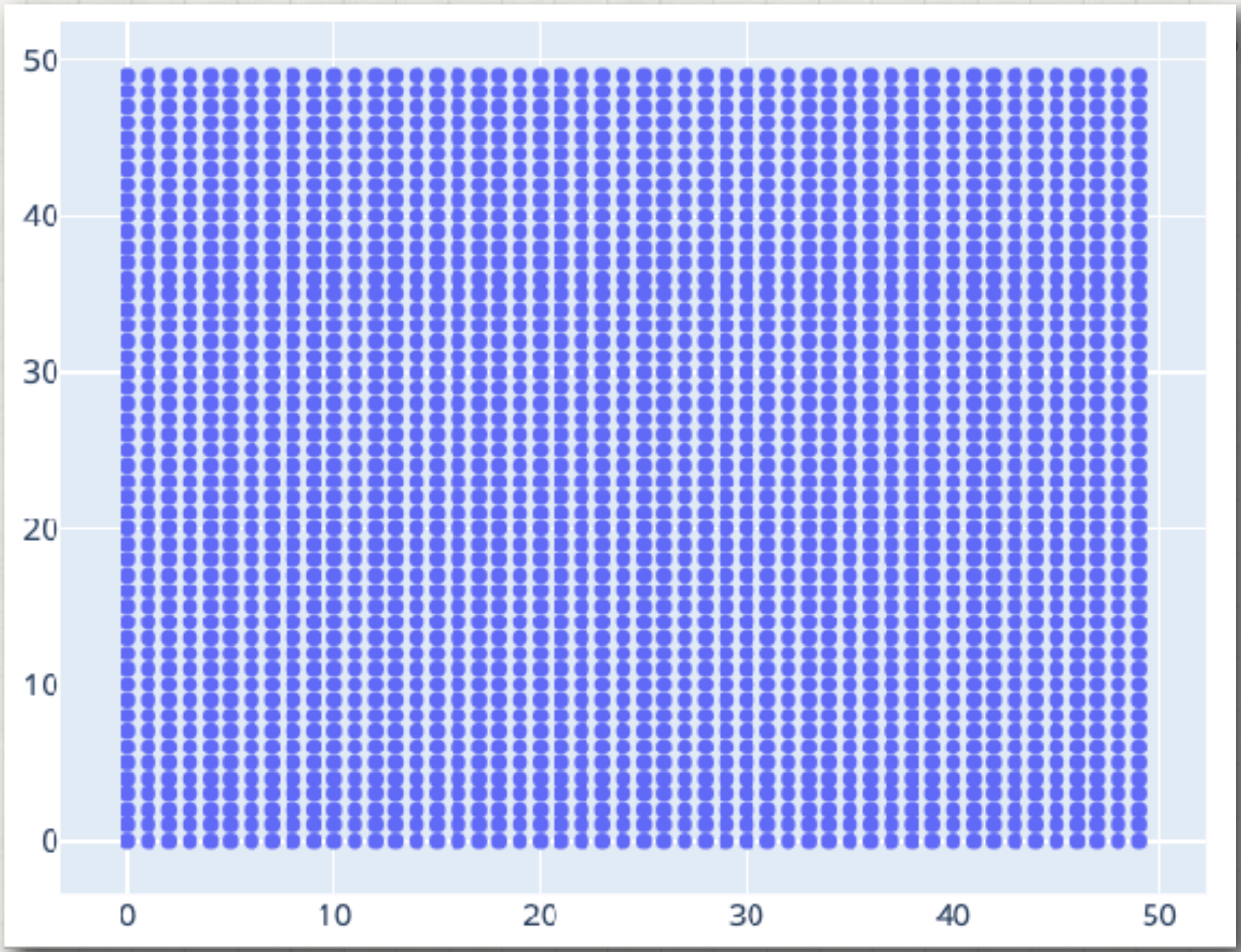
A. 2,500

B. 5,000

C. 20~30 Billion

D. None of above

A (SLIGHTLY) MORE COMPLEX CASE



Q: How many rows in the corresponding table?

