Lecture 2
Machine learning framework: terms, definitions, jargon

Serena Yeung
BIODS 388

Machine Learning

- A family of statistical and mathematical modeling techniques that uses a variety of approaches to automatically learn and improve the prediction of a target objective, without explicit programming.

Machine Learning

- A family of statistical and mathematical modeling techniques that uses a variety of approaches to automatically learn and improve the prediction of a target objective, without explicit programming.

- Concisely: systems that improve their performance in a given task, through exposure to experience, or data.

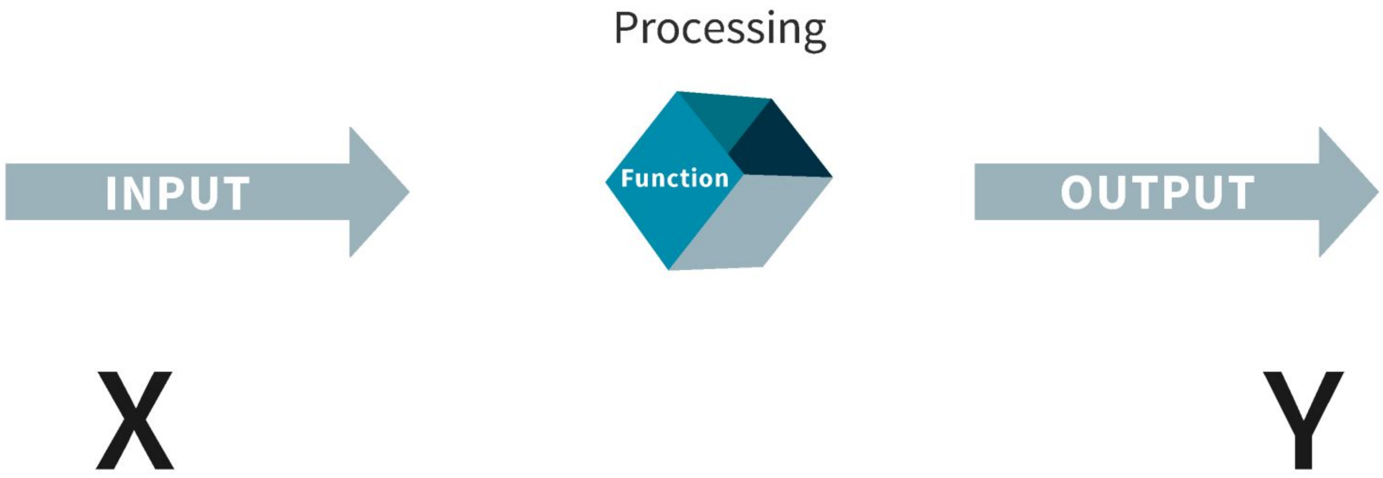
Different paradigms of machine learning

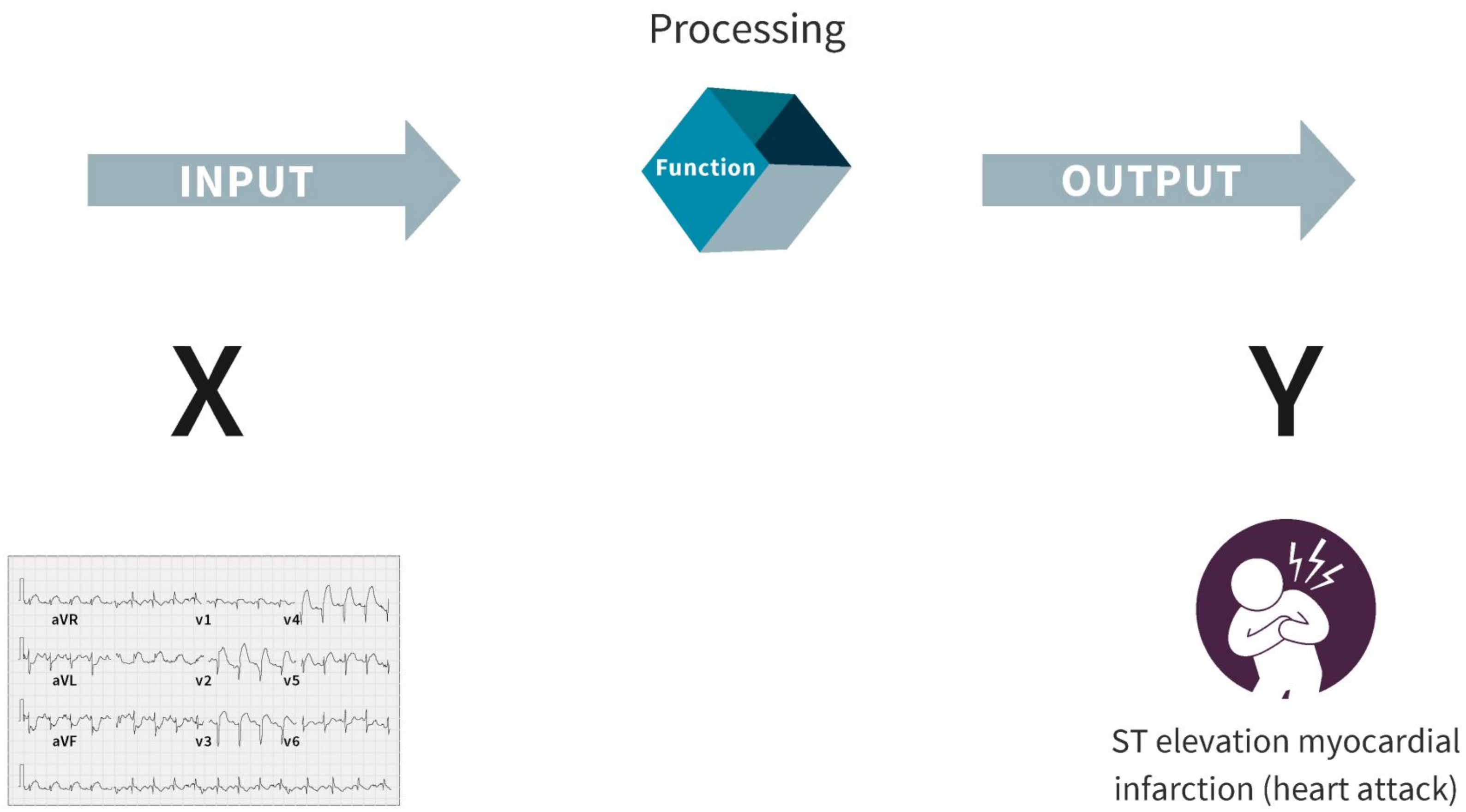
- Supervised learning
- Unsupervised learning
- Reinforcement learning

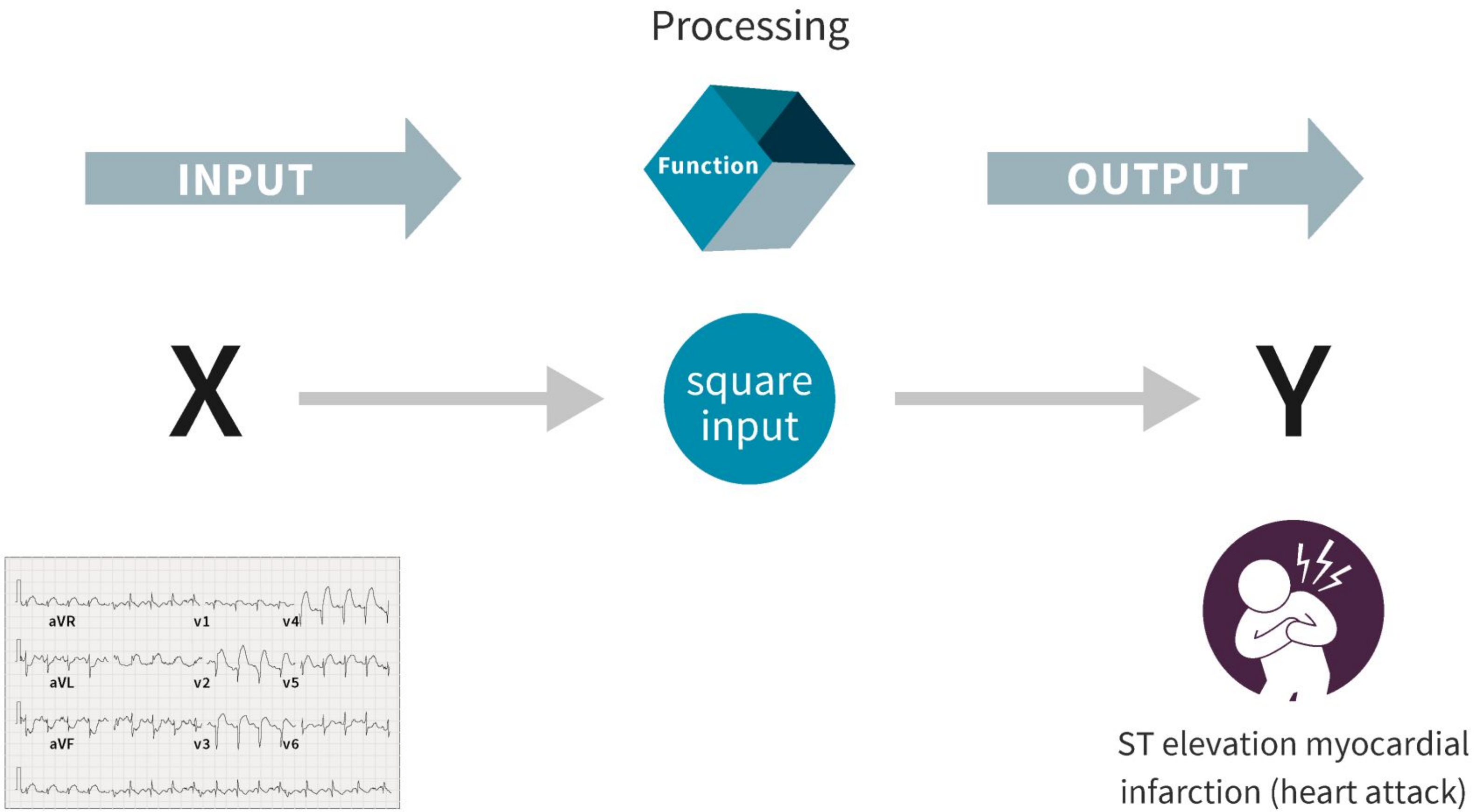
Different paradigms of machine learning

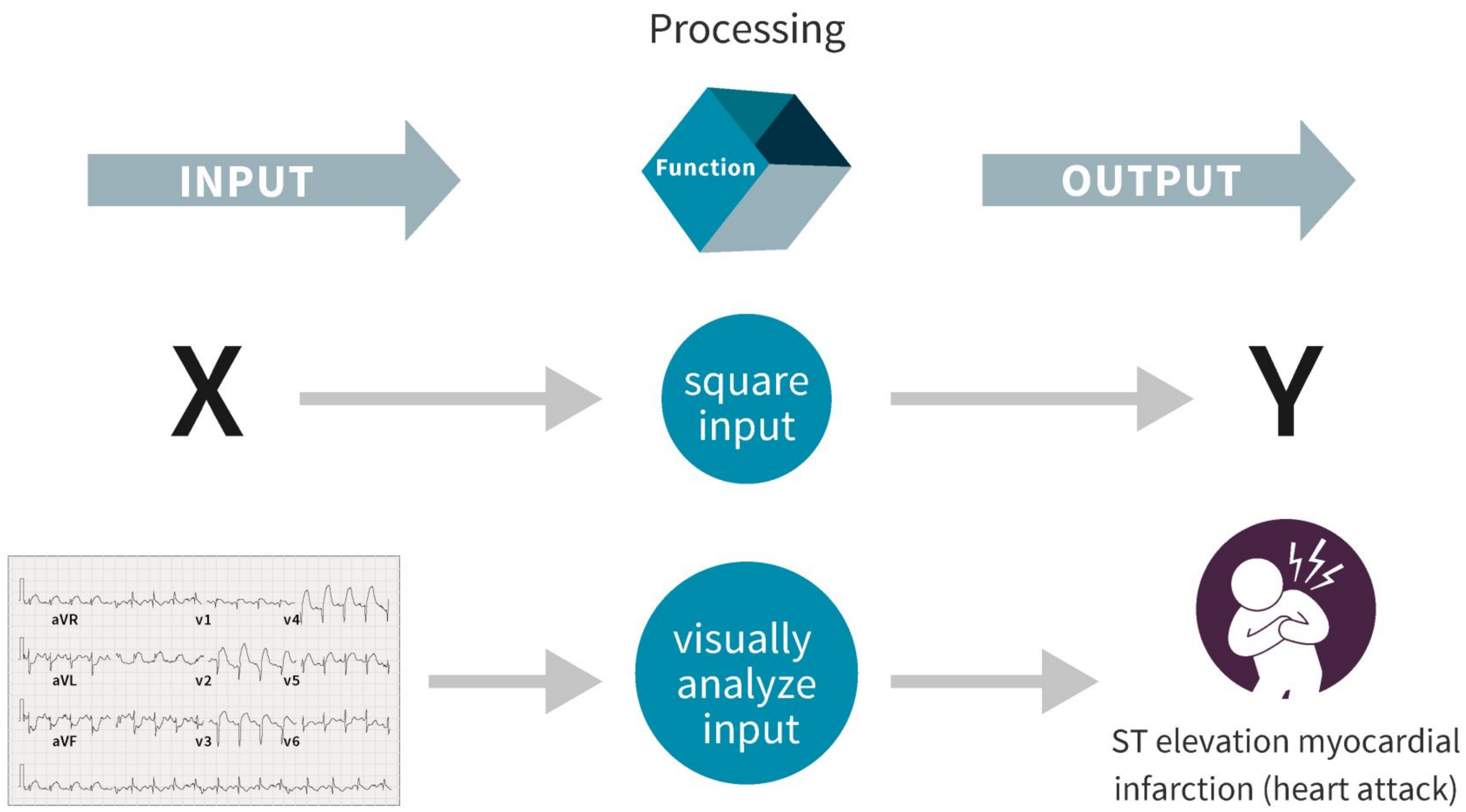
- **Supervised learning**
- Unsupervised learning
- Reinforcement learning

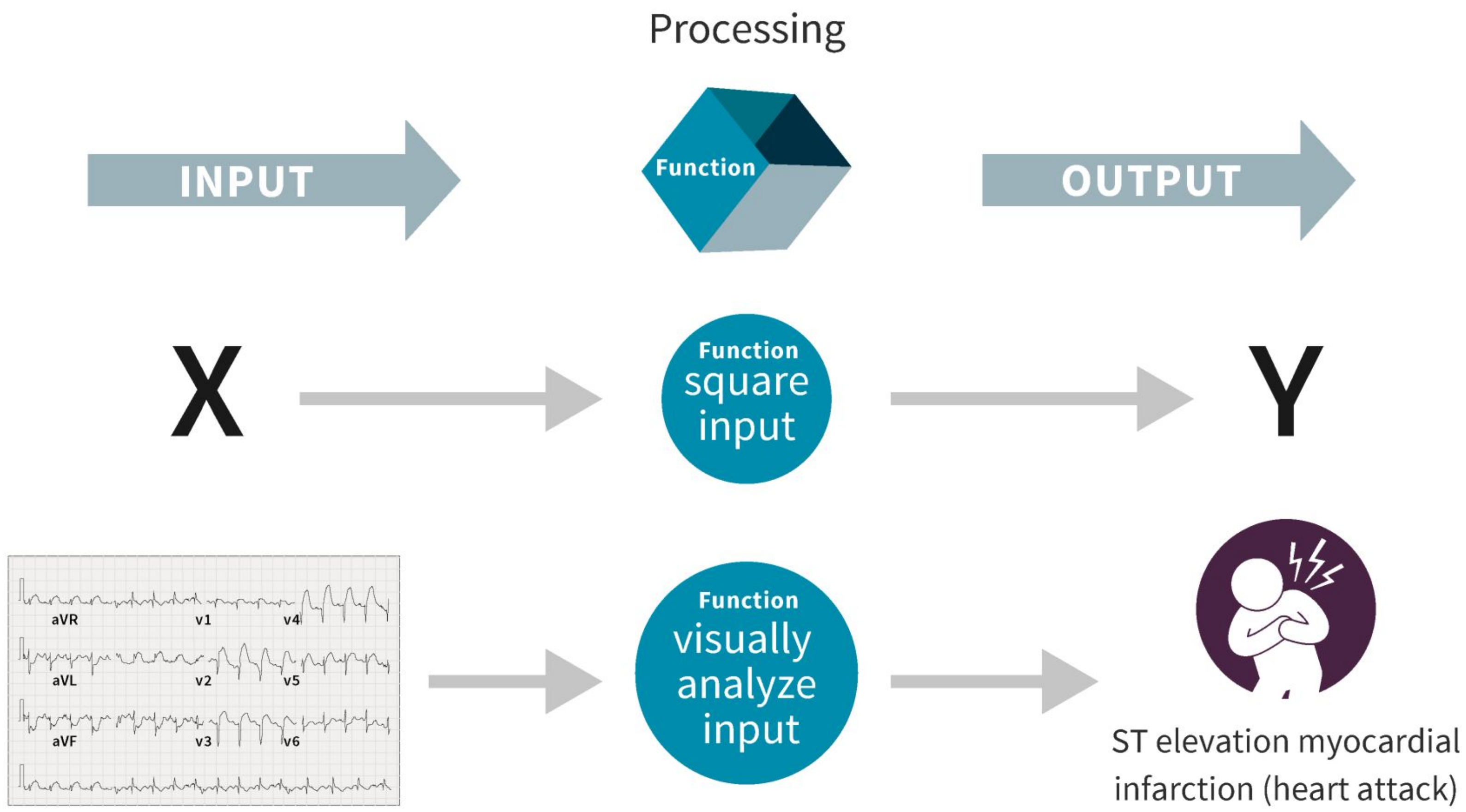
Processing


Processing
INPUT
Function
OUTPUT
X
Y


Processing
INPUT
Function
OUTPUT
X
Y
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
ST elevation myocardial
infarction (heart attack)


Processing
INPUT
Function
OUTPUT
X
square input
Y
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
ST elevation myocardial infarction (heart attack)


Processing
INPUT
Function
OUTPUT
X
square input
Y
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
visually analyze input
ST elevation myocardial infarction (heart attack)



Processing
INPUT
Function
OUTPUT
X
Function square input
Y
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
Function visually analyze input
ST elevation myocardial infarction (heart attack)

Traditional computer programming approach:
Write rules to process the inputs to produce the outputs

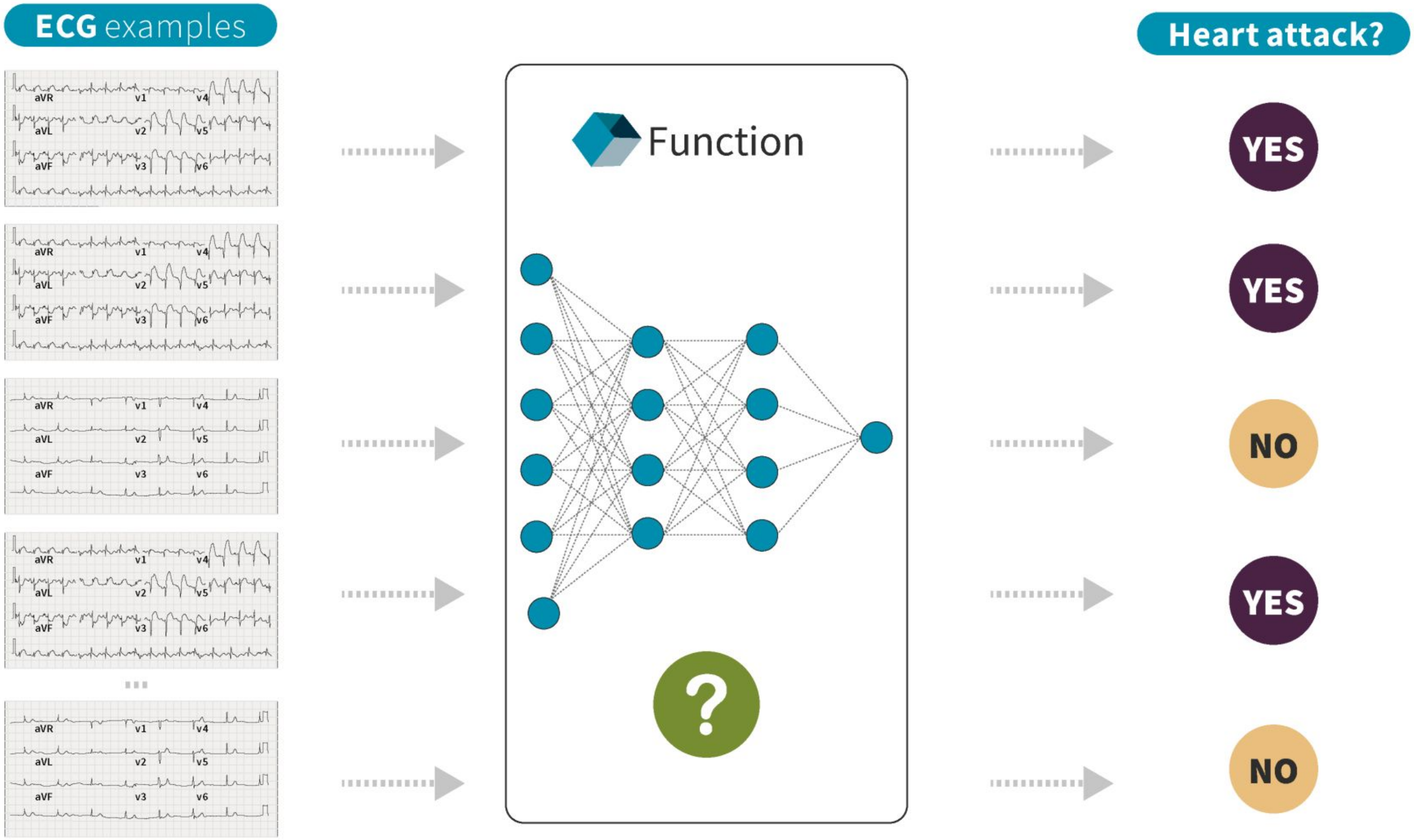
```
def example (x):
y = x^2
return  y
```

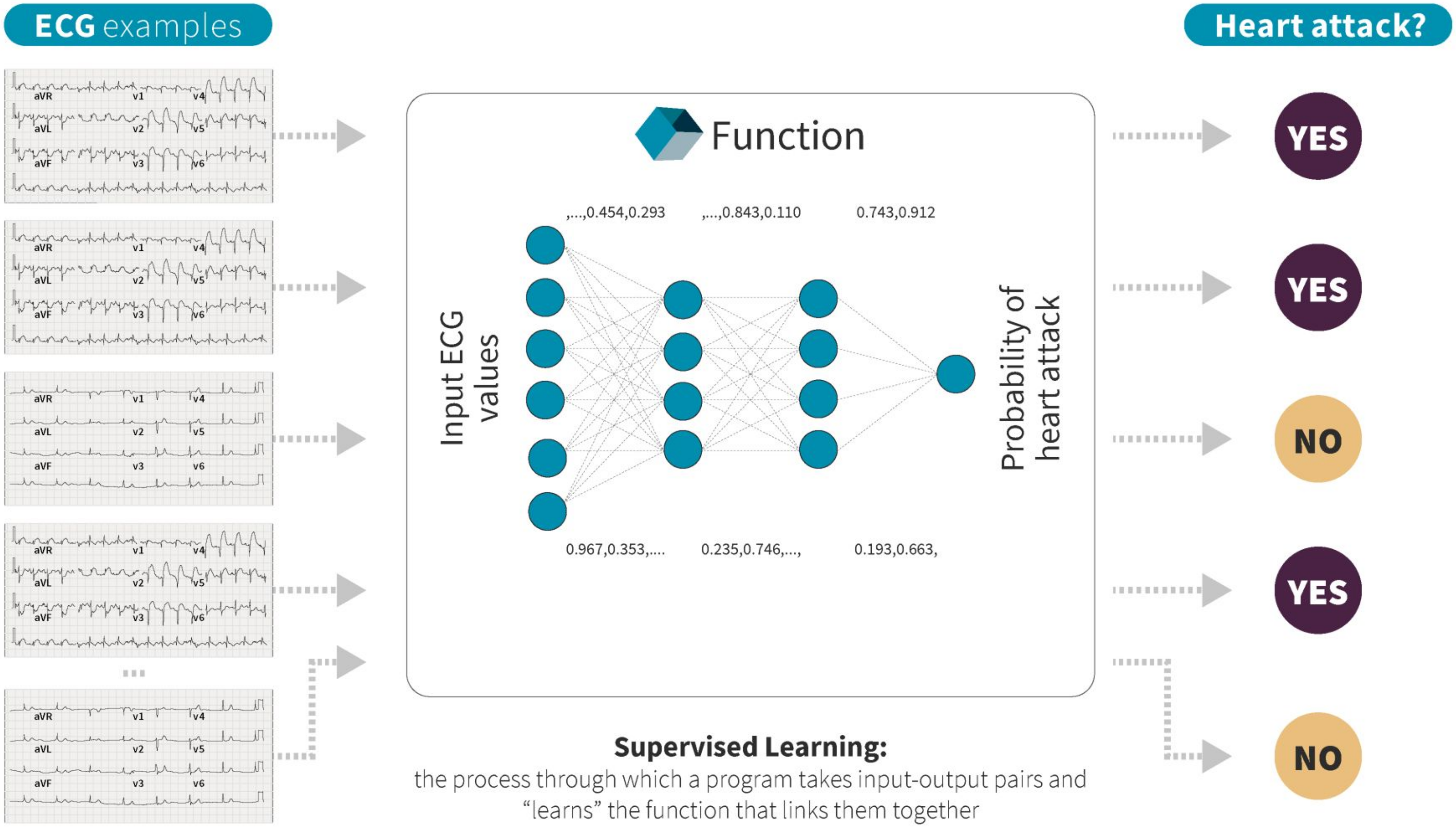
```
def example (x):
y = x^2
return y

def example2 (ECG_data):
####
#### Fill in lots of steps
#### of processing !
return diagnosis
```

(Supervised) machine learning approach:
Collect a dataset of examples linking inputs to outputs, and search for (i.e., ***learn***) a function that can accurately map this data of inputs to the correct outputs