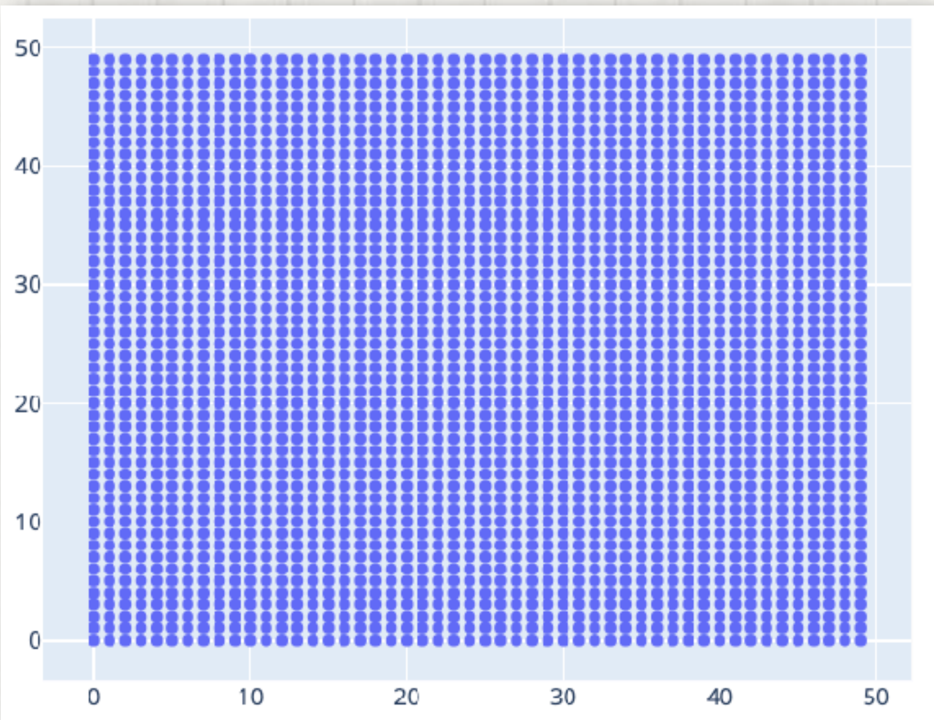
A. 2,500

B. 5,000

C. 20~30 Billion

D. None of above

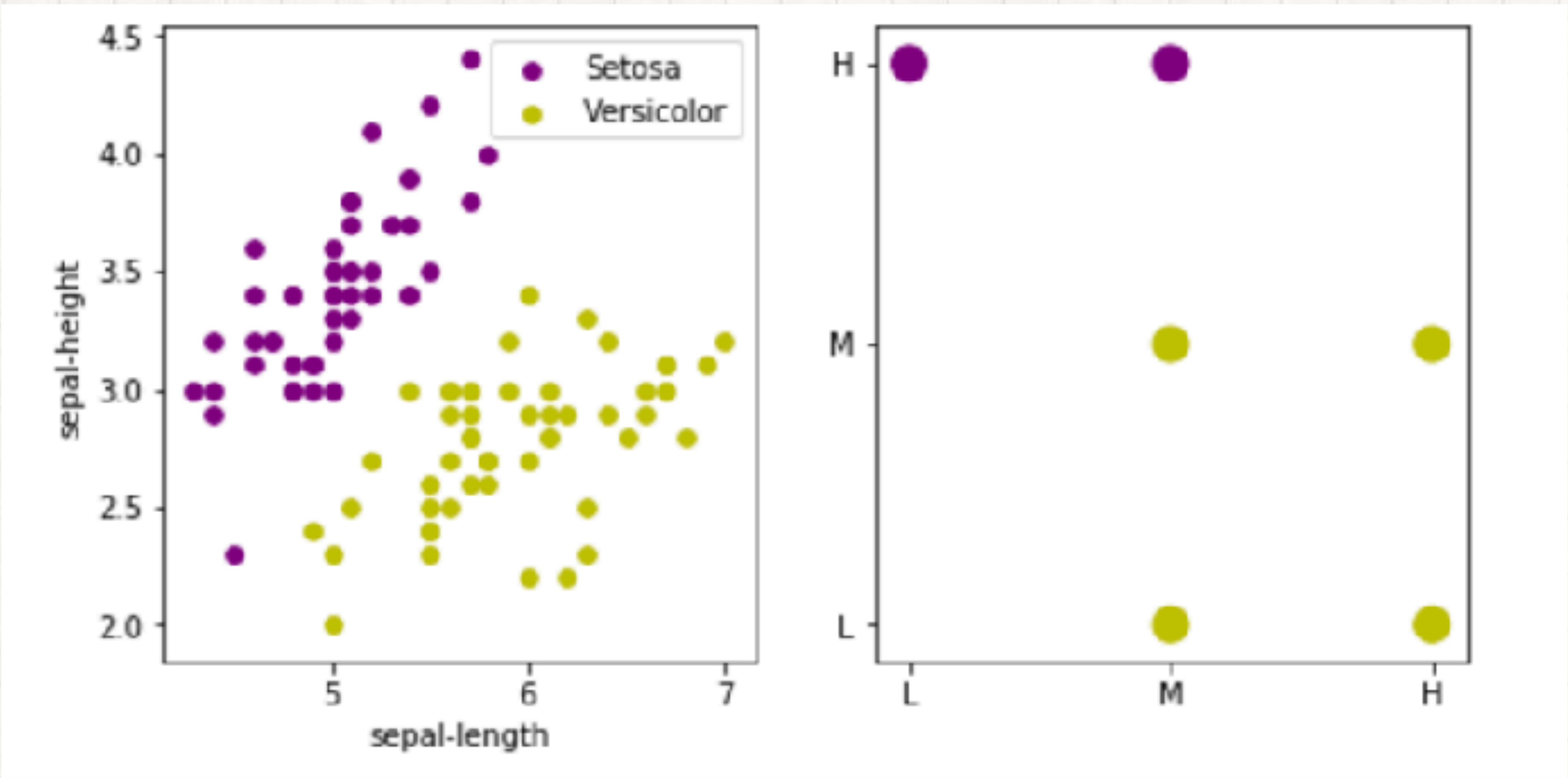
LARGE NUMBER OF POSSIBILITIES



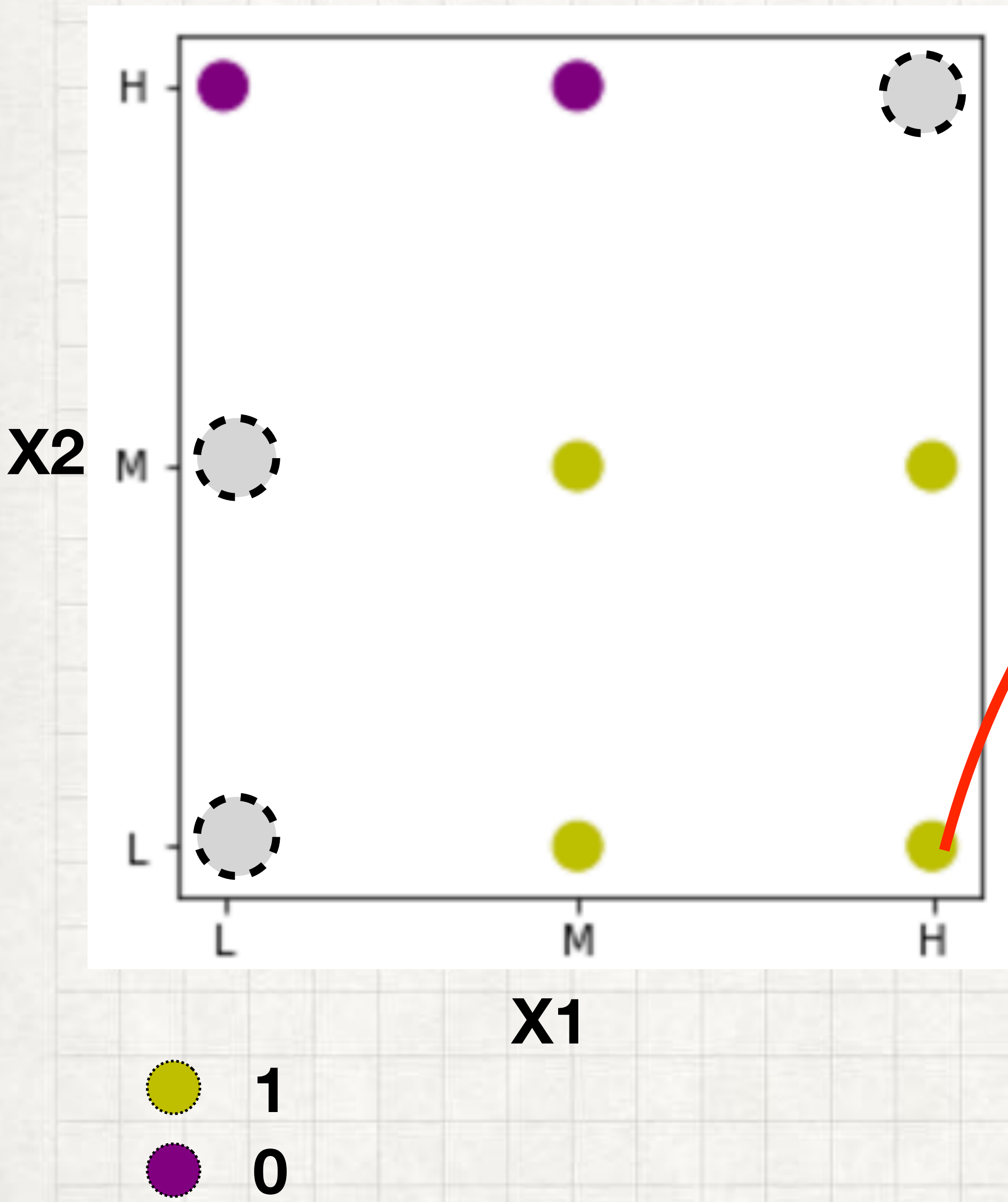
- 2^{2500}=
3758280234548012036833624189723865048677365517592586770565238397822316814983377085357327257526588443337024577495260577603092278913516177656519073109687802364646940433162365621467244164785911318325937291112215801805317492327775155799698990751422139691179948773438020494216249544022145293907816475633395350247725849016076668629825679186228496361602088773658349501637901885230262474405073903820321888923861099058697067531432439211984822120754440224333665547868565593896895856381265823772240377217022399914414660261857526515029364722809110185003203754963367499515695215418504417479258440662952796718726052857925526601307020479982183347493563216774695296825517658582675027158940078877272500707803502629523772140288422974862635978797921763382209326194895
09376

SIMPLIFIED IRIS DATA

- Now we “match” our complete table with data.

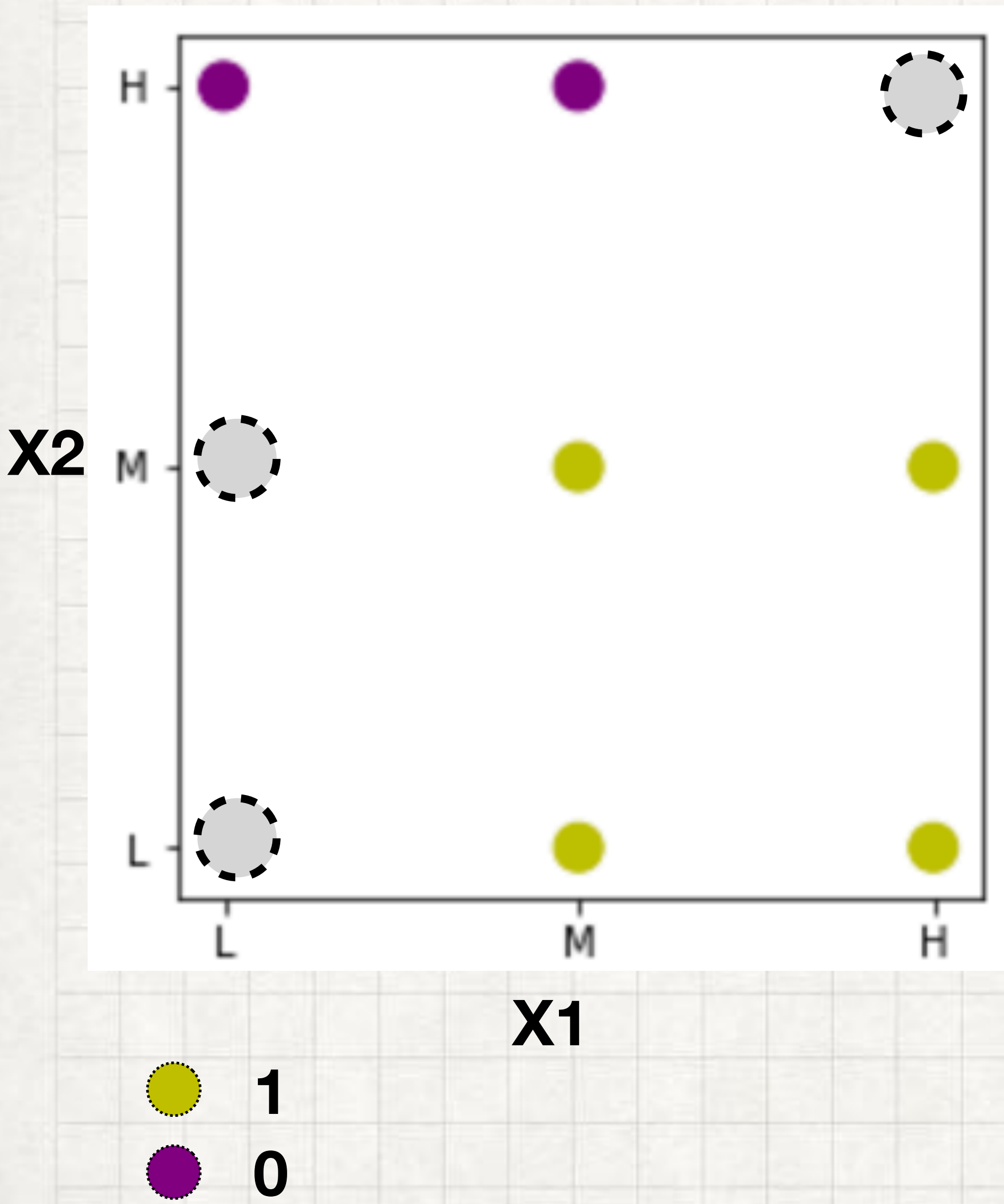


CONSISTENCY CHECK



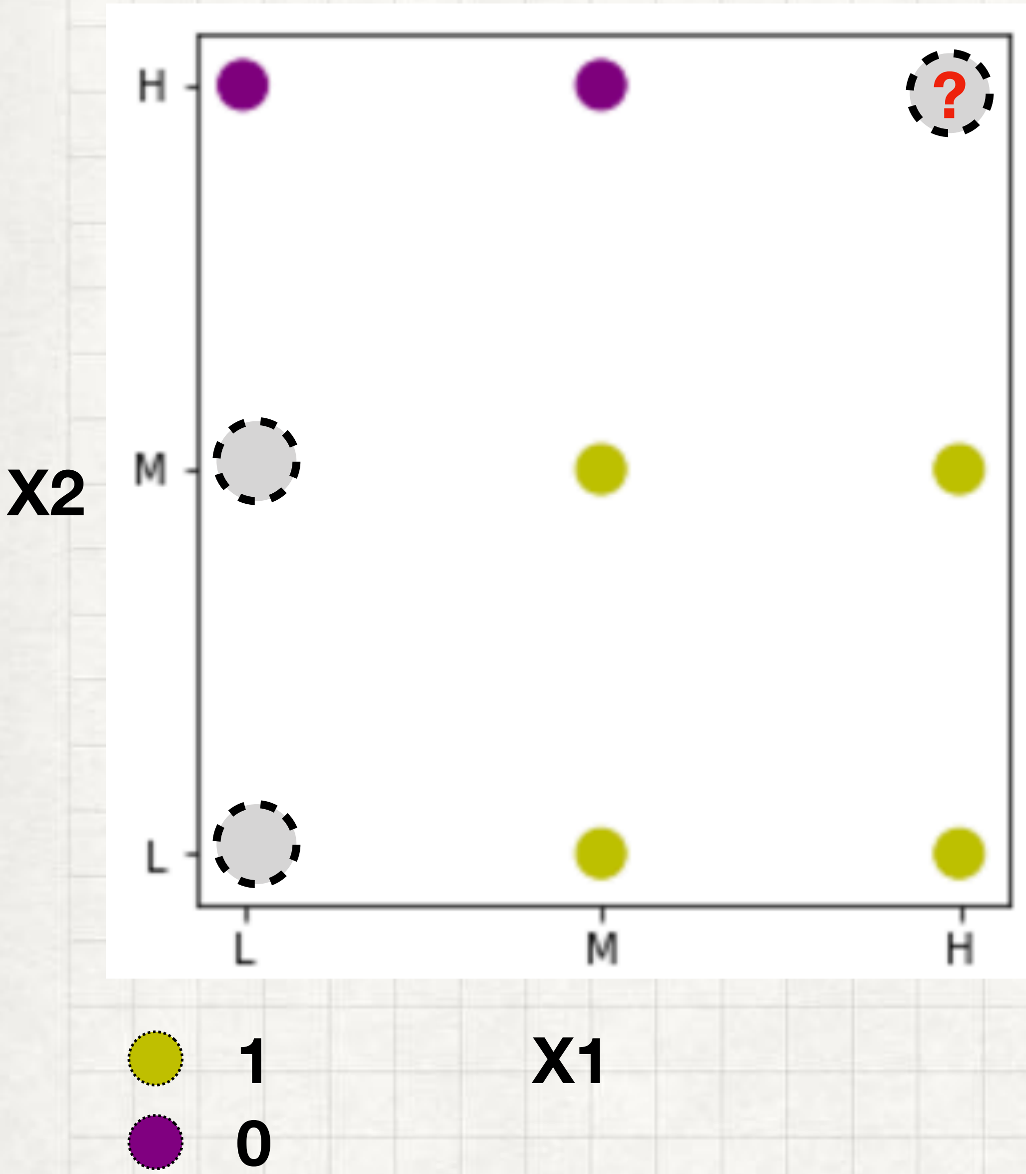
	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
		0	0	0	1	1	0	0	1	0
		0	0	0	1	1	0	1	1	0
		0	0	0	1	1	0	1	1	1
		0	1	0	1	1	0	1	1	0
		0	1	0	1	1	0	1	1	1
		1	0	0	1	1	0	1	1	0
		1	0	0	1	1	0	1	1	1
		1	1	0	1	1	0	1	1	0
		1	1	0	1	1	0	1	1	1
...										

FITTING TO DATA



	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
		0	0	0	1	1	0	0	1	0
		0	0	0	1	1	0	1	1	0
		0	0	0	1	1	0	1	1	1
		0	1	0	1	1	0	1	1	0
		0	1	0	1	1	0	1	1	1
		1	0	0	1	1	0	1	1	0
		1	0	0	1	1	0	1	1	1
		1	1	0	1	1	0	1	1	0
		1	1	0	1	1	0	1	1	1
...										

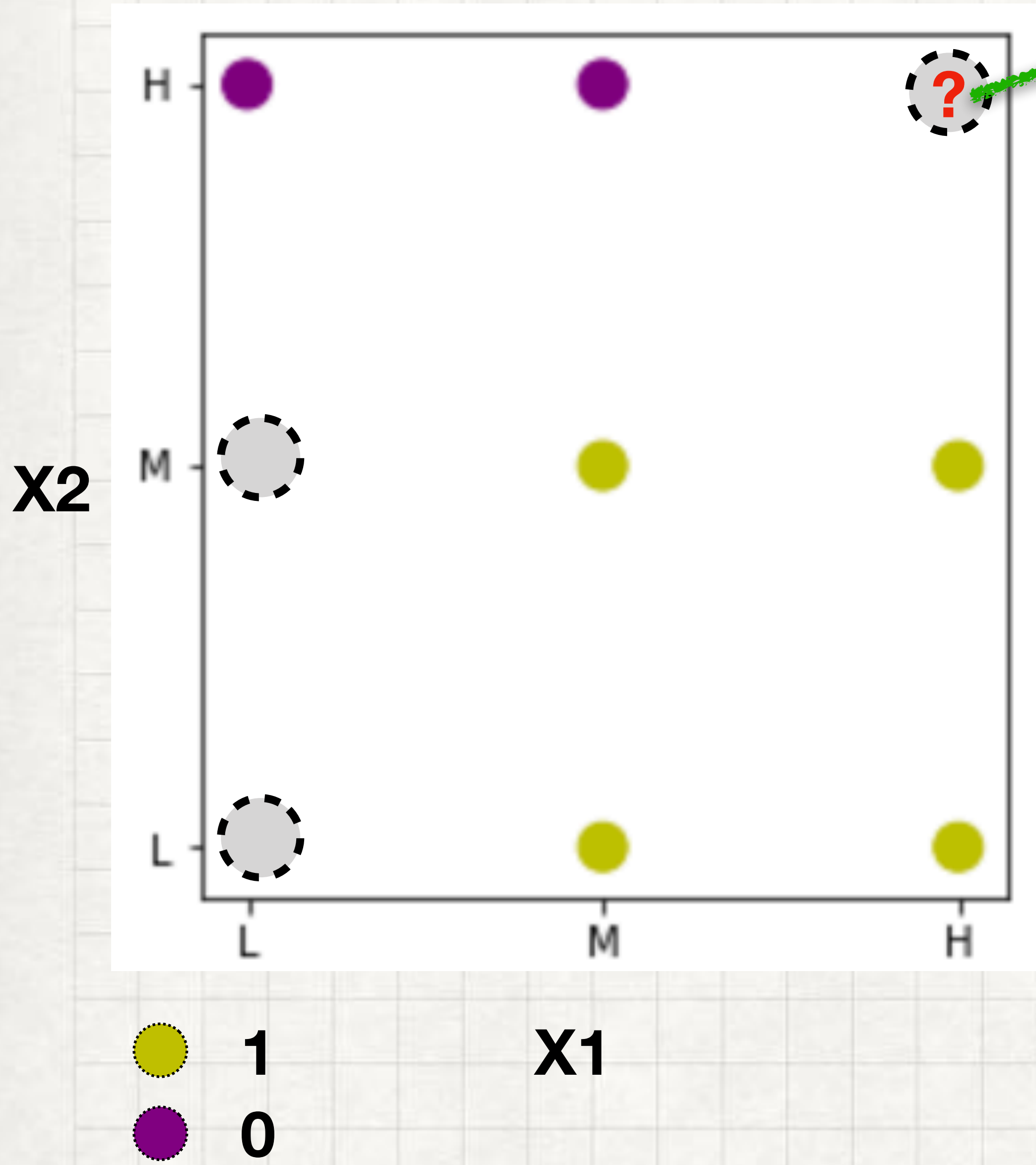
TO PREDICT A TEST DATA SAMPLE



	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
		0	0	0	1	1	0	1	1	0
		0	0	0	1	1	0	1	1	1
		0	1	0	1	1	0	1	1	0
		0	1	0	1	1	0	1	1	1
		1	0	0	1	1	0	1	1	0
		1	0	0	1	1	0	1	1	1
		1	1	0	1	1	0	1	1	0
		1	1	0	1	1	0	1	1	1

- Q: What is the prediction on this unseen data? (X1=H, X2=H)
- A. 0
- B. 1
- C. undecided.

TO PREDICT A TEST DATA SAMPLE



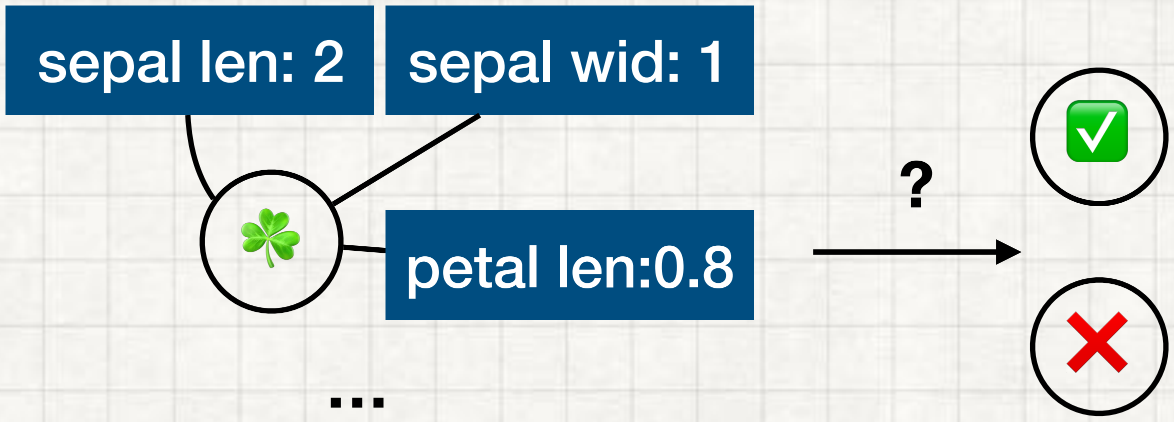
	X1	L	L	L	M	M	M	H	H	H
Y|X	X2	L	M	H	L	M	H	L	M	H
		0	0	0	1	1	0	1	1	0
		0	0	0	1	1	0	1	1	1
		0	1	0	1	1	0	1	1	0
		0	1	0	1	1	0	1	1	1
		1	0	0	1	1	0	1	1	0
		1	0	0	1	1	0	1	1	1
		1	1	0	1	1	0	1	1	0
		1	1	0	1	1	0	1	1	1

- Q: What is the prediction on this unseen data? (X1=H, X2=H)
- A. 0
- B. 1
- C. undecided.

3. LEARNING WITH HYPOTHESES

LEARNABLE CONCEPTS

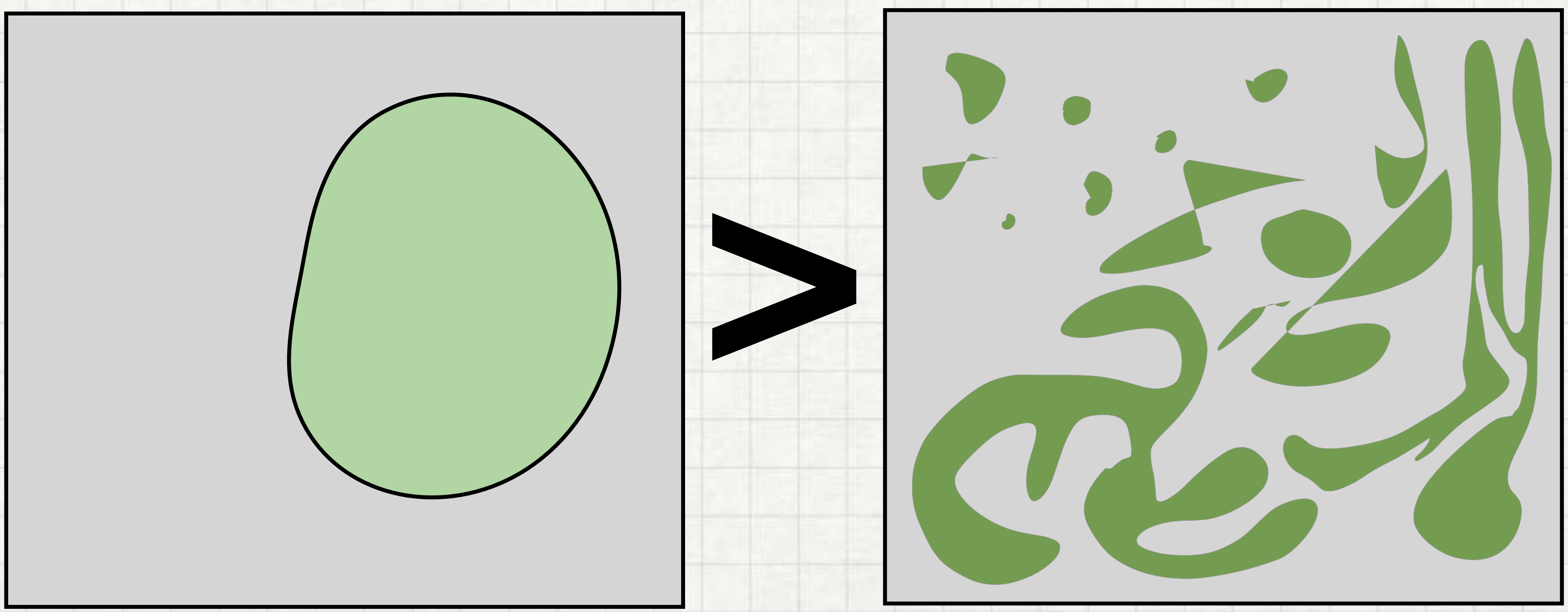
- Determine the X-Space
- The concept in $\mathscr{X}$