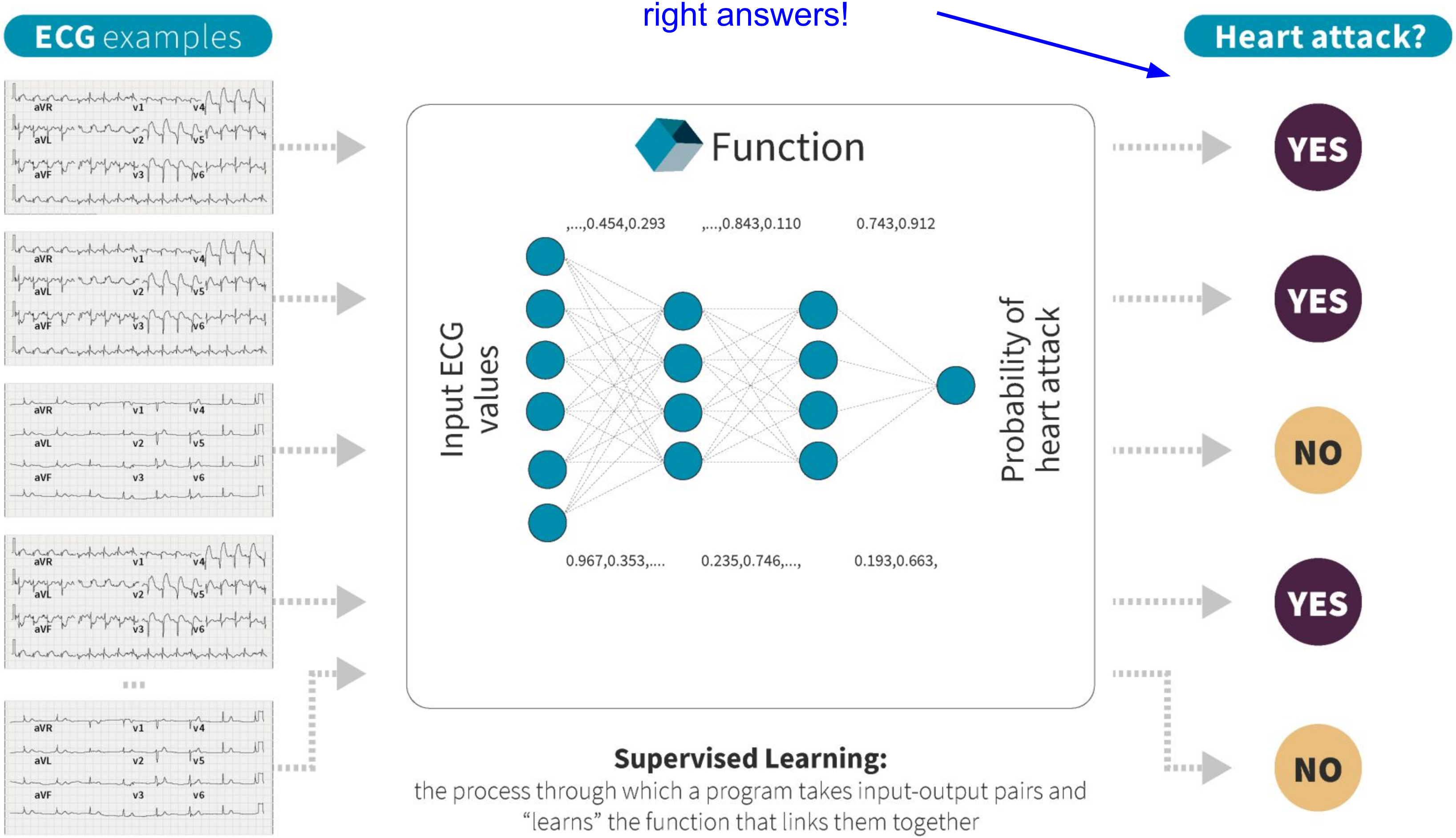

ECG examples
Heart attack?
YES
YES
NO
YES
NO


ECG examples
Function
Heart attack?
YES
YES
NO
YES
NO

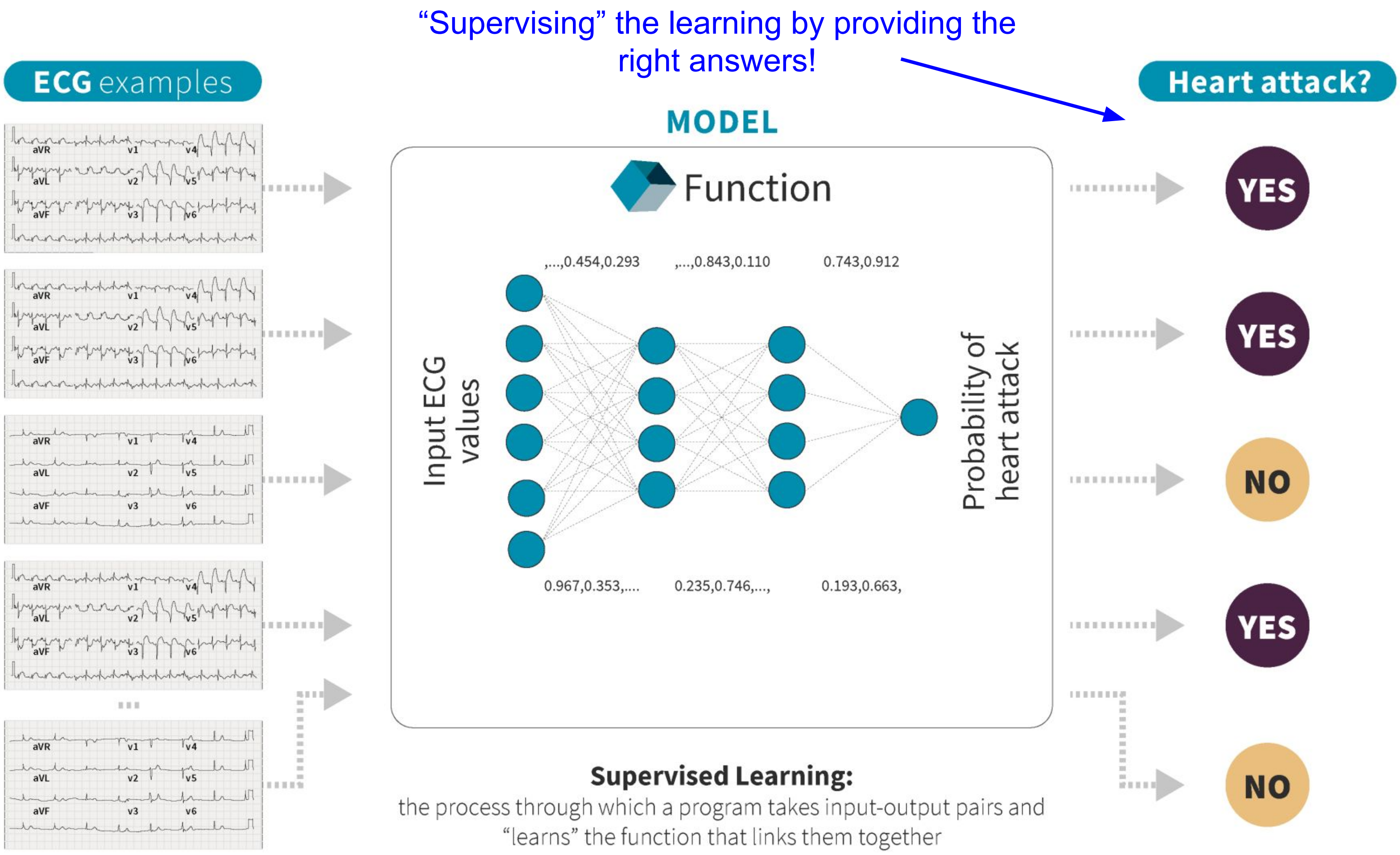


Supervised Learning:
the process through which a program takes input-output pairs and "learns" the function that links them together

“Supervising” the learning by providing the right answers!

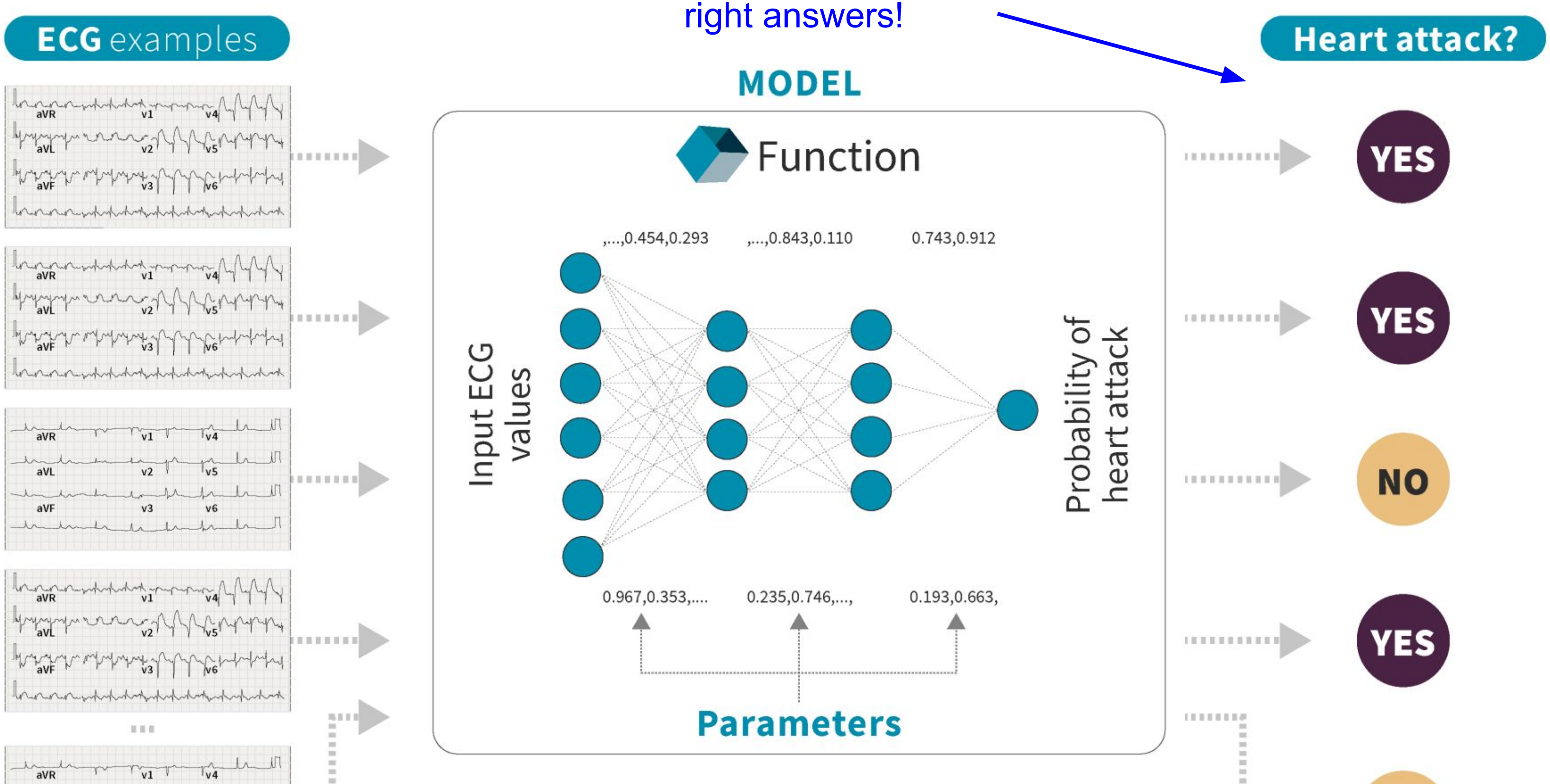


Supervised Learning:
the process through which a program takes input-output pairs and “learns” the function that links them together



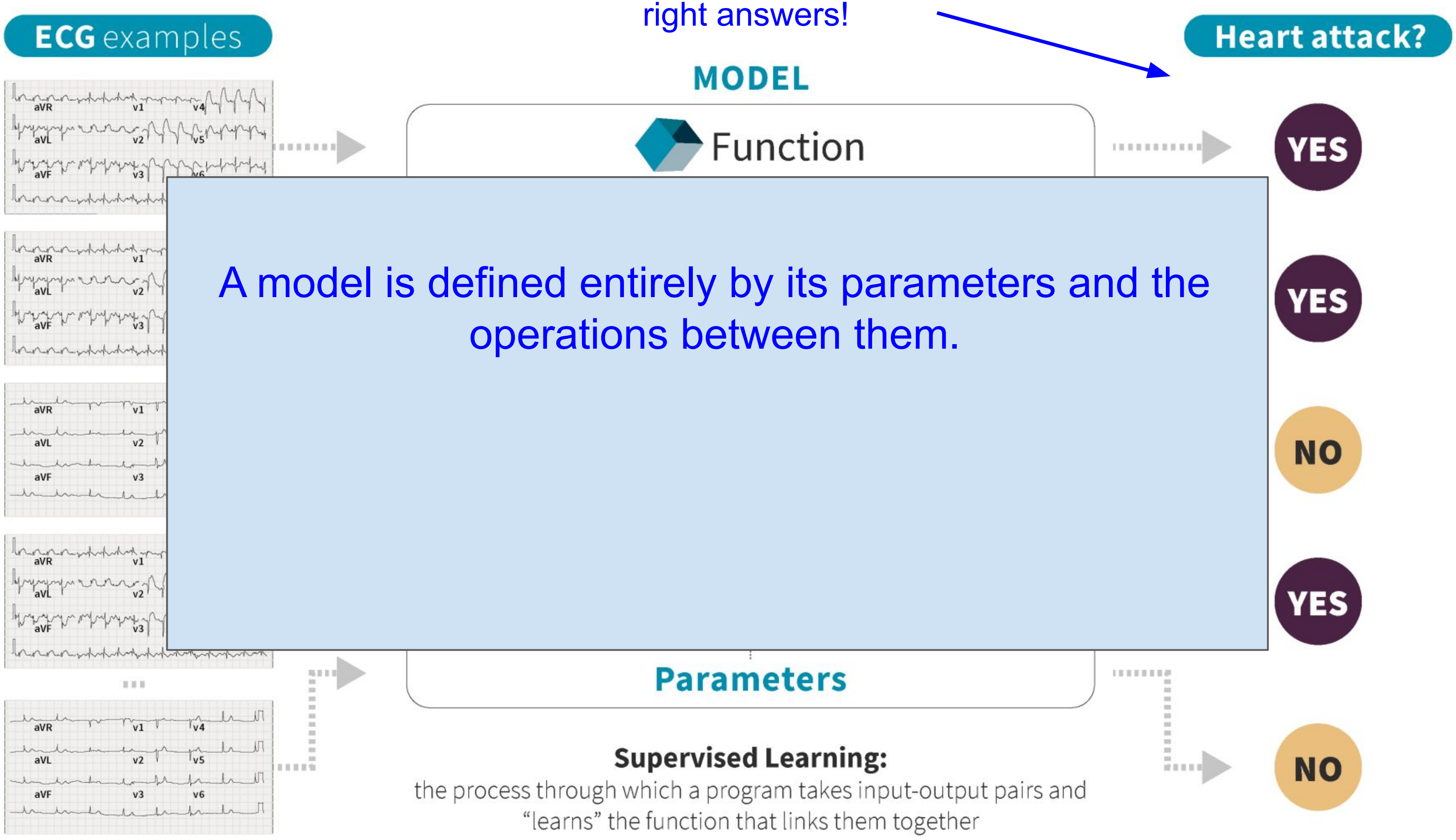
“Supervising” the learning by providing the right answers!
ECG examples
aVR v1 v4
aVL v2 v5
aVF v3 v6
MODEL
Function
Input ECG values
,...,0.454,0.293
,...,0.843,0.110
0.743,0.912
0.967,0.353,....
0.235,0.746,...,
0.193,0.663,
Probability of heart attack
Heart attack?
YES
YES
NO
YES
NO
Supervised Learning:
the process through which a program takes input-output pairs and “learns” the function that links them together

“Supervising” the learning by providing the right answers!



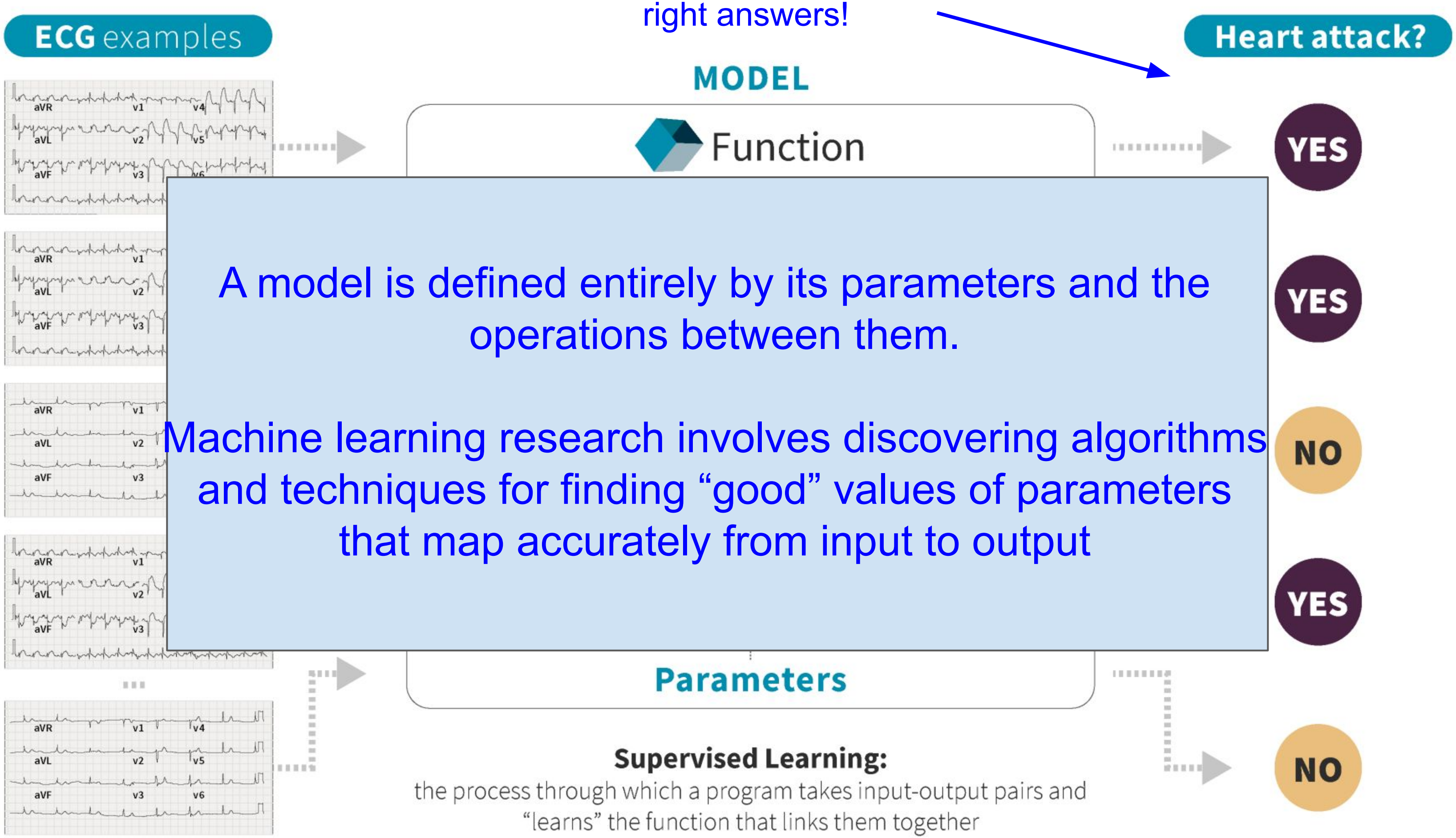
Supervised Learning:
the process through which a program takes input-output pairs and “learns” the function that links them together

“Supervising” the learning by providing the right answers!



A model is defined entirely by its parameters and the operations between them.

"Supervising" the learning by providing the right answers!



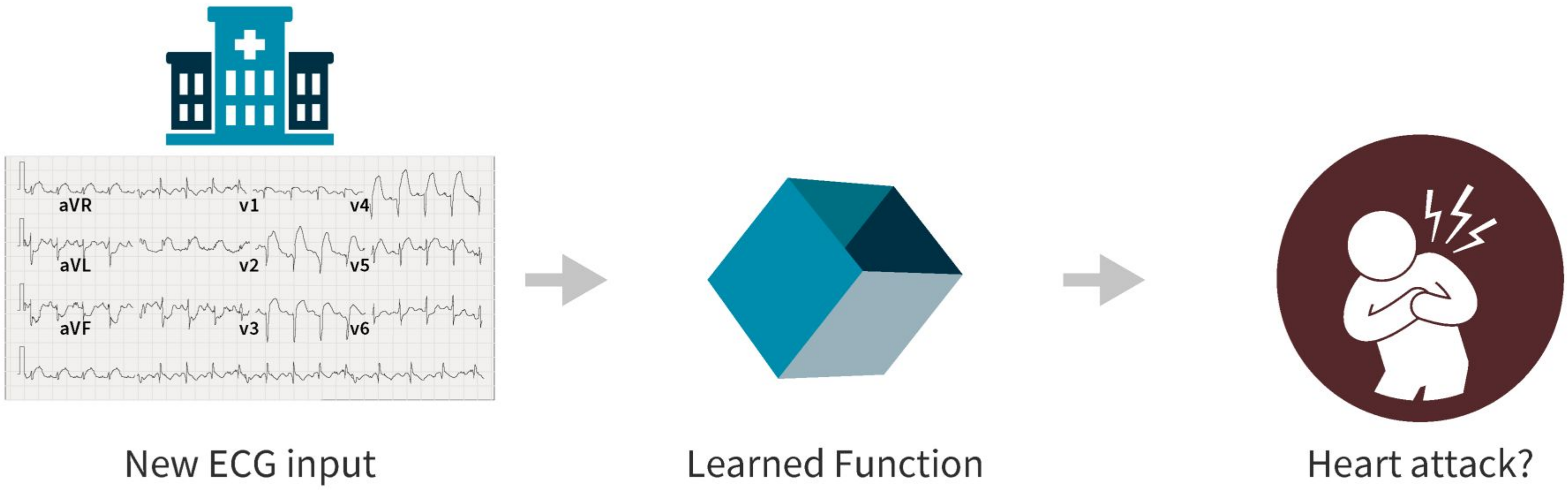
A model is defined entirely by its parameters and the operations between them.

Machine learning research involves discovering algorithms and techniques for finding "good" values of parameters that map accurately from input to output

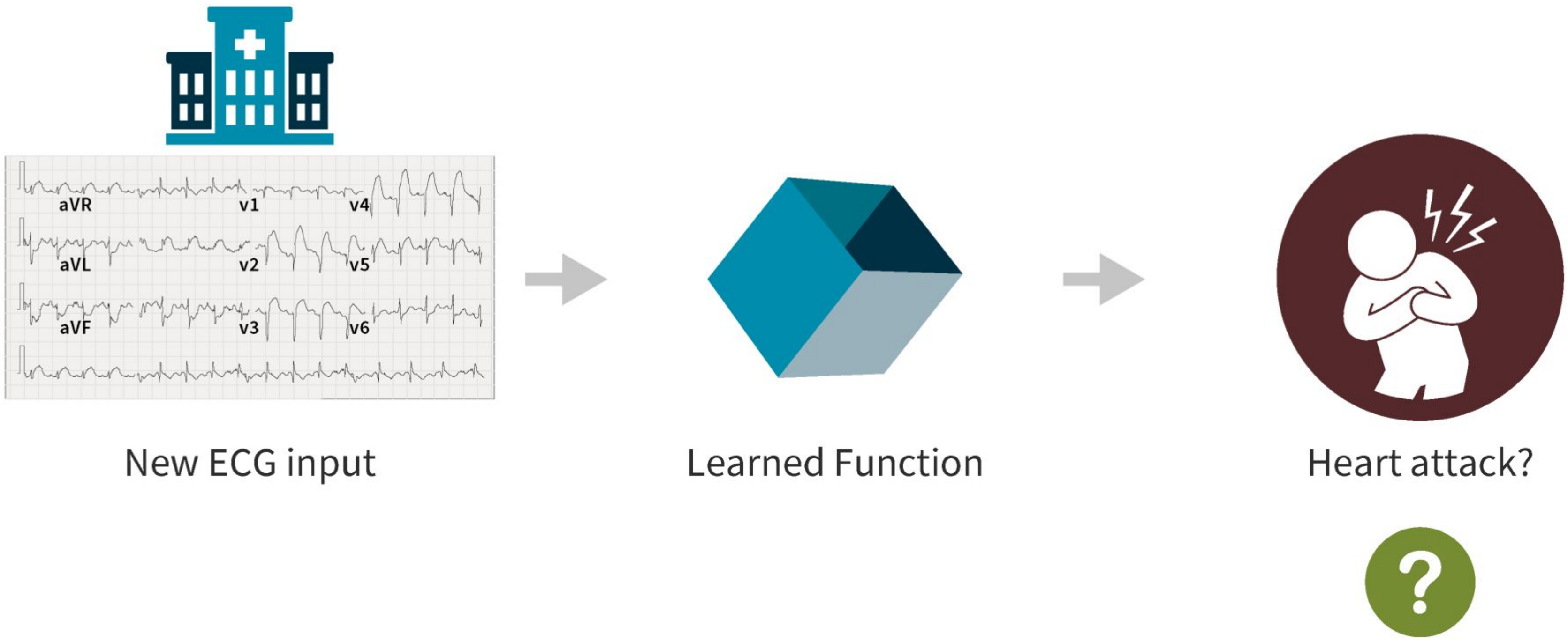
Zooming back out… supervised machine learning

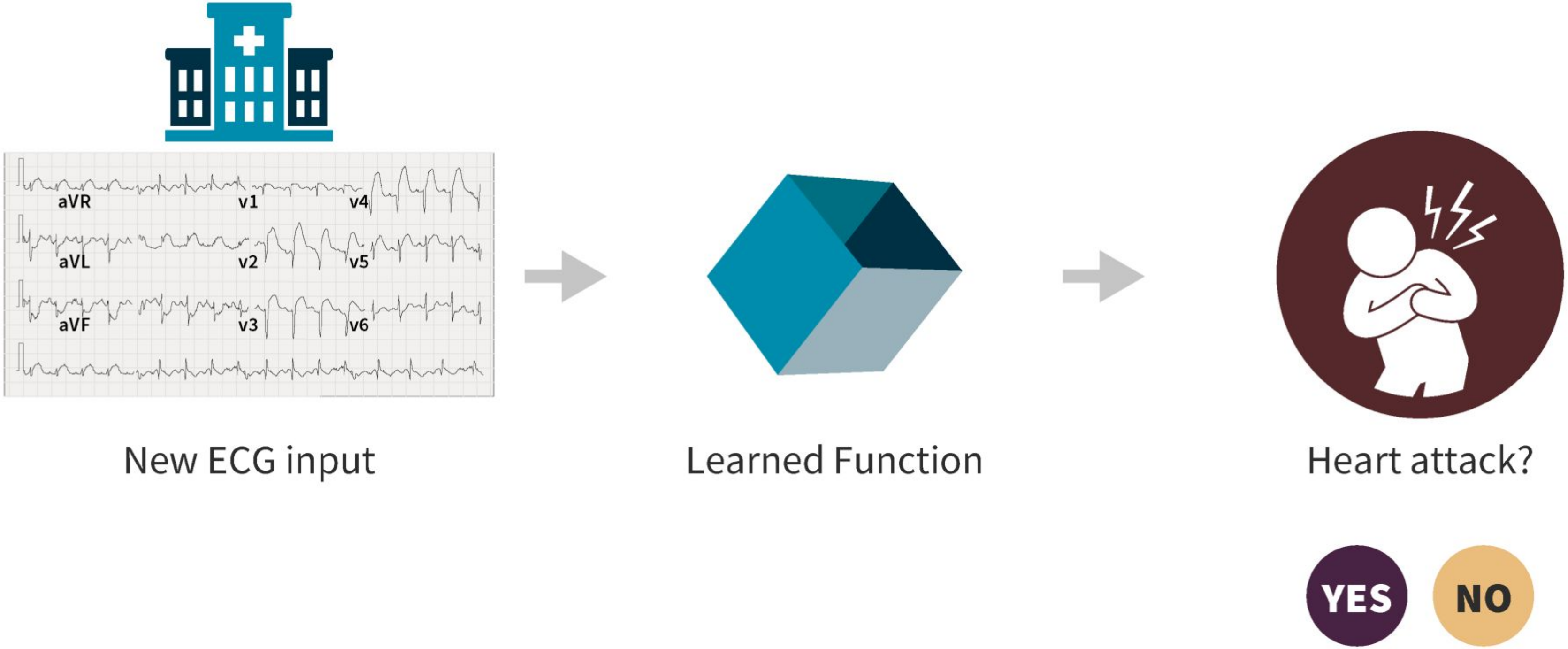
- Once we have performed supervised machine learning and obtained a “trained” model, the model can be used to take new inputs, and produce new outputs (predictions)

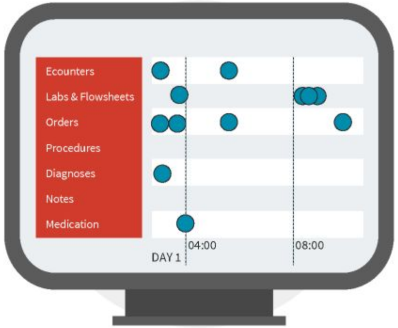
- This replaces the need for the hand-written rules in traditional computer programming!



aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
New ECG input
Learned Function
Heart attack?


aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
New ECG input
Learned Function
Heart attack?
?


aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
New ECG input
Learned Function
Heart attack?
YES
NO



Electronic health record data

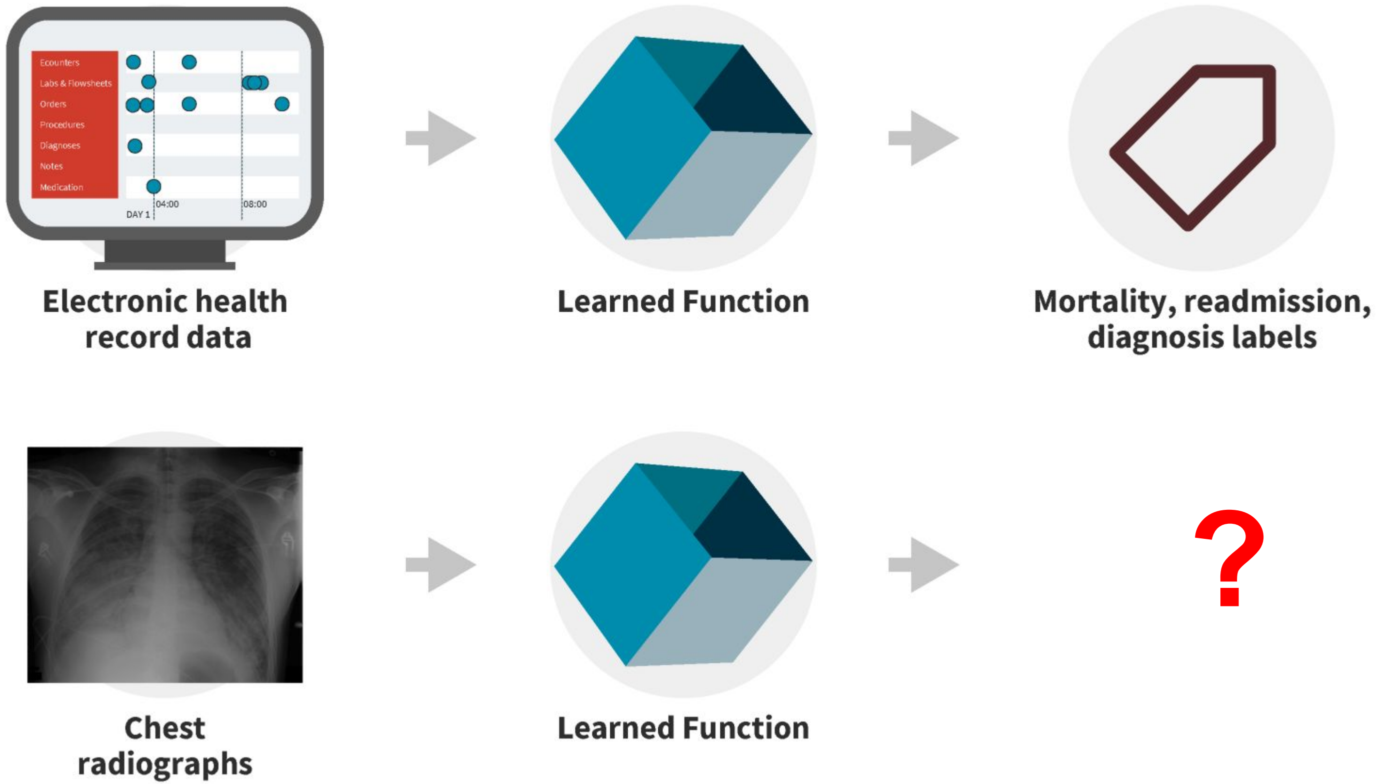
Learned Function

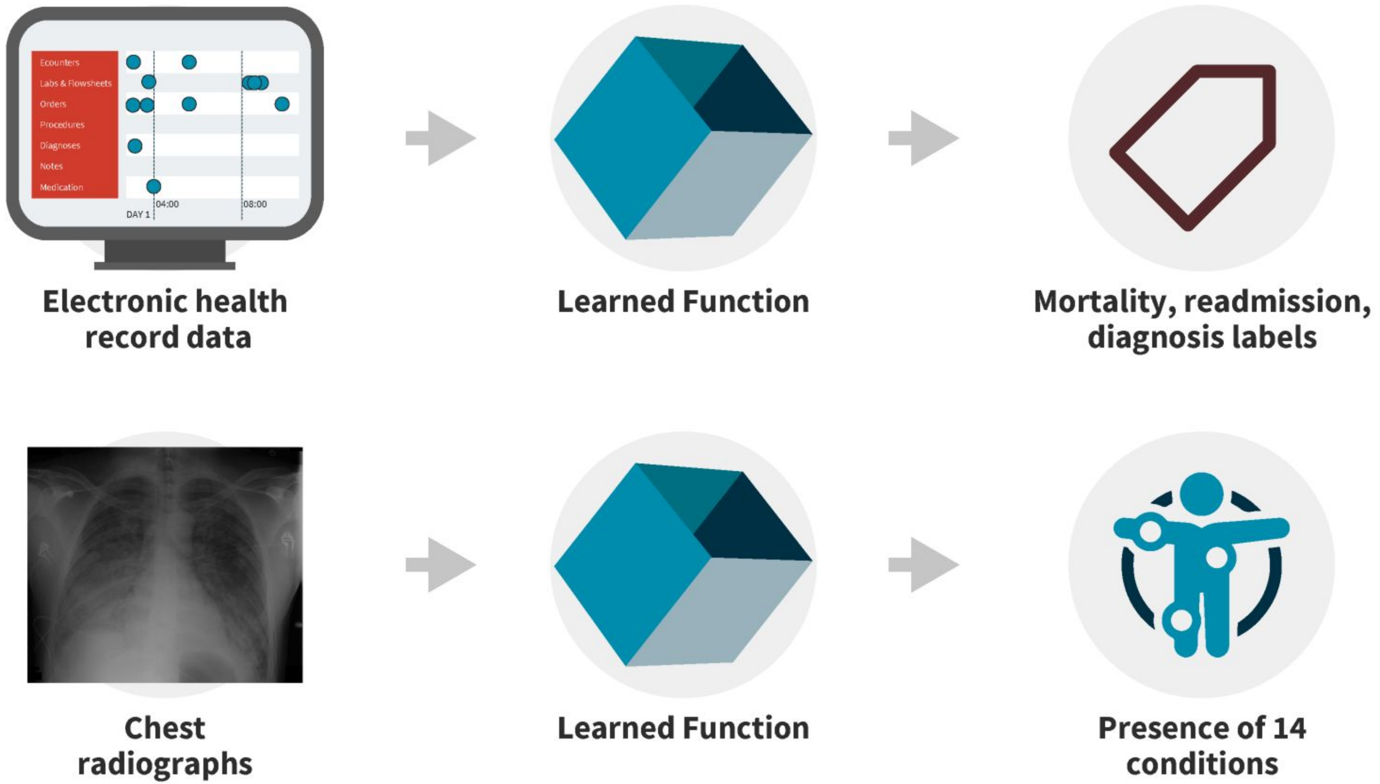
?