NOT ALL CONCEPTS ARE LEARNABLE

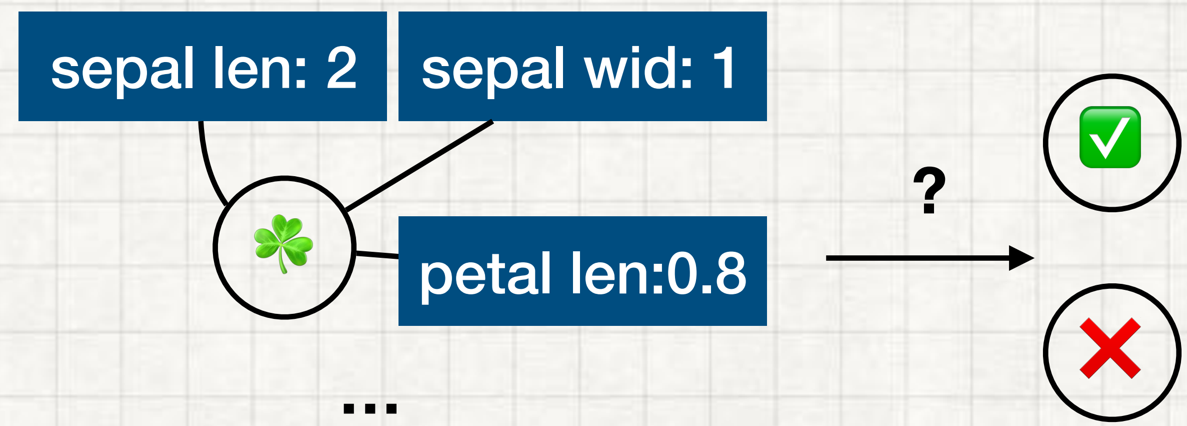
DO NOT TRY TO LEARN ALL CONCEPTS

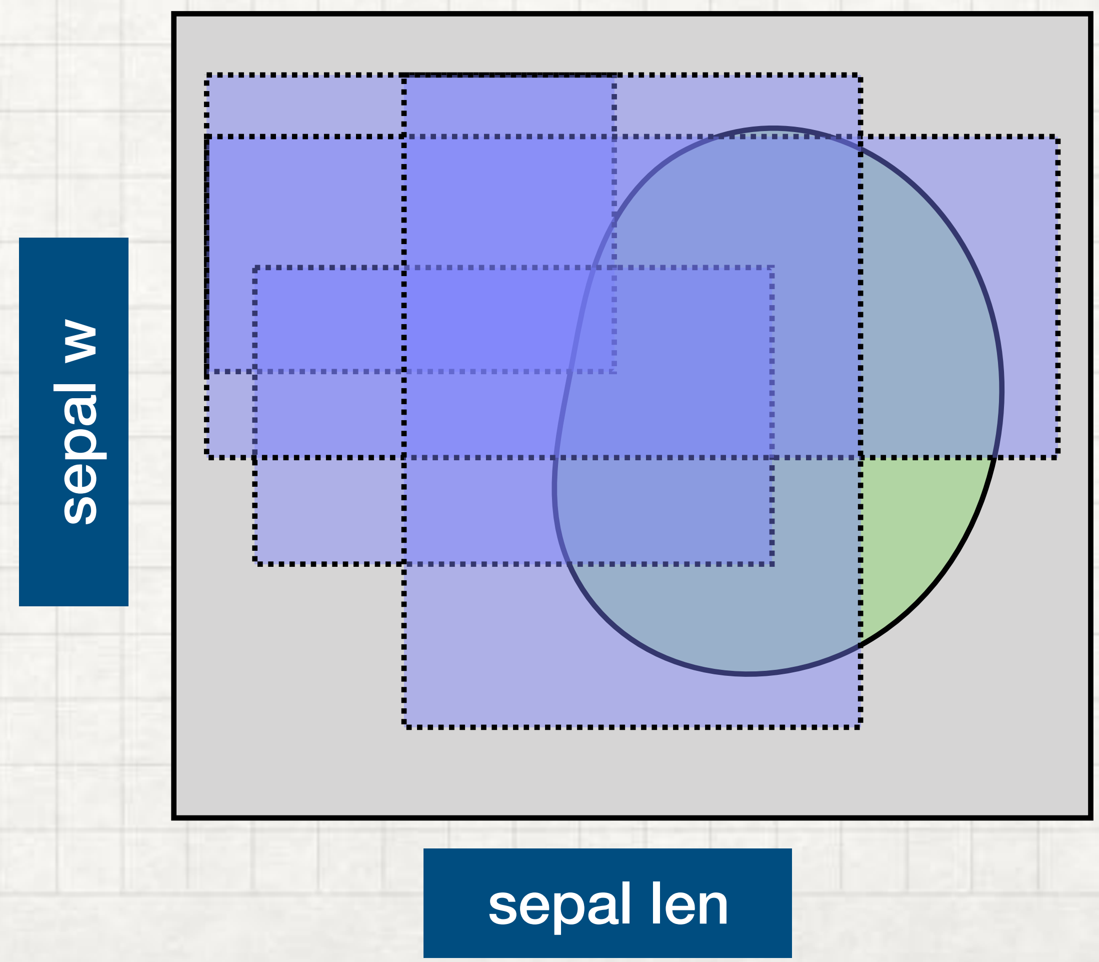
CONCEPT LEARNING ESSENTIALS – PREFERRING REGULARITY



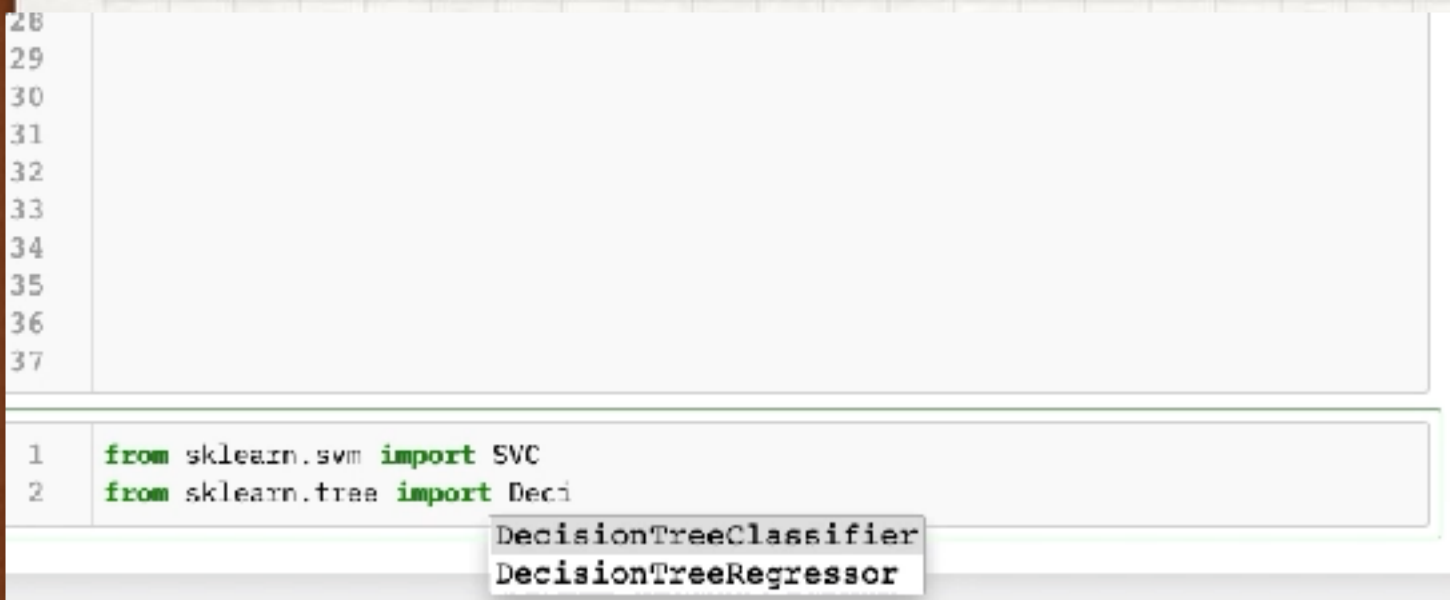
CONCEPT LEARNING ESSENTIALS – (BIASED) HYPOTHESES

- Determine the X-Space
- The concept in $\mathscr{X}$
- Your hypotheses about learnable concepts
  - Preference about the UNKNOWN concept
  - Mostly WRONG — i.e. you could not find a single hypothesis matches the concept perfectly!