Electronic health record data

Learned Function

Mortality, readmission, diagnosis labels


Ecounters
Labs & Flowsheets
Orders
Procedures
Diagnoses
Notes
Medication
04:00
08:00
DAY 1
Electronic health record data
Learned Function
Mortality, readmission, diagnosis labels
Chest radiographs
Learned Function
?


Ecounters
Labs & Flowsheets
Orders
Procedures
Diagnoses
Notes
Medication
DAY 1
04:00
08:00
Electronic health record data
Learned Function
Mortality, readmission, diagnosis labels
Chest radiographs
Learned Function
Presence of 14 conditions

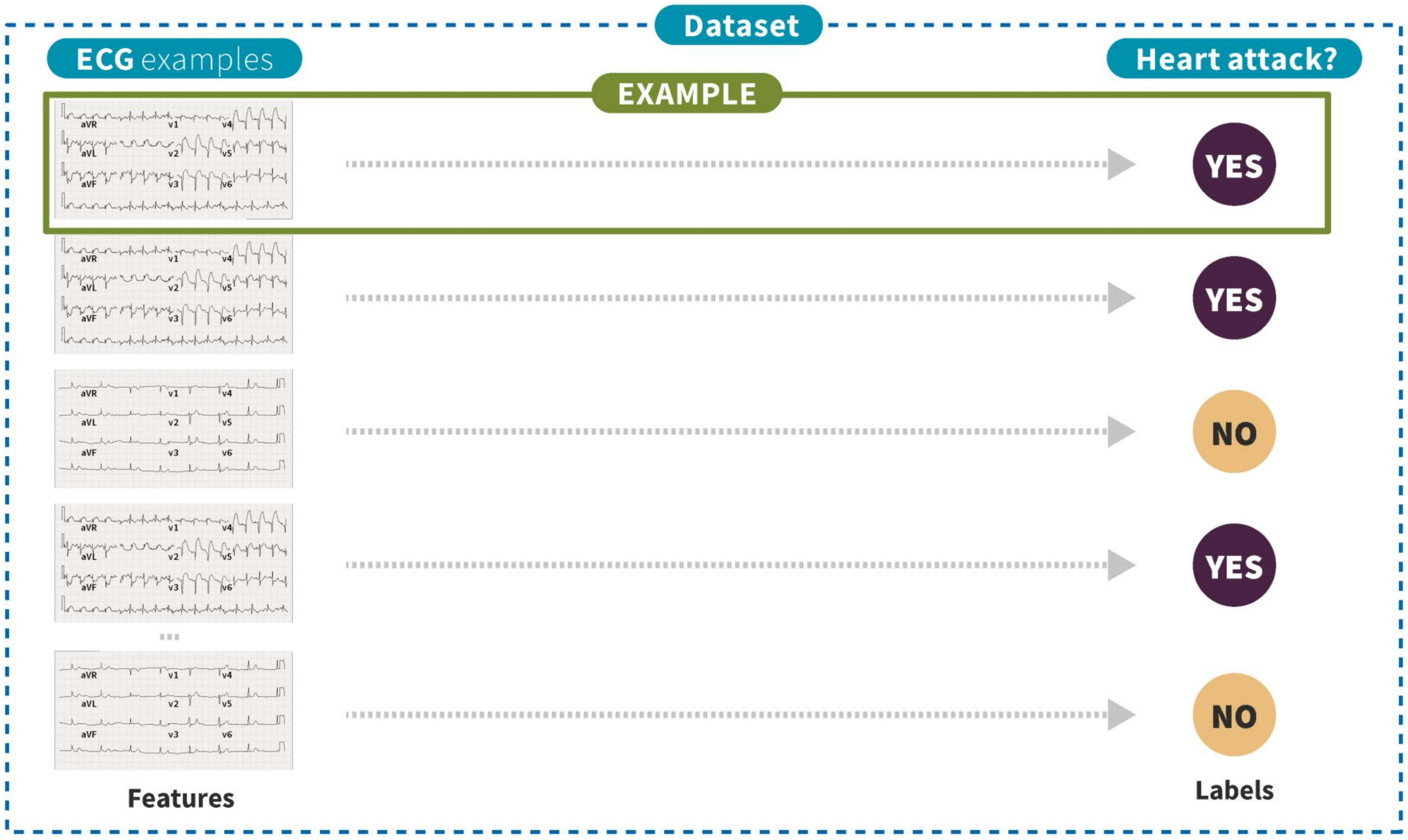
Going deeper into some machine learning terminology...


ECG examples
Heart attack?
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
YES
YES
NO
YES
...
NO

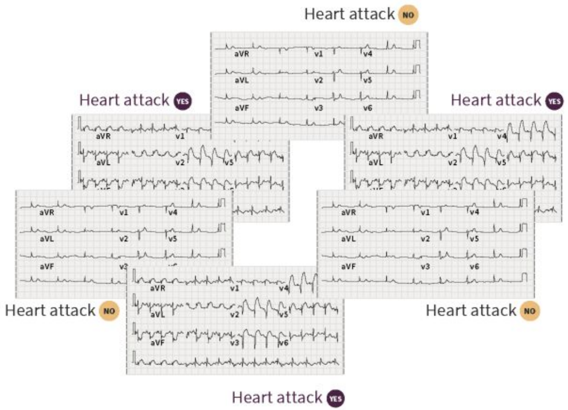

ECG examples
Heart attack?
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
YES
YES
NO
YES
NO
Features


ECG examples
Heart attack?
aVR
v1
v4
aVL
v2
v5
aVF
v3
v6
YES
YES
NO
YES
NO
Features
Labels

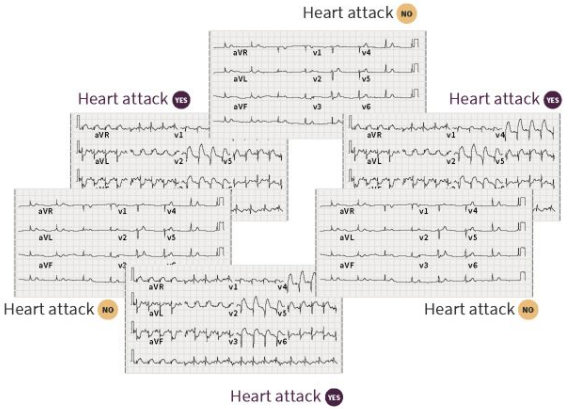

ECG examples
Heart attack?
EXAMPLE
YES
YES
NO
YES
NO
Features
Labels


Dataset
ECG examples
Heart attack?
EXAMPLE
YES
YES
NO
YES
NO
Features
Labels

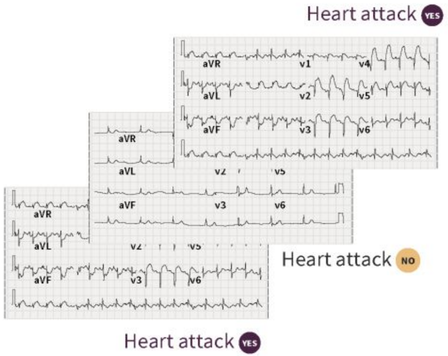
Training set:
examples used to learn function



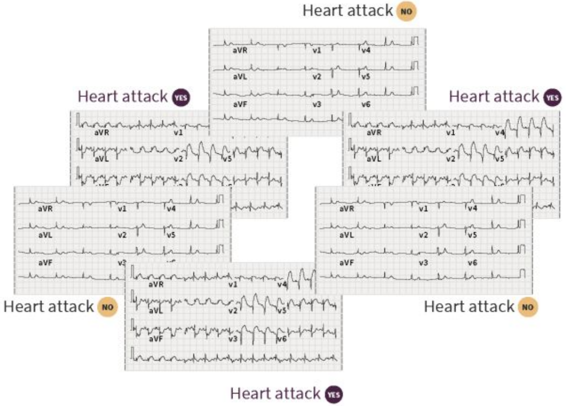
Training set:
examples used to learn function



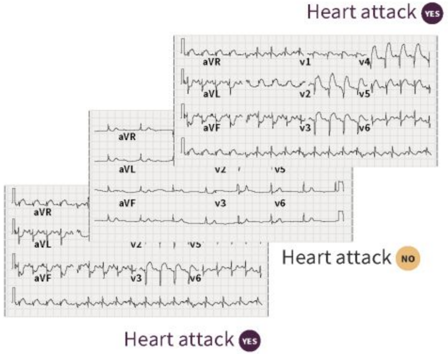
Validation set:
examples used to periodically assess generalization performance and choose hyperparameters



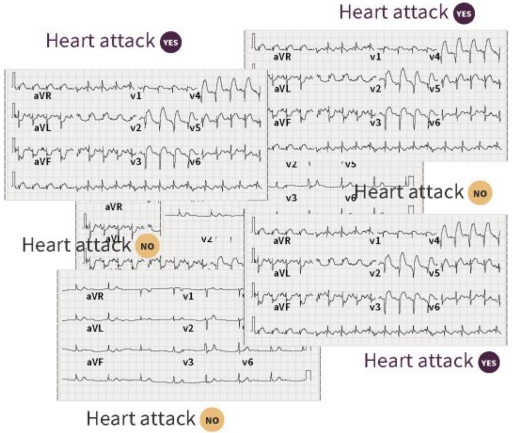
Training set:
examples used to learn function



Validation set:
examples used to periodically assess generalization performance and choose hyperparameters



Test set:
examples evaluated only at the very end of model development (completely unseen during the development process)



Model training

Training loop:

1. Start the program, initialize model with random function
   - Set model best performance = 0
2. Expose the model to training examples, to update function
3. Evaluate the function on the validation set
   - If performance > previous best, update this and save model
4. Repeat steps 2 and 3 until validation performance no longer improves

Model training

Training loop:

1. Start the program, initialize model with random function
    - Set model best performance = 0
2. Expose the model to training examples, to update function
3. Evaluate the function on the validation set
    - If performance > previous best, update this and save model
4. Repeat steps 2 and 3 until validation performance no longer improves

Hyperparameter tuning: repeat training loop for various hyperparameter settings (design choices in program and model)

Model training

Training loop:

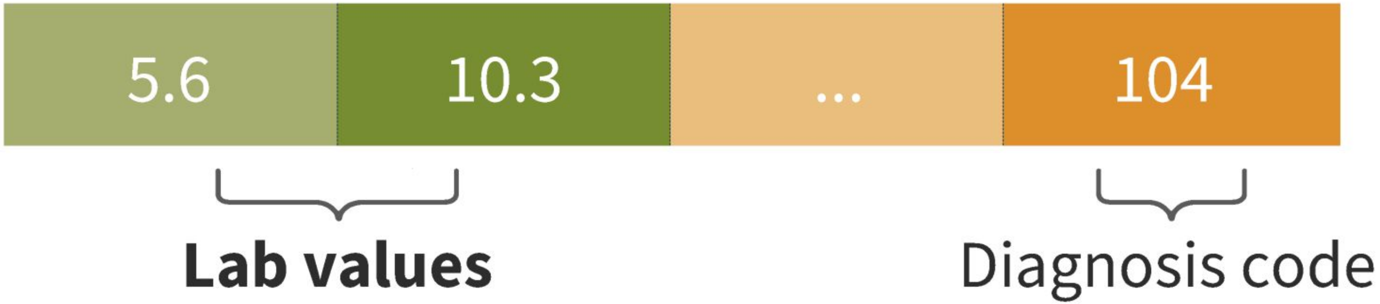
1. Start the program, initialize model with random function
    - Set model best performance = 0
2. Expose the model to training examples, to update function
3. Evaluate the function on the validation set
    - If performance > previous best, update this and save model
4. Repeat steps 2 and 3 until validation performance no longer improves

Hyperparameter tuning: repeat training loop for various hyperparameter settings (design choices in program and model)

Assess final model performance on test set

More on input features: numeric data representing the input, that are given to the model in order to make a prediction

EXAMPLES OF INPUT FEATURES



EXAMPLES OF INPUT FEATURES

Heart attack?

Unstructured data

Physician Note

"...PMH of **metastatic breast cancer, R lung malignant** effusion, and **R lung empyema** who presents with increased drainage from **R lung pleurx** tract..."

EXAMPLES OF INPUT FEATURES

Heart attack?

5.6	10.3	...	104

$\underbrace{5.6 \quad 10.3}_{\text{Lab values}}$ ... $\underbrace{104}_{\text{Diagnosis code}}$

```
[[  2   1   1  37   1  10  66  60  77  94  78  69  64  23  12  45  28  45]
 [ 58   1   9  13  17  29  56  72  65  64  59  58  39  18  15  12   7   1]
 [ 71  49  53  38  30  41  73  73  80  71  69  69  72  45  45  49  36  59]
 [ 88  60  73  50  59  59  54  51  71  81  69  50  54  75  56  61  80  67]
 [ 94  91  86  59  65  57  57  52  64  88  66  56  55  54  70  64 109 114]
 [ 94  95  84  74  70  41  48  55  74  85  84  60  50  46  70  82  92 122]
 [ 85  85  95  83  54  37  59  60  84  97  82  50  38  44  56  92 111 112]
 [ 81  87  94  92  54  54  56  54  79  96  79  48  36  44  62 103 107 145]
 [ 67  83  91  87  60  59  61  71  91 108  86  65  53  40  63 101 110 121]
 [ 49  73  88  72  66  73  78  84 107 120 102  71  57  39  56  89 114 103]
 [ 31  61  84  65  73  80  92 103 117 128 114  76  66  57  52  89 111  91]
 [  6  51  82  84  92  90  92 114 128 135 122 109  73  69  69  84 109  66]
 [  2  44  72  87  95 104 113 124 138 141 130 122  96  77  68  76 104  10]
 [  0  37  74  84 102 113 115 131 146 146 133 124 113  94  83  96  90   1]
 [  0  33  67  90 113 126 130 140 148 147 136 130 117  95  91  81  71   1]
 [  0  33  68  98 122 139 141 144 153 149 135 127 122 108  96  76  65   1]
 [  0  36  81 105 127 144 151 151 155 149 125 114 113 121 105  76  49   1]
 [  0  39  90 114 131 151 155 157 161 153 122  96 102 107 110  66  50   0]]
```

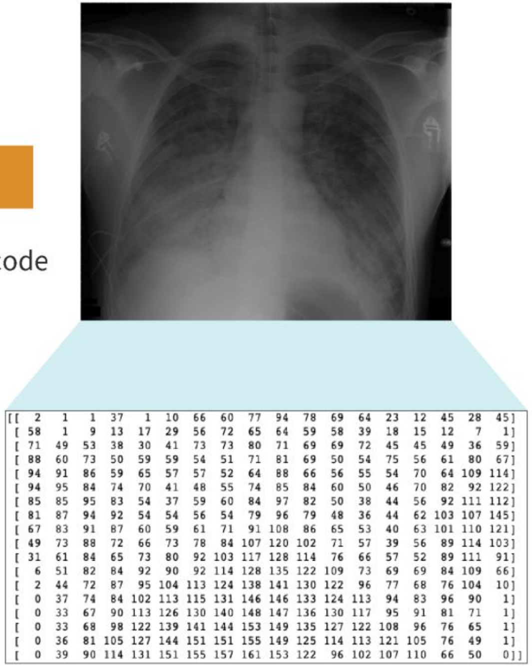
Unstructured data

Physician Note

"...PMH of **metastatic breast cancer, R lung malignant** effusion, and **R lung empyema** who presents with increased drainage from **R lung pleurx** tract..."

EXAMPLES OF INPUT FEATURES

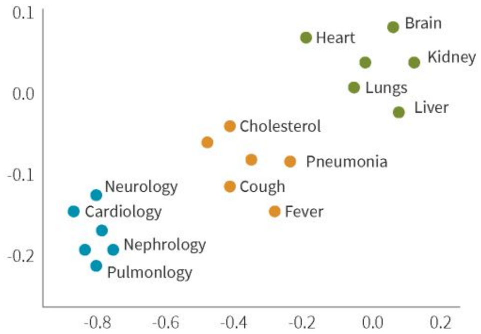
Heart attack?

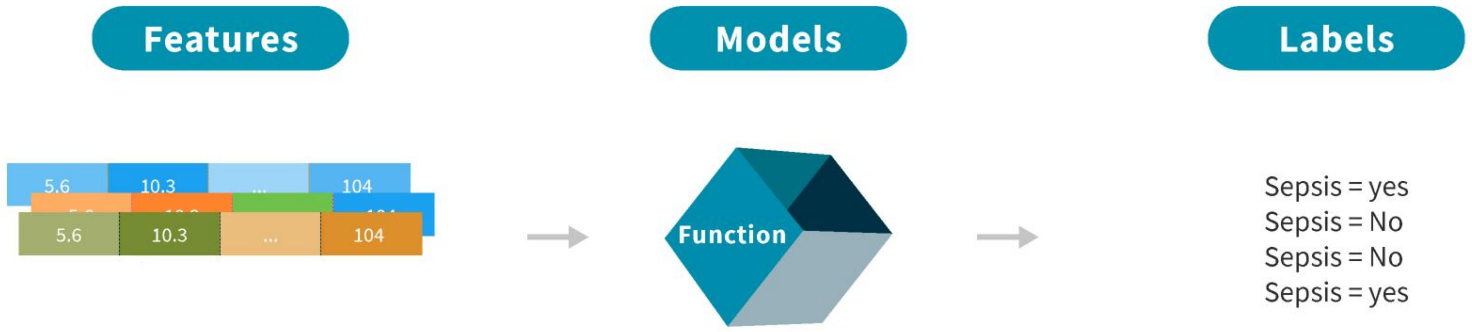


Unstructured data

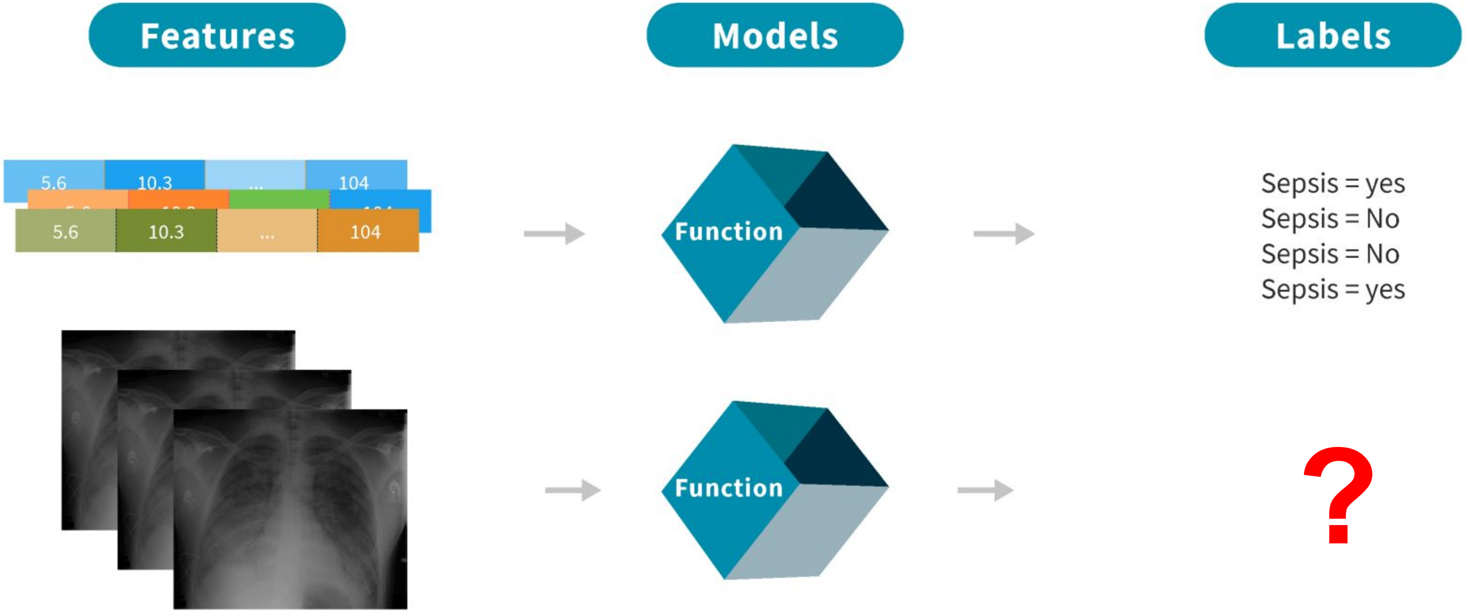
Physician Note

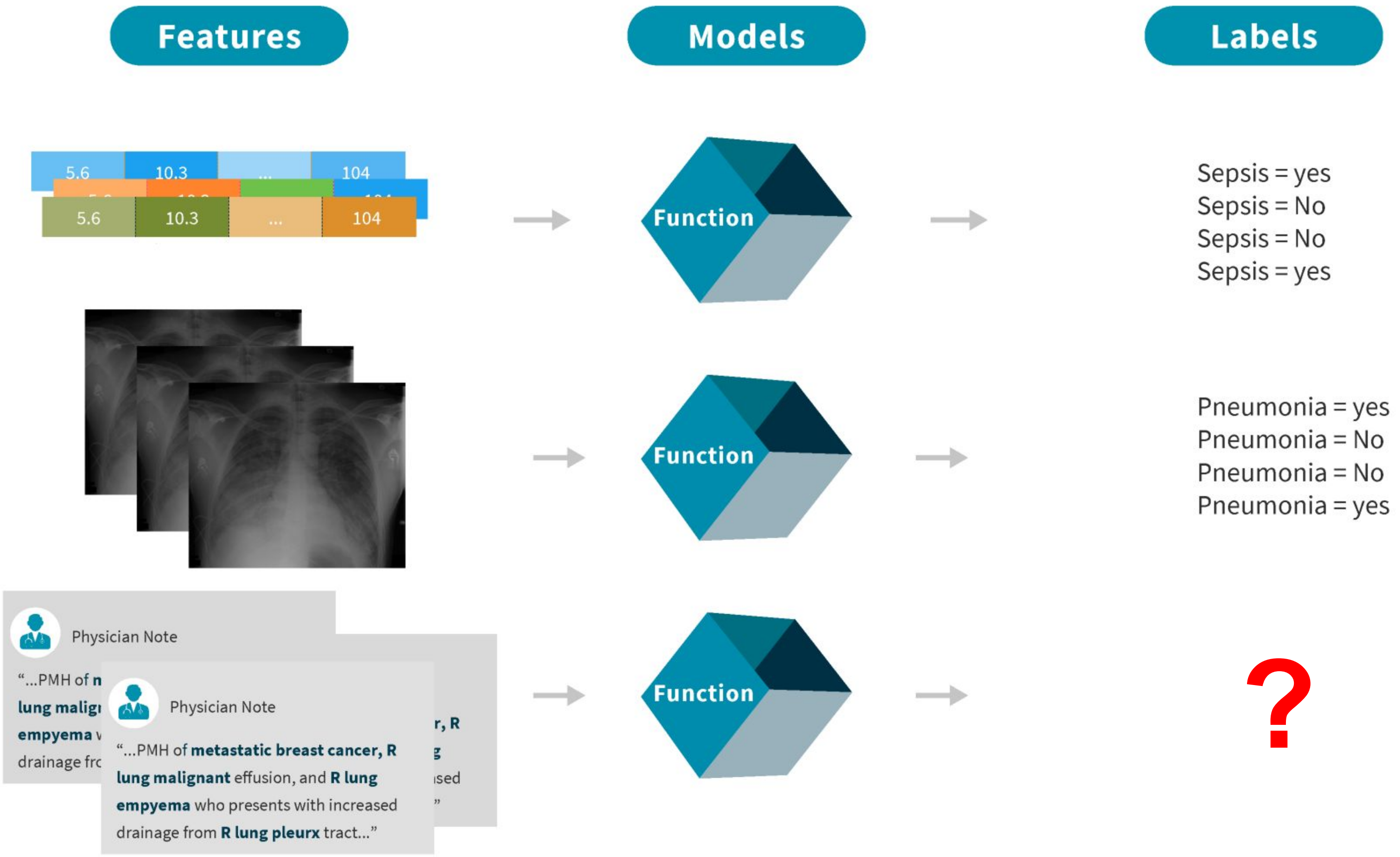
"...PMH of **metastatic breast cancer, R lung malignant** effusion, and **R lung empyema** who presents with increased drainage from **R lung pleurx** tract..."

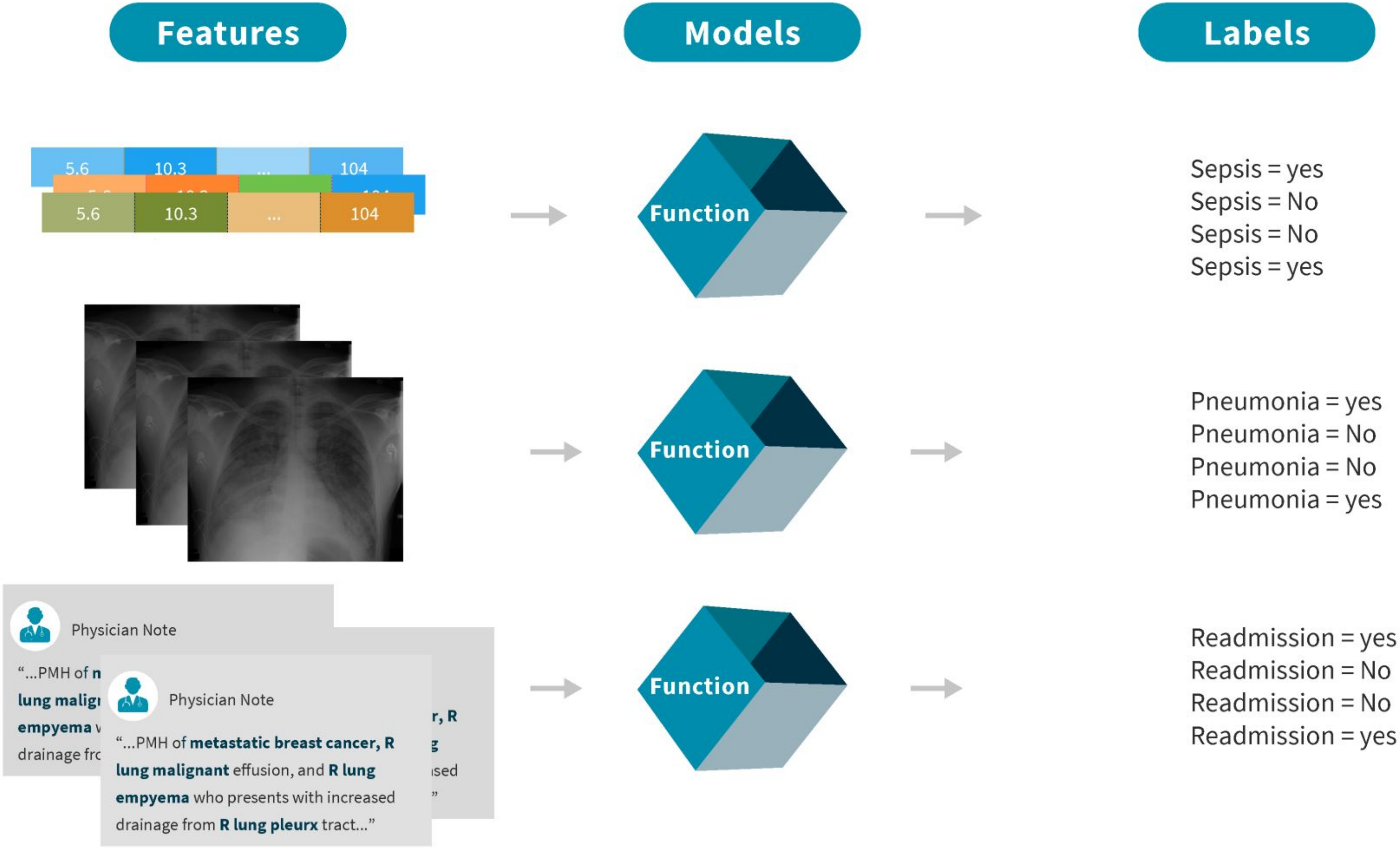




Features
Models
Labels
5.6
10.3
...
104
5.6
10.3
...
104
Function
Sepsis = yes
Sepsis = No
Sepsis = No
Sepsis = yes


Features
Models
Labels
5.6
10.3
...
104
5.6
10.3
...
104
Function
Function
Sepsis = yes
Sepsis = No
Sepsis = No
Sepsis = yes
?


Features
Models
Labels
5.6
10.3
...
104
Function
Function
Function
Sepsis = yes
Sepsis = No
Sepsis = No
Sepsis = yes
Pneumonia = yes
Pneumonia = No
Pneumonia = No
Pneumonia = yes
?
Physician Note
Physician Note
“...PMH of metastatic breast cancer, R lung malignant effusion, and R lung empyema who presents with increased drainage from R lung pleurx tract...”



Features
Models
Labels
5.6
10.3
...
104
Function
Sepsis = yes
Sepsis = No
Sepsis = No
Sepsis = yes
Pneumonia = yes
Pneumonia = No
Pneumonia = No
Pneumonia = yes
Physician Note
"...PMH of metastatic breast cancer, R lung malignant effusion, and R lung empyema who presents with increased drainage from R lung pleurx tract..."
Readmission = yes
Readmission = No
Readmission = No
Readmission = yes

Examples of machine learning tasks corresponding to different types of desired outputs (labels)

REGRESSION

Real Numbers

CLASSIFICATION

Categories

Let’s look at some first examples of traditional machine learning algorithms for these tasks

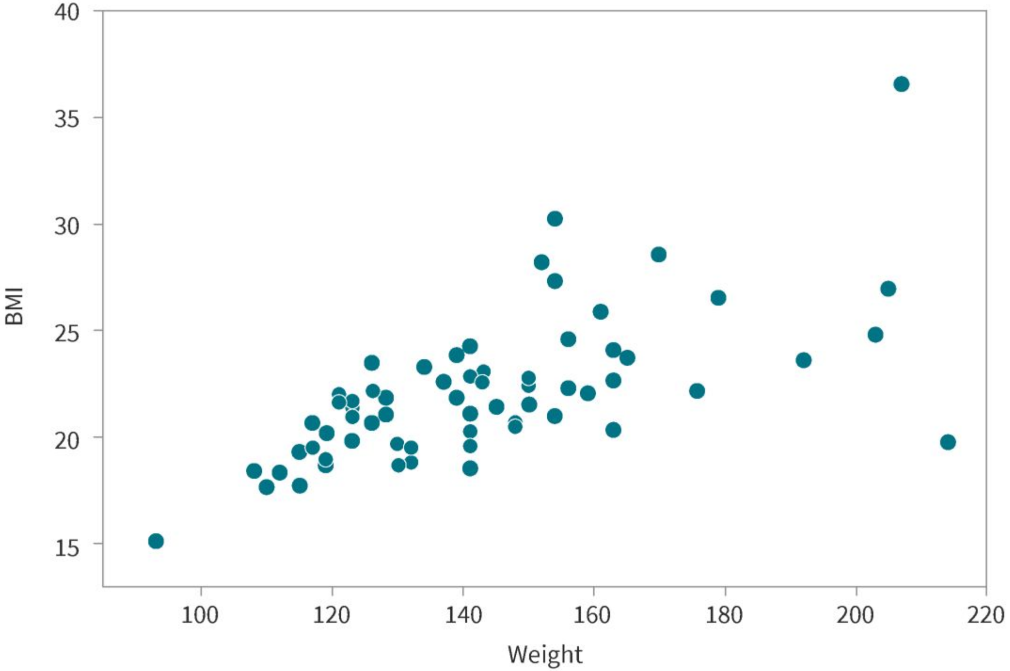
EXAMPLE: *USING BODY WEIGHT TO PREDICT BODY MASS INDEX (BMI)*