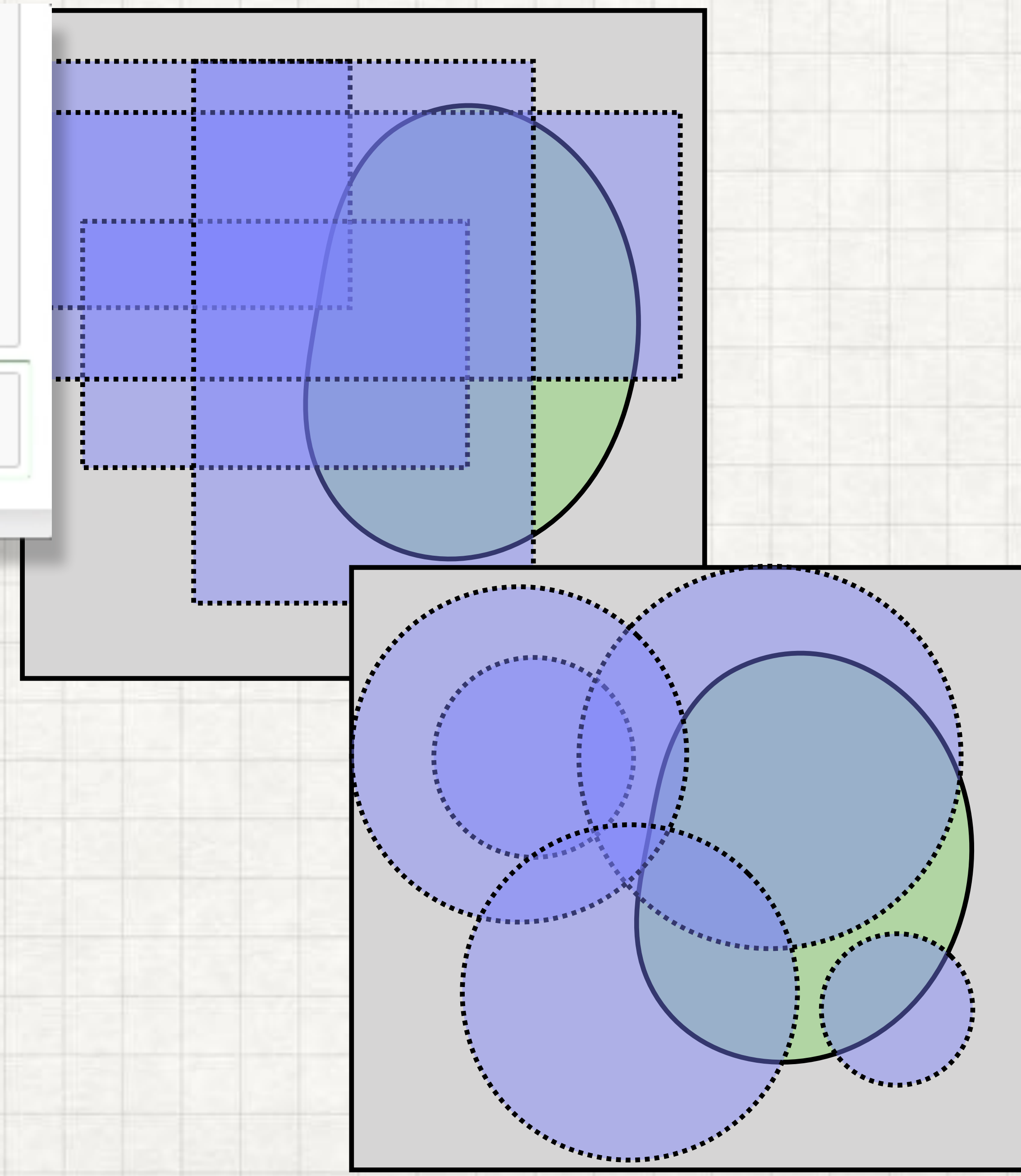


HYPOTHESES IN PRACTICE

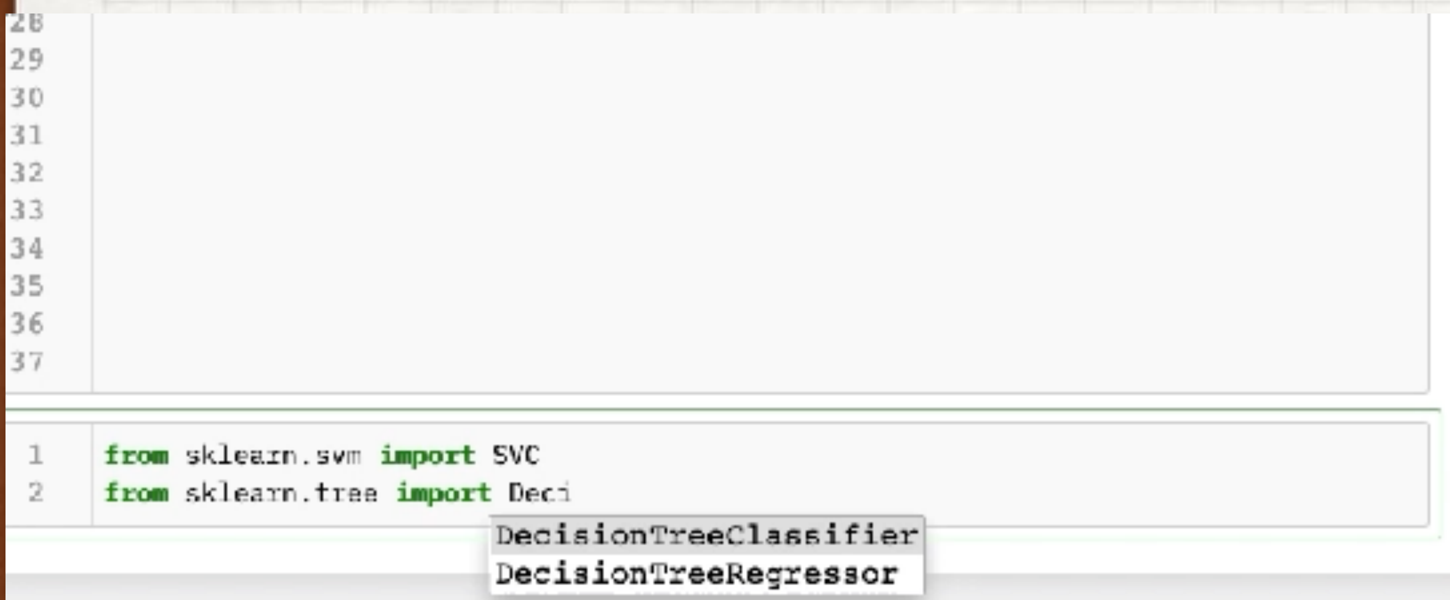


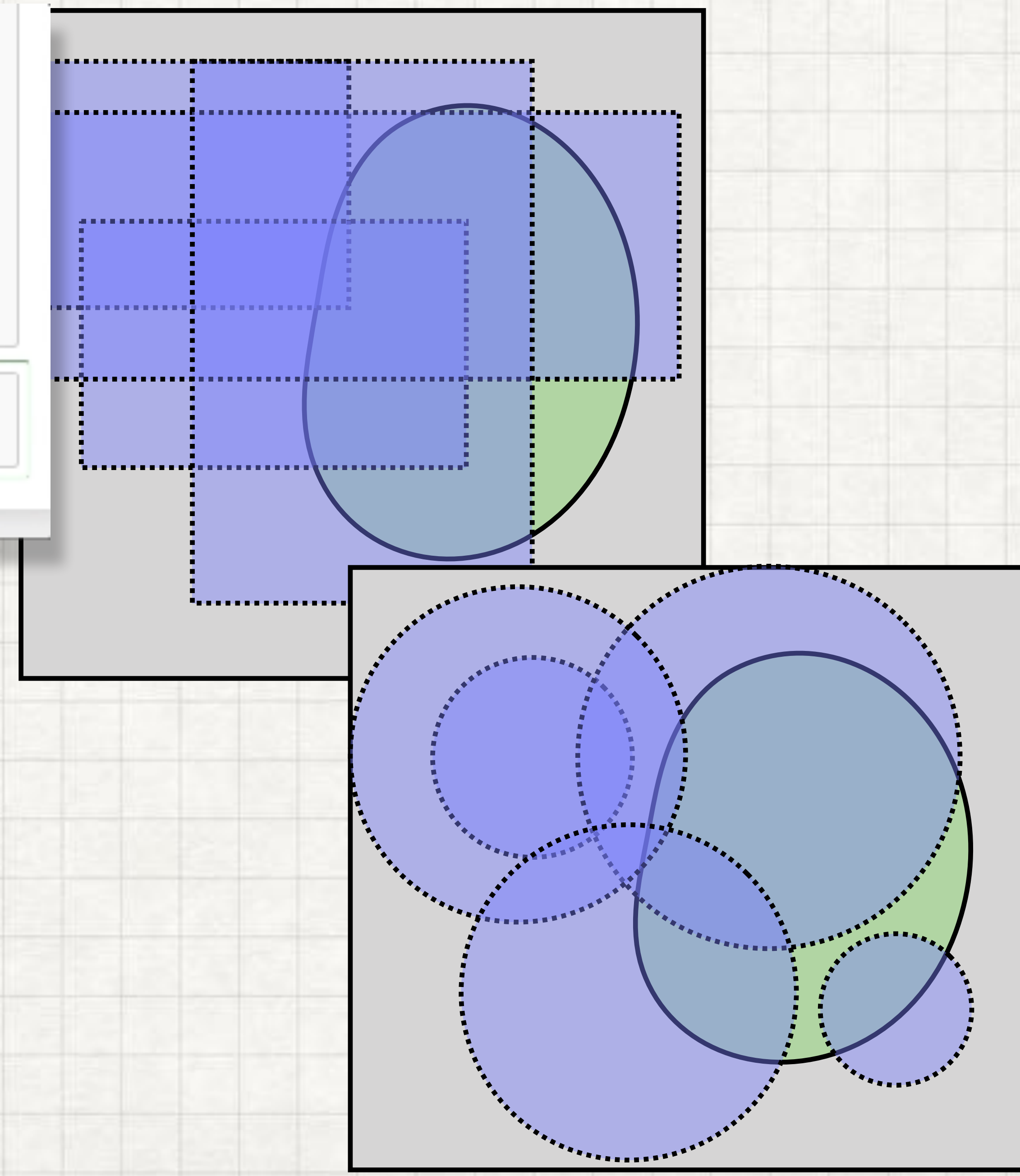
he
ot

i.e. you
gle
s the



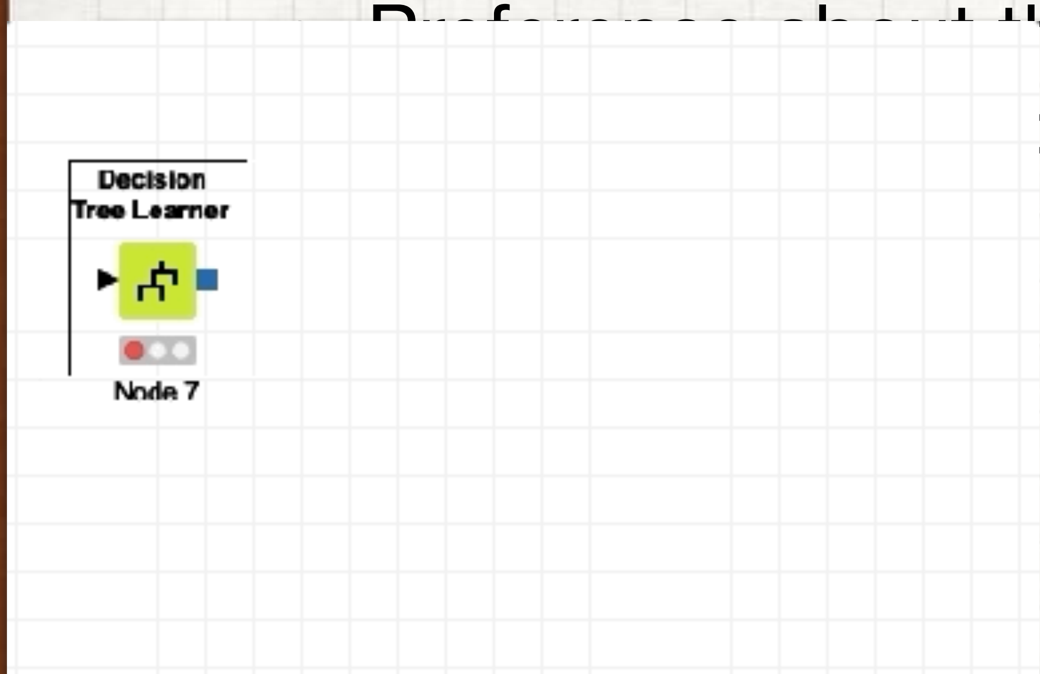
HYPOTHESES IN PRACTICE





Preference about the
ot

i.e. you
gle
s the



HYPOTHESIS SET AND CONCEPT

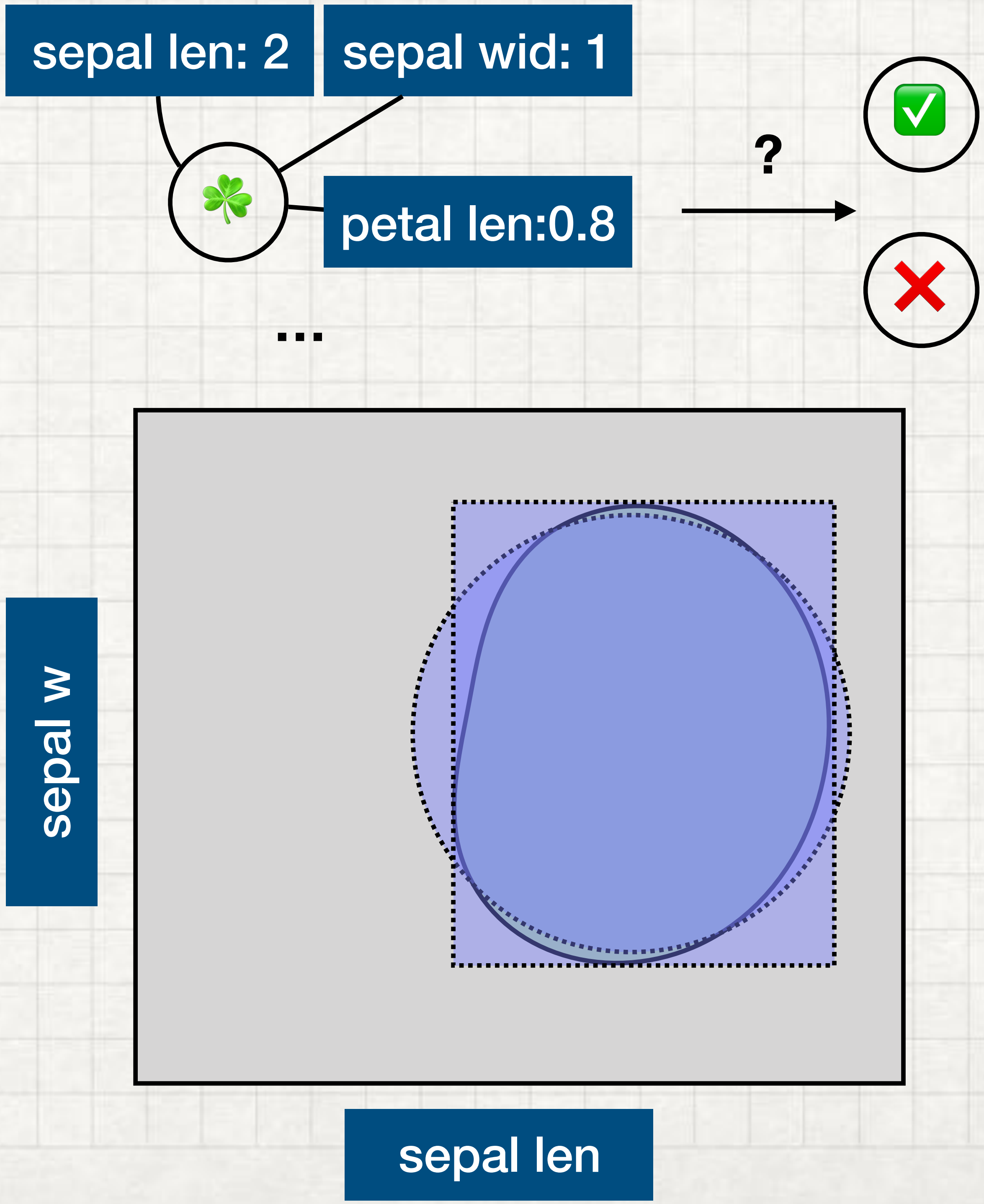
- **Q**: Which of the following understanding of the hypothesis set is correct?

A. A hypothesis set is a principled mathematical notion, it leads to the precise target concept.

B. Hypotheses are subjective, we should not use them if we have sufficient data.

C. In most cases, the hypothesis set does not contain the precise target concept. But the error is inevitable, it is the price for generalisation.

D. We can only decide what hypothesis family to use after seeing the data.



HYPOTHESIS SET AND CONCEPT

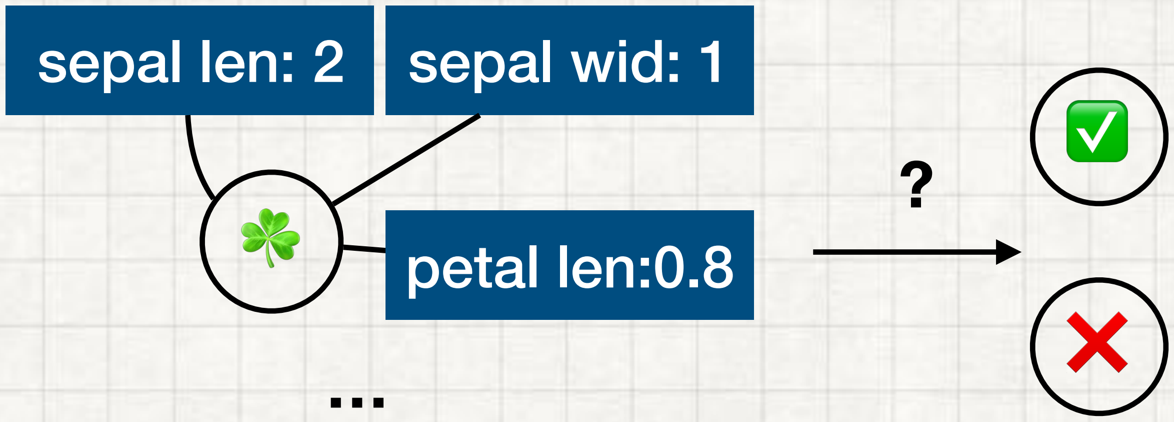
- **Q**: Which of the following understanding of the hypothesis set is correct?

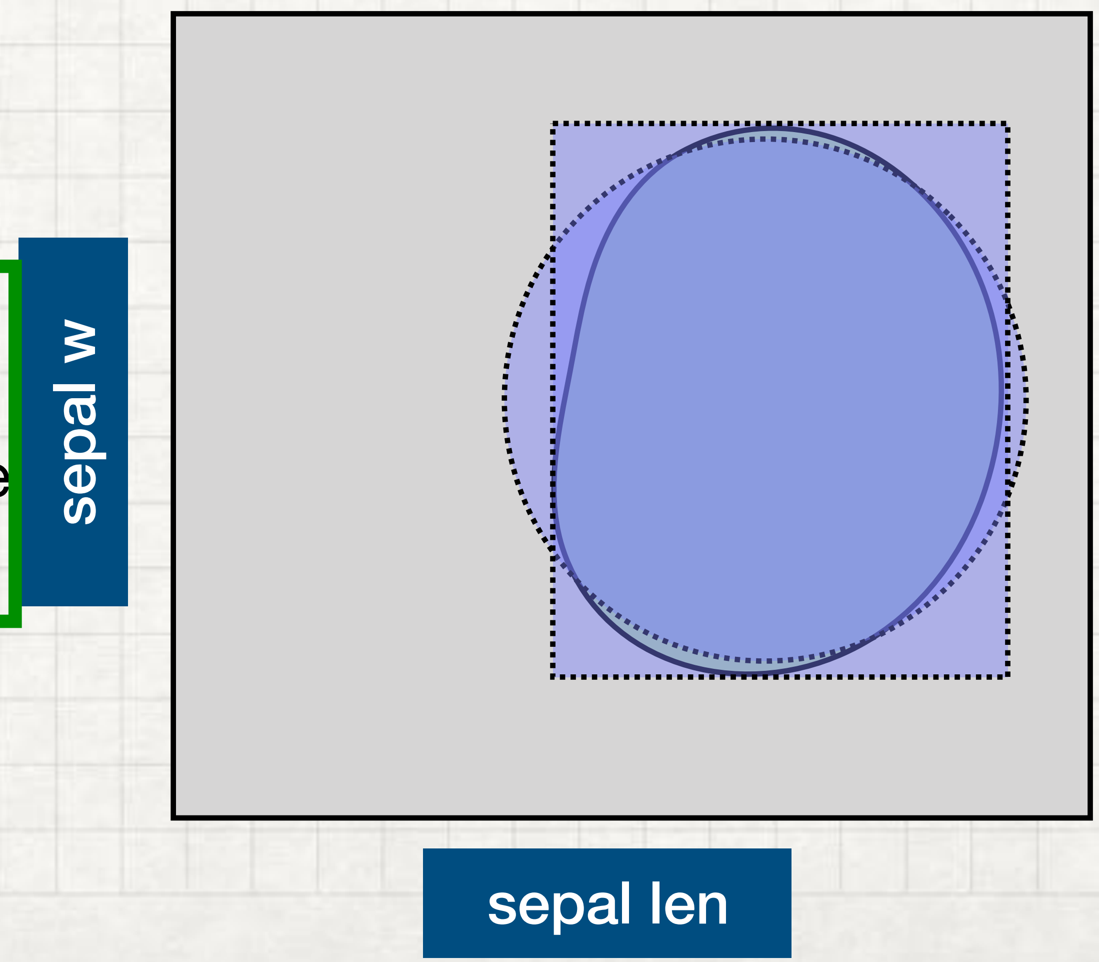
A. A hypothesis set is a principled mathematical notion, it leads to the precise target concept.

B. Hypotheses are subjective, we should not use them if we have sufficient data.

C. In most cases, the hypothesis set does not contain the precise target concept. But the error is inevitable, it is the price for generalisation.

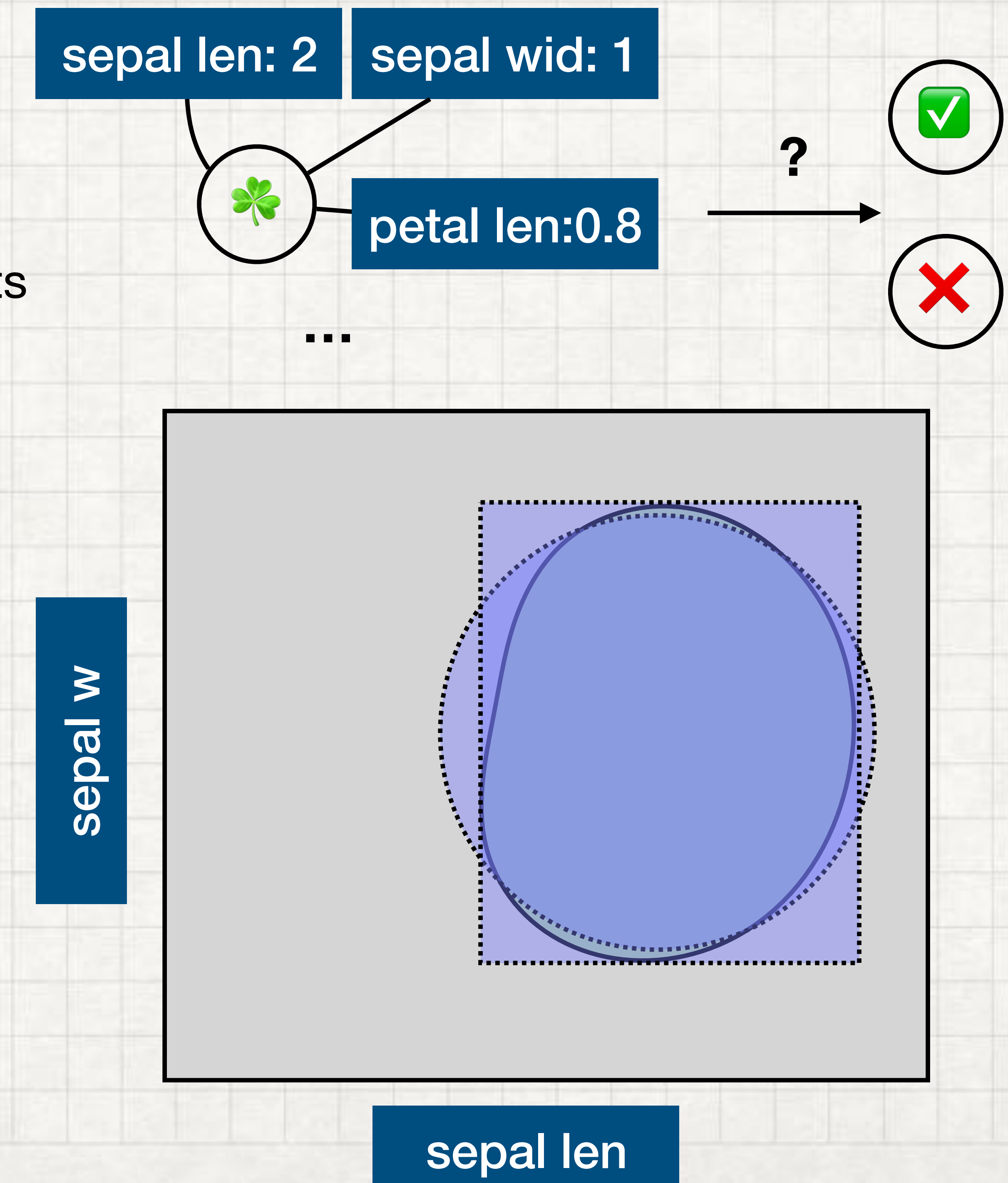
D. We can only decide what hypothesis family to use after seeing the data.