	WEIGHT	BMI
0	159	22.02
1	214	19.70
2	163	24.09
3	205	26.97
4	150	21.51
...	...	...
62	143	22.51
63	165	23.69
64	93	15.08
65	163	22.64
66	207	36.57

Source: https://people.sc.fsu.edu/~jburkardt/data/csv/

EXAMPLE: *USING BODY WEIGHT TO PREDICT BODY MASS INDEX (BMI)*

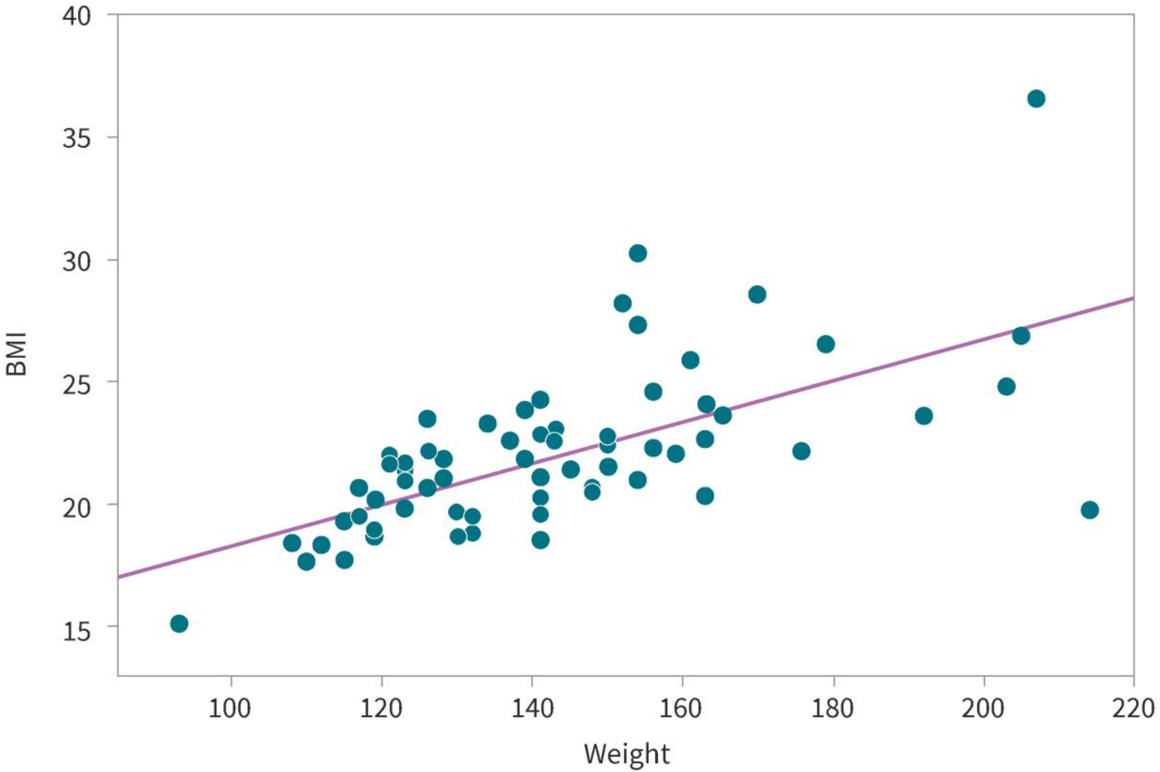
	WEIGHT	BMI
0	159	22.02
1	214	19.70
2	163	24.09
3	205	26.97
4	150	21.51
...	...	...
62	143	22.51
63	165	23.69
64	93	15.08
65	163	22.64
66	207	36.57



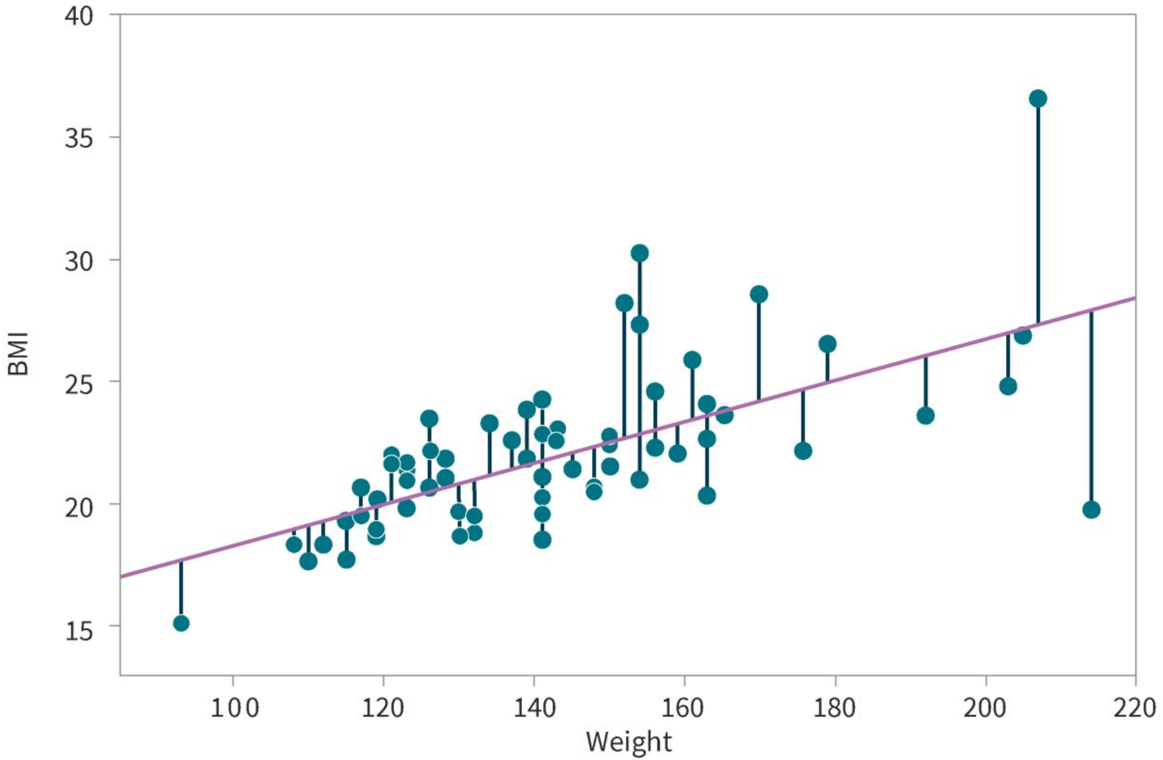
Source: https://people.sc.fsu.edu/~jburkardt/data/csv/

EXAMPLE: *USING BODY WEIGHT TO PREDICT BODY MASS INDEX (BMI)*

Example: linear regression algorithm to learn a (linear) function



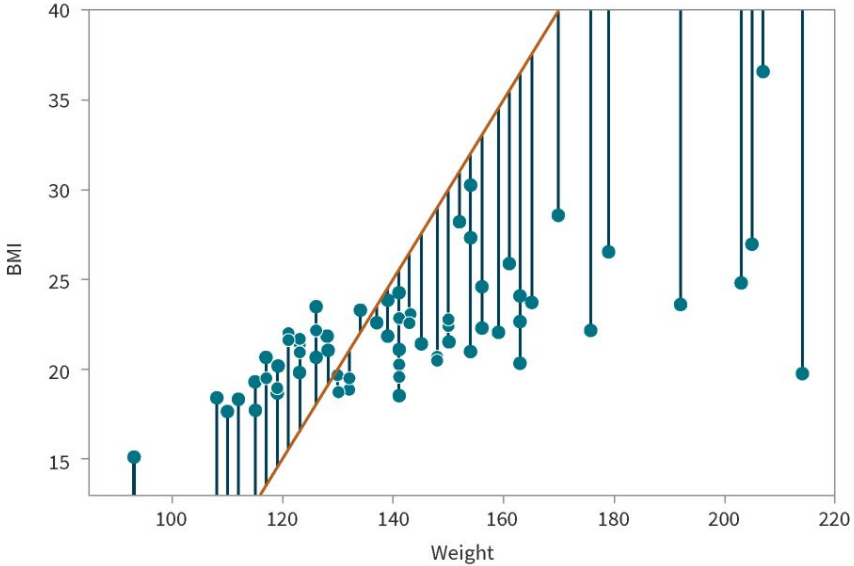
EXAMPLE: *USING BODY WEIGHT TO PREDICT BODY MASS INDEX (BMI)*



Loss: the difference between the function and the actual y values.

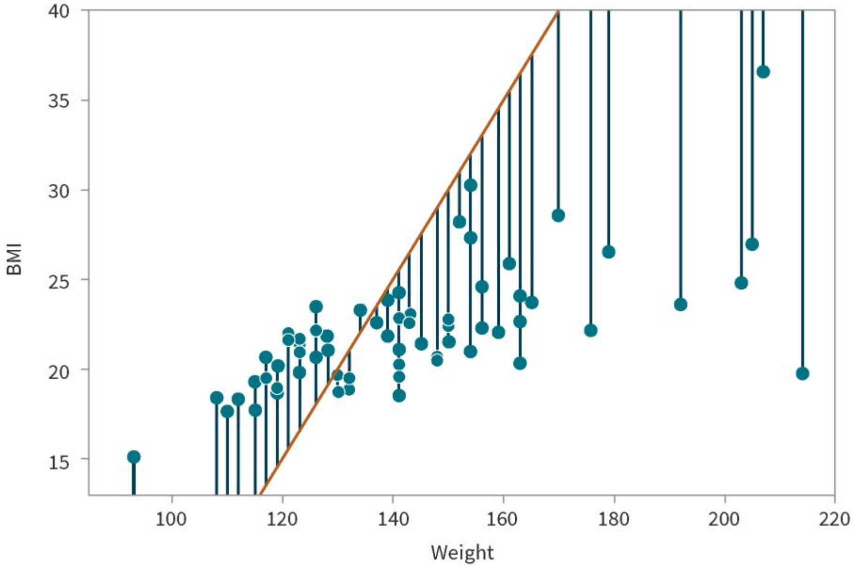
Training a machine learning model means to minimize the loss!

EXAMPLE: *USING BODY WEIGHT TO PREDICT BODY MASS INDEX (BMI)*

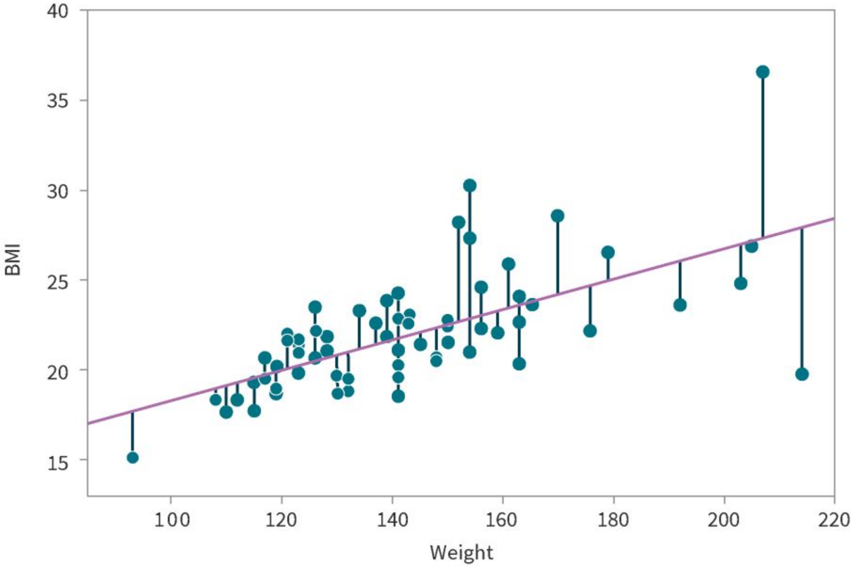


A poorly trained or (poorly fitted) model has high cumulative loss

EXAMPLE: *USING BODY WEIGHT TO PREDICT BODY MASS INDEX (BMI)*



A poorly trained or (poorly fitted) model has high cumulative loss



A well trained or (well fitted) model minimizes the cumulative loss.

$$y = mx + b$$

$$y = \overbrace{\overset{\text{WEIGHT}}{m}x + \overset{\text{BIAS}}{b}}^{\text{PARAMETERS}}$$

CLASSIFICATION: *PREDICTING CATEGORICAL LABELS*

CLASSIFICATION: *PREDICTING CATEGORICAL LABELS*

Examples:

INPUT FEATURES	LABELS
Image pixels	What is present in the image: lung nodule, skin cancer, knee arthritis, etc.

CLASSIFICATION: *PREDICTING CATEGORICAL LABELS*

Examples:

INPUT FEATURES	LABELS
Image pixels	What is present in the image: lung nodule, skin cancer, knee arthritis, etc.
Structured data (e.g. lab values, diagnosis codes, age)	Heart attack, sepsis, mortality, etc.

CLASSIFICATION: *PREDICTING CATEGORICAL LABELS*

Examples:

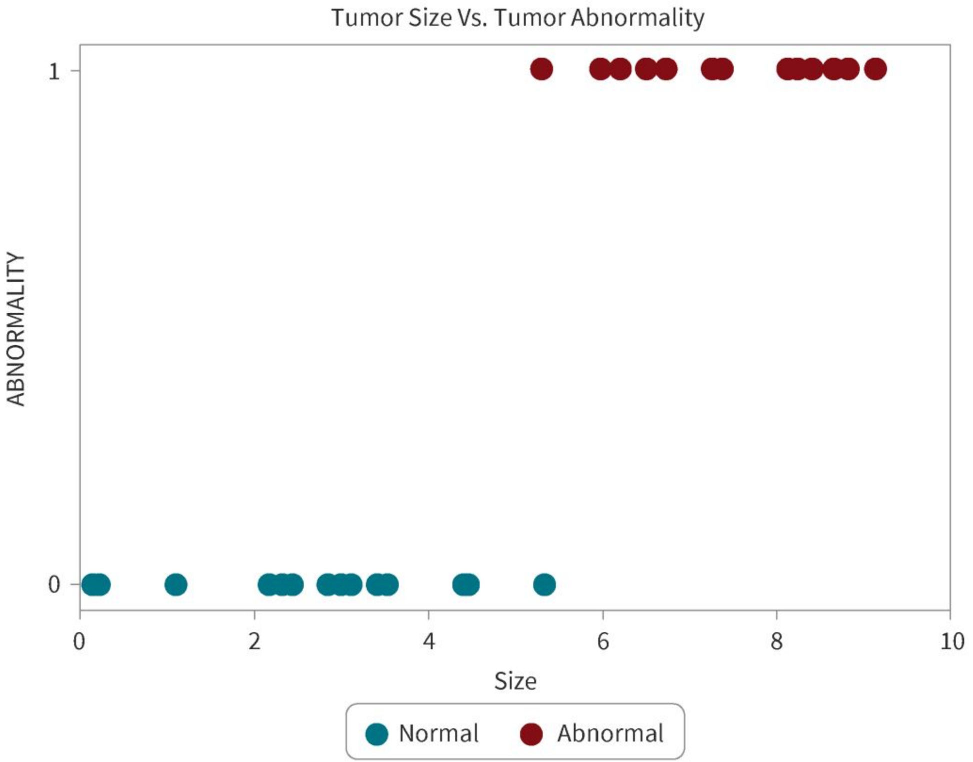
INPUT FEATURES	LABELS
Image pixels	What is present in the image: lung nodule, skin cancer, knee arthritis, etc.
Structured data (e.g. lab values, diagnosis codes, age)	Heart attack, sepsis, mortality, etc.
Unstructured text in nursing notes or pathology reports	Final diagnosis: stroke, appendicitis, etc.