HYPOTHESIS SET AND CONCEPT

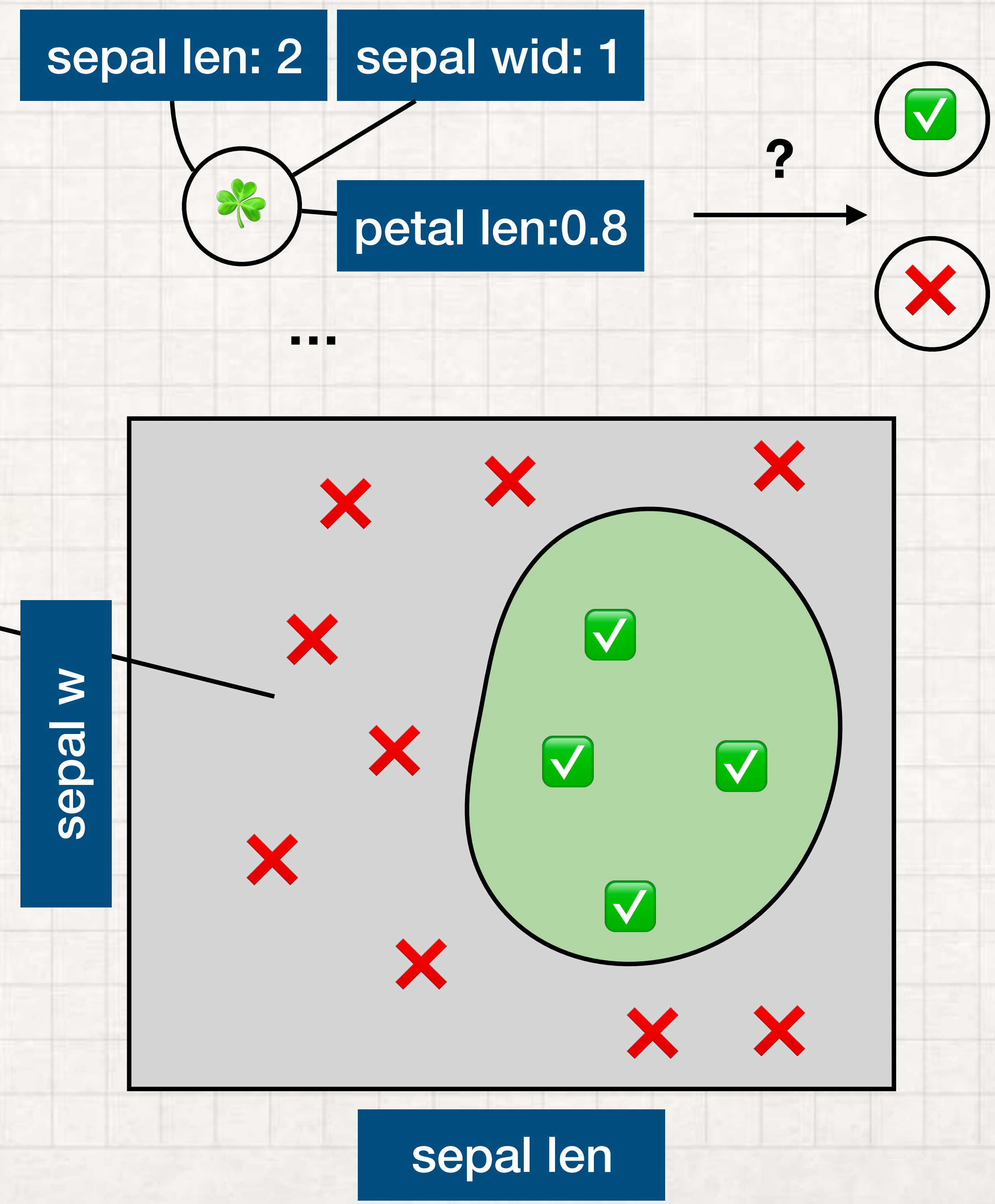
- Determine the X-Space
- The concept in ***X***
- Your hypotheses about learnable concepts
  - Some family may perform better than the others. **You don't know now.**
  - SUBJECT TO
    - the concept / ***X***
    - how you evaluate "performance"
    - learning algorithm (below)
    - data availability (below)



ENCOUNTER DATA

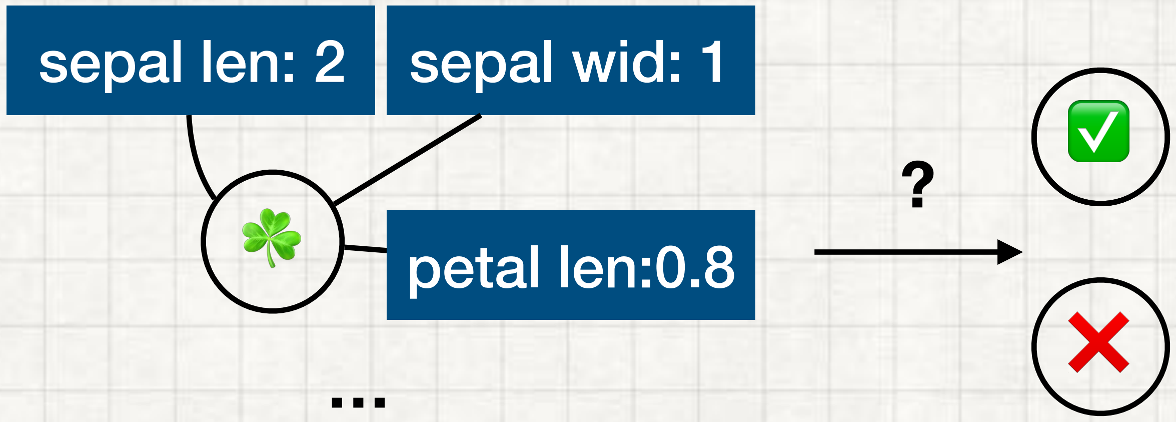
- Determine the X-Space
- The concept in $\mathscr{X}$
- Your hypotheses

Now Data Joins the Game!



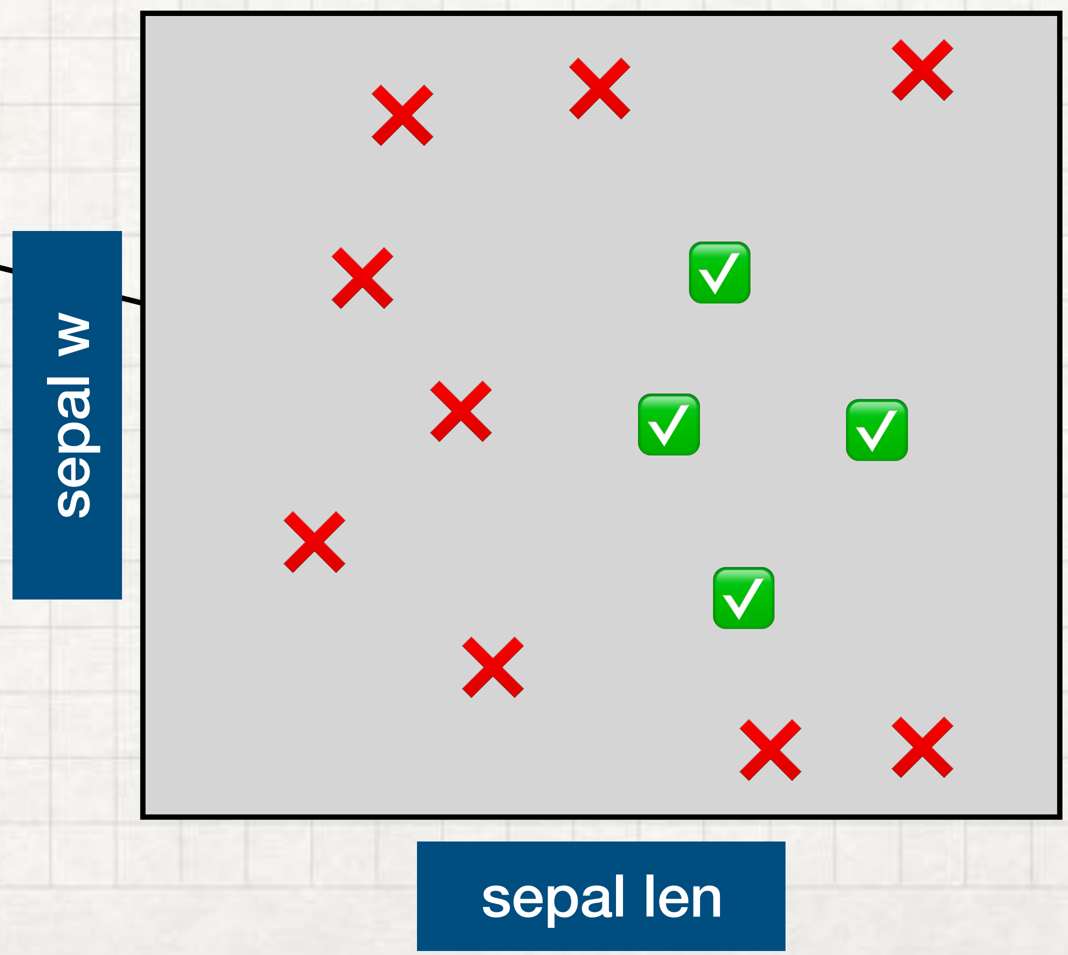
ENCOUNTER DATA

- Determine the X-Space
- The concept in $\mathscr{X}$
- Your hypotheses



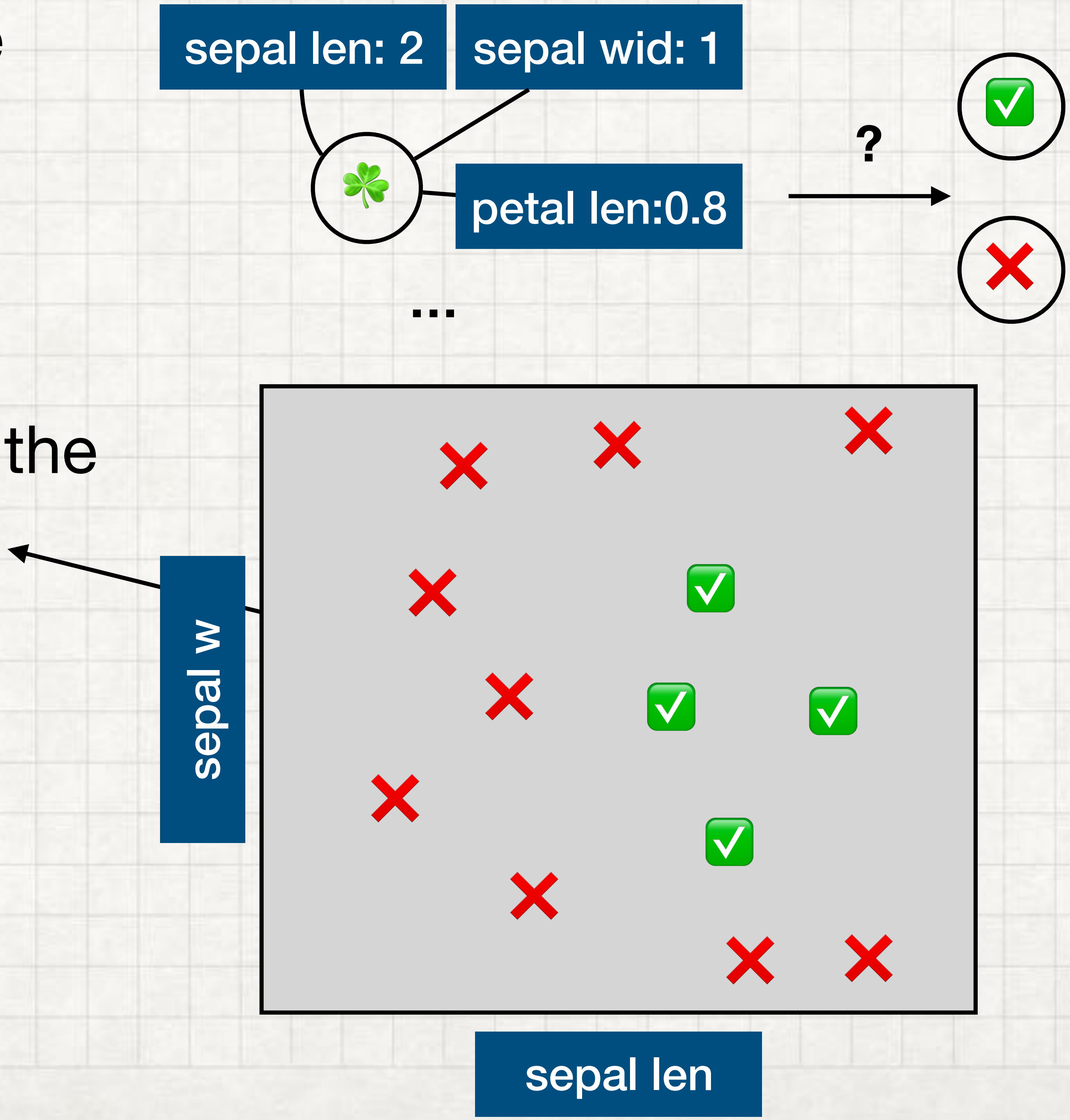
Now Data Joins the Game!

Concept not accessible!



DATA IN WORK

- Determine the X-Space
- The concept in $\mathscr{X}$
- Your hypothesis set — presumably containing the target concepts
- Data decides THE one hypothesis.



REPRESENTING A HYPOTHESIS

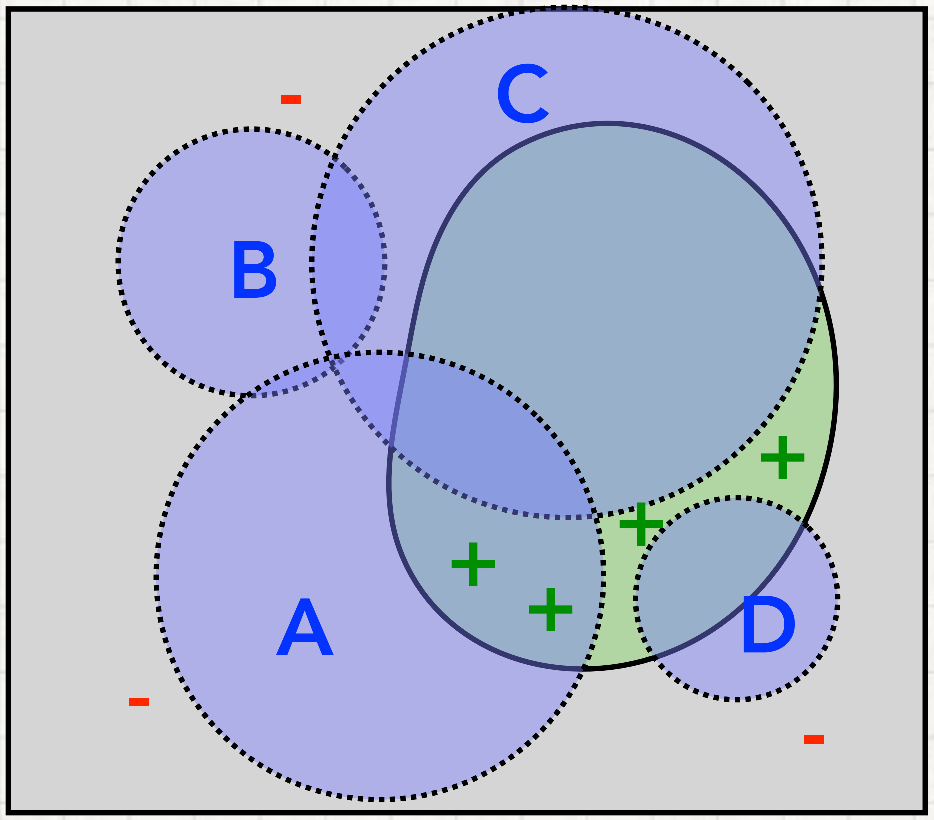
- Concept description by contour of functions

$$c : \{x \,|\, f(x) < 0\}$$

4. LEARNING FROM DATA [NEXT]

LEARNING IS TO SELECT

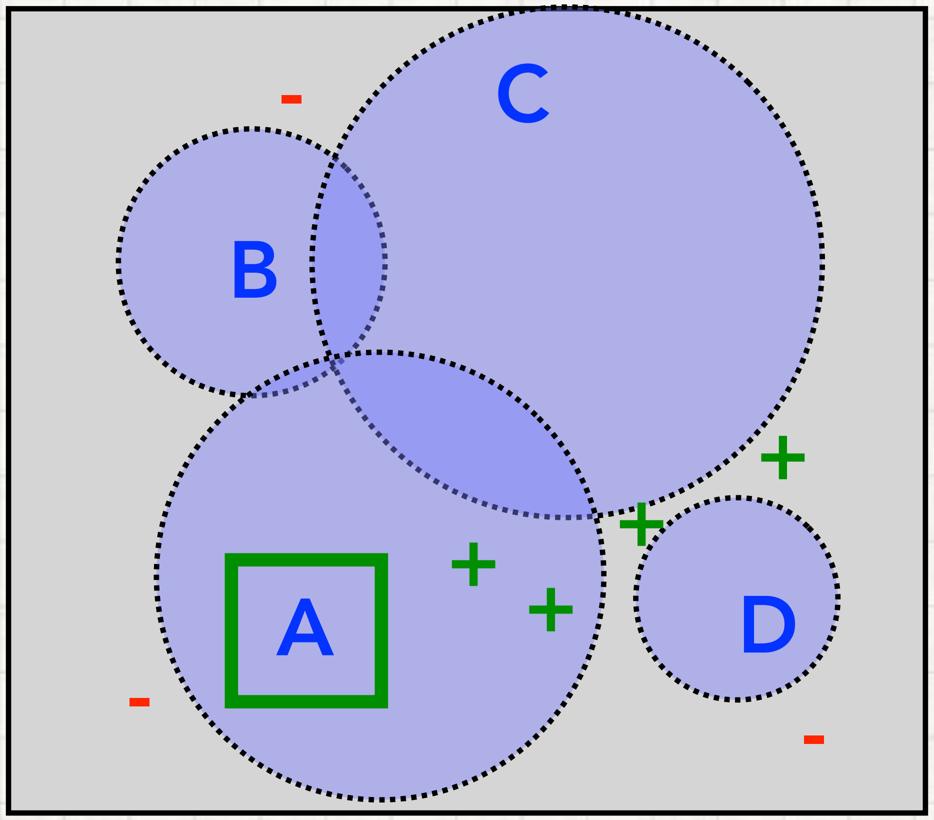
- Rather than "to come up with ideas" about the target concept, learning is to select a member hypothesis from a specified family as the estimation of the target.



- **Q**: Which hypothesis will be selected?

LEARNING IS TO SELECT

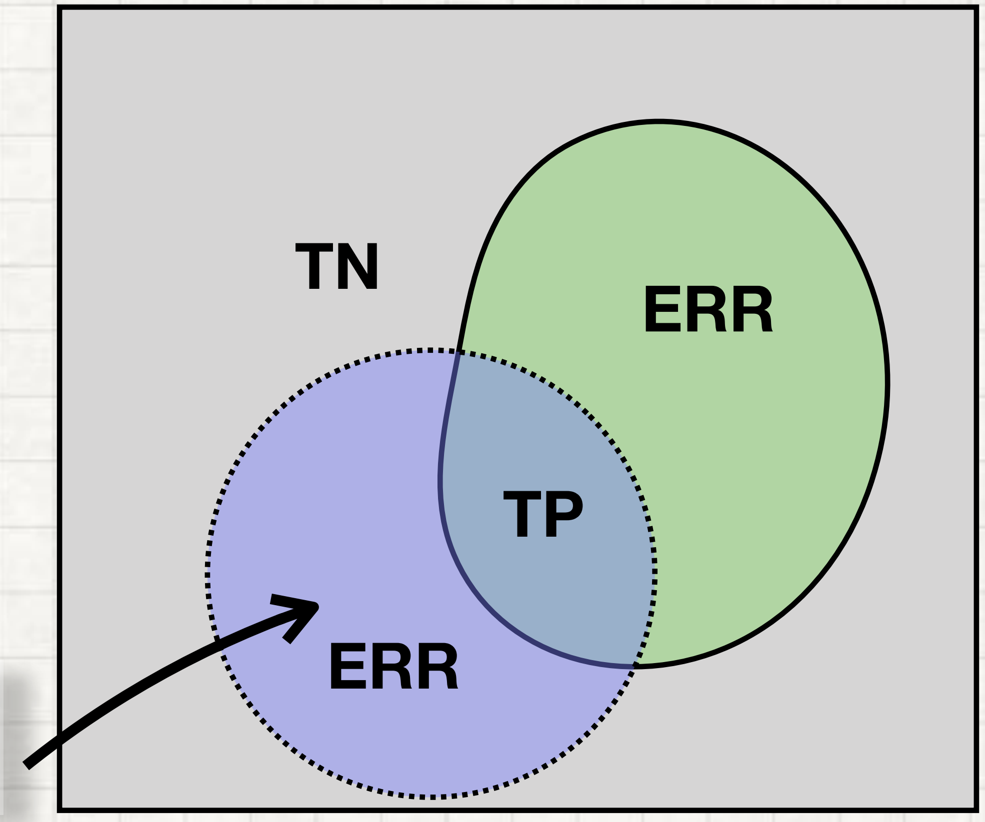
- Rather than "to come up with ideas" about the target concept, learning is to select a member hypothesis from a specified family as the estimation of the target.



- Q: Which hypothesis will be selected?
- A. Remember that we don't have access to the concept. The selection is according to data alone.

WHAT IS LEARNED?

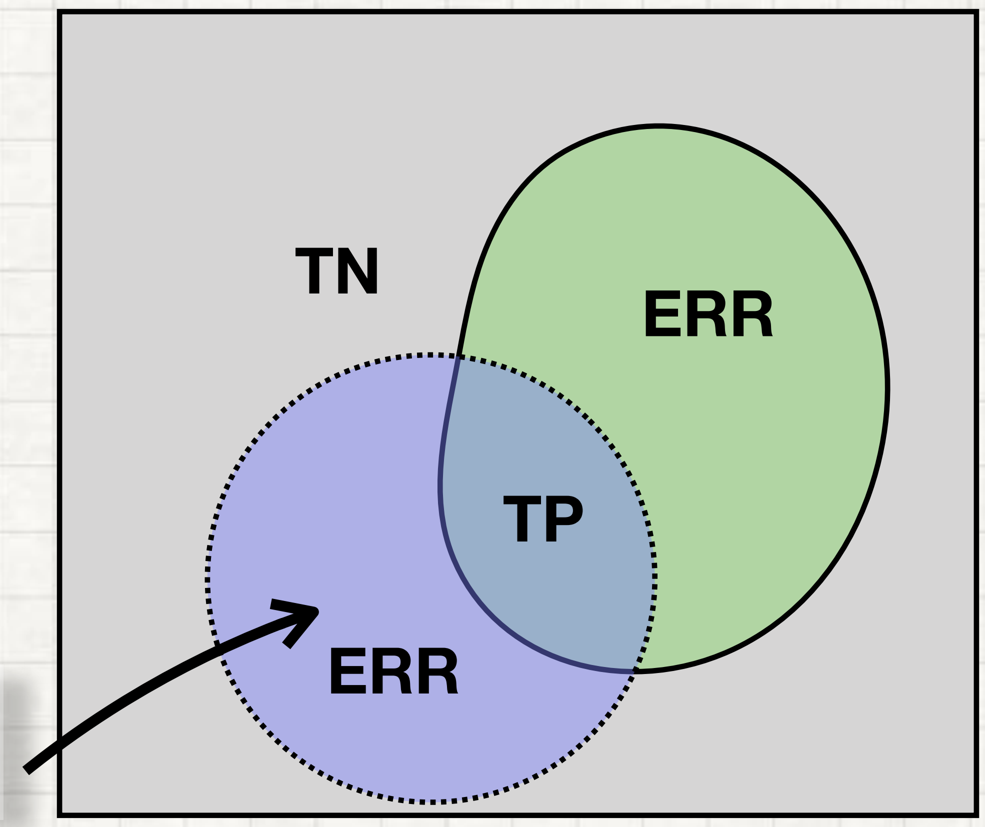
- Errors a hypothesis made on training samples reflects the “risk” of employing it in future.



- **Q**: What type of error?

WHAT IS LEARNED?

- Errors a hypothesis made on training samples reflects the "risk" of employing it in future.



- **Q**: What type of error?
- False positive.