EXAMPLE: *Using tumor size to classify as normal or abnormal*

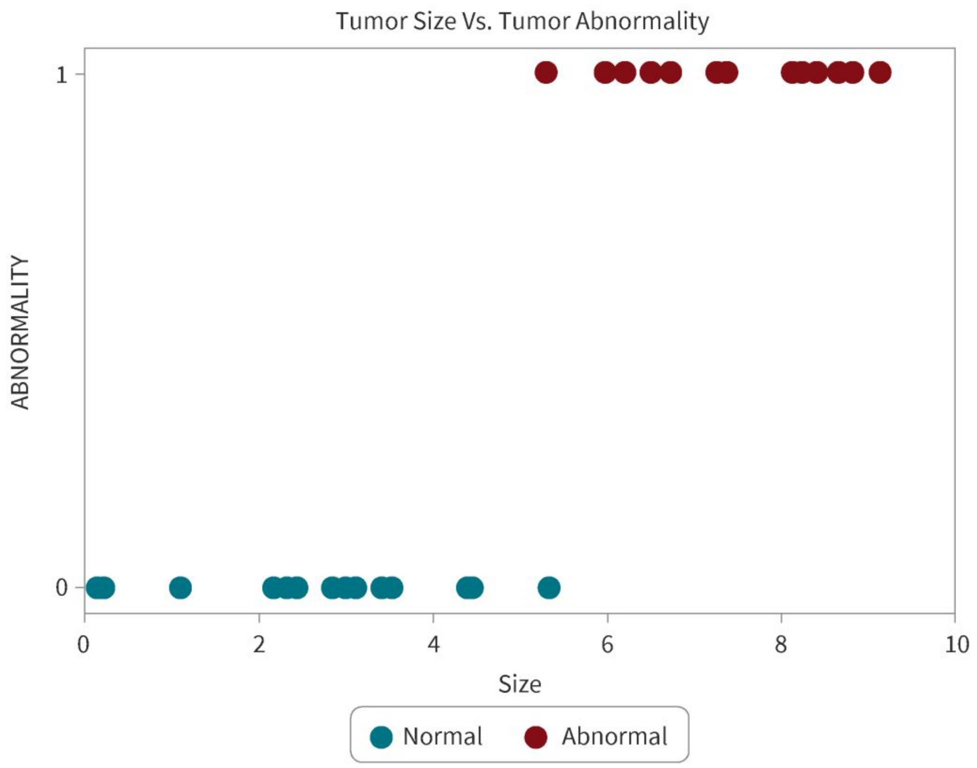
	SIZE	ABNORMALITY
0	4.093385	0
1	9.764256	1
2	7.187037	1
3	2.320848	0
4	6.273131	1
5	2.088424	0
6	3.380568	0
7	2.306047	0
8	6.427853	1

EXAMPLE: *Using tumor size to classify as normal or abnormal*

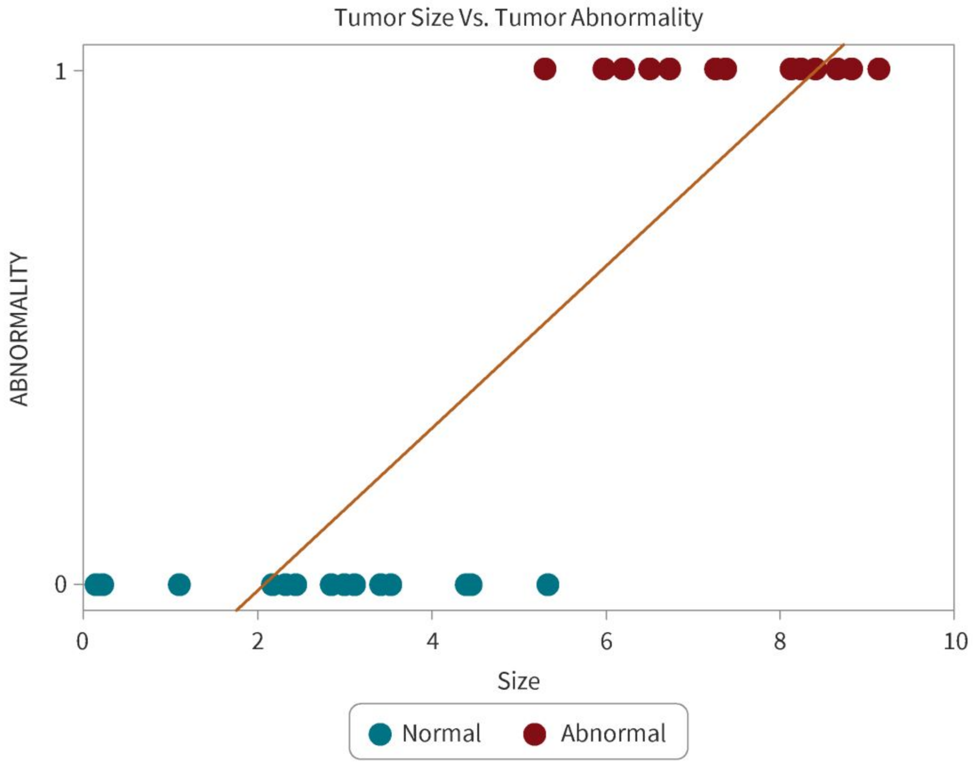
	SIZE	ABNORMALITY
0	4.093385	0
1	9.764256	1
2	7.187037	1
3	2.320848	0
4	6.273131	1
5	2.088424	0
6	3.380568	0
7	2.306047	0
8	6.427853	1



EXAMPLE: *Using tumor size to classify as normal or abnormal*



EXAMPLE: *Using tumor size to classify as normal or abnormal*

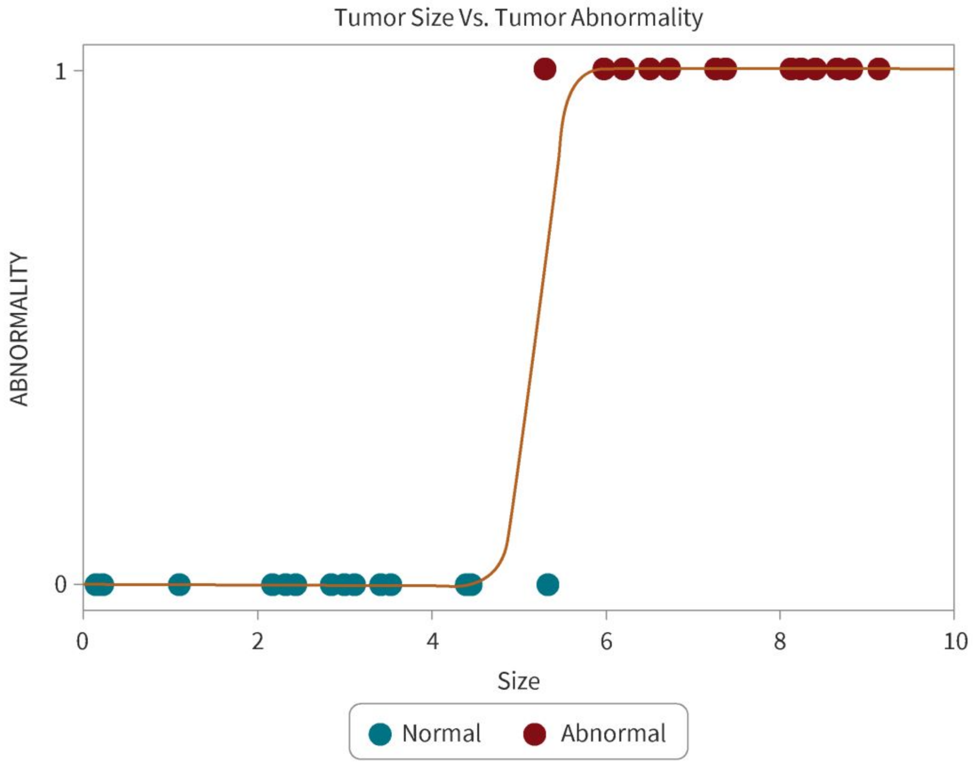


Potential problems with a linear function:

- What happens to outliers?

- How do we interpret the values of the function?

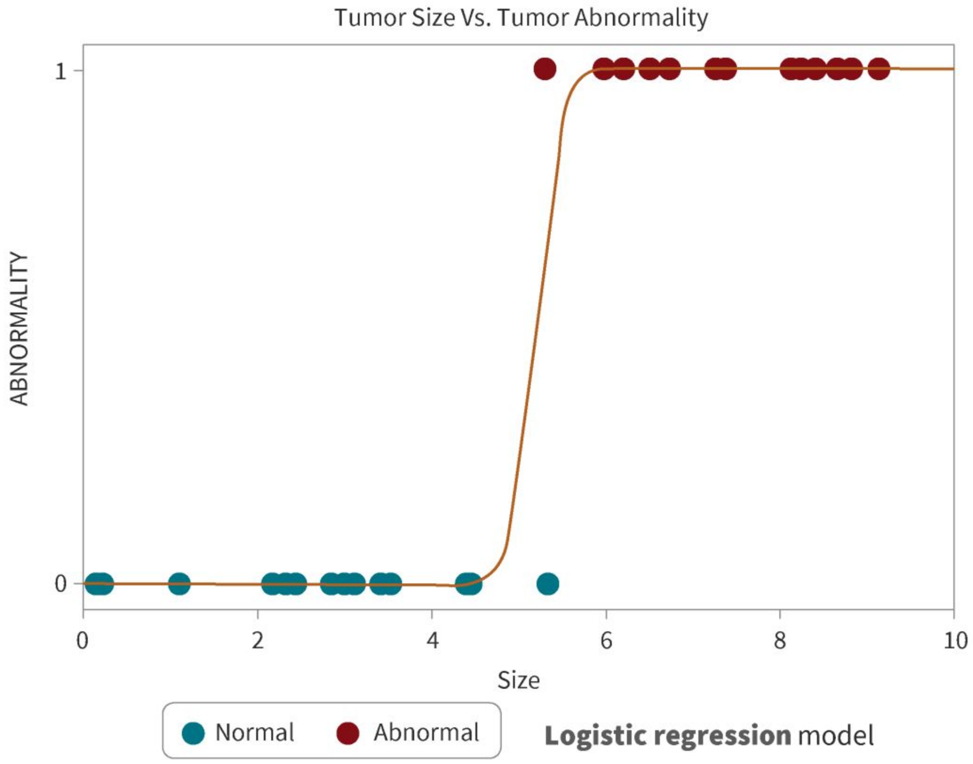
EXAMPLE: *Using tumor size to classify as normal or abnormal*



Sigmoid function:

A non-linear transformation that squashes all values to be between 0 and 1. The function values can be interpreted as probabilities.

EXAMPLE: *Using tumor size to classify as normal or abnormal*

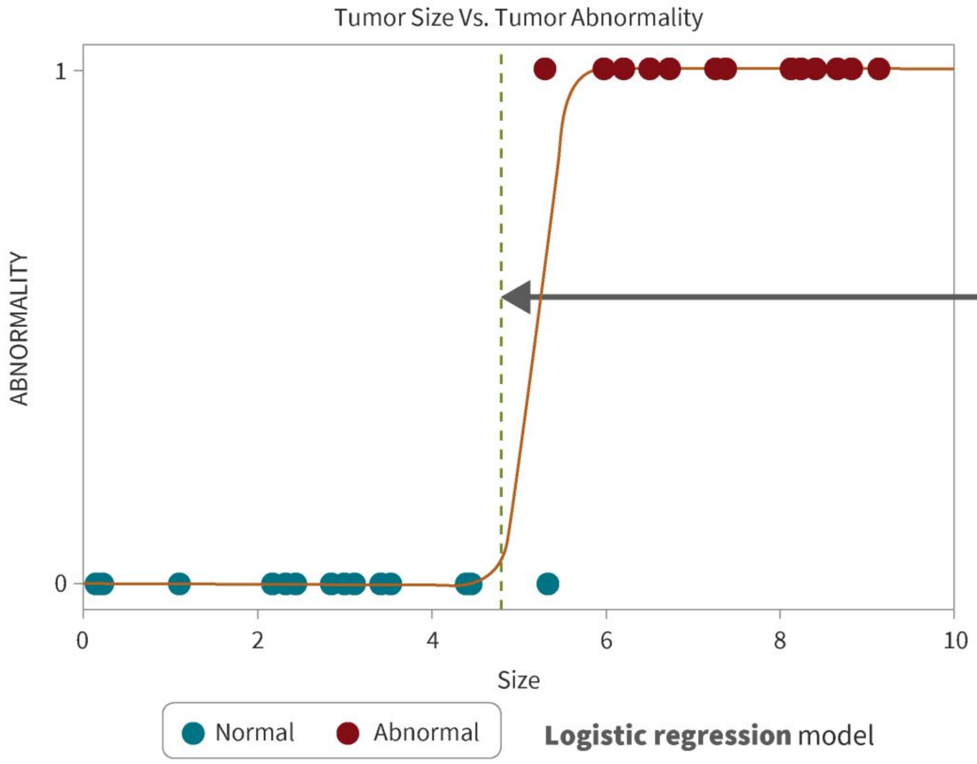


Logistic regression model

Sigmoid function:

A non-linear transformation that squashes all values to be between 0 and 1. The function values can be interpreted as probabilities.

EXAMPLE: *Using tumor size to classify as normal or abnormal*

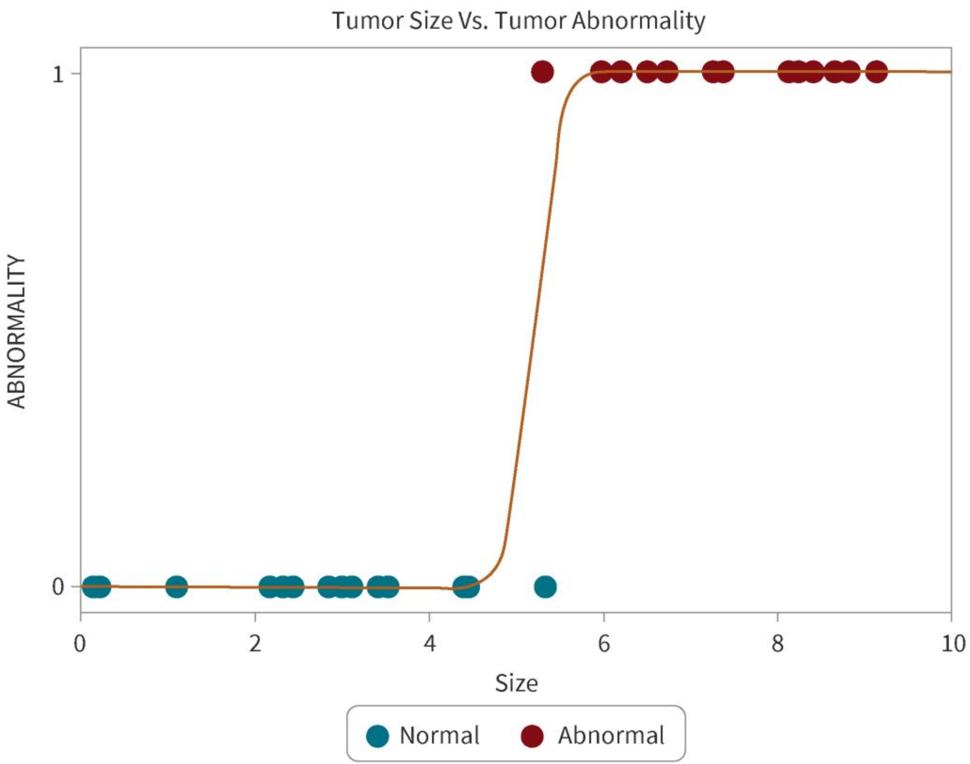


Adjust the operating point in order to calibrate for the clinical setting.

Logistic regression model

1 Feature = One-Dimensional

	SIZE	ABNORMALITY
0	4.093385	0
1	9.764256	1
2	7.187037	1
3	2.320848	0
4	6.273131	1
5	2.088424	0
6	3.380568	0
7	2.306047	0
8	6.427853	1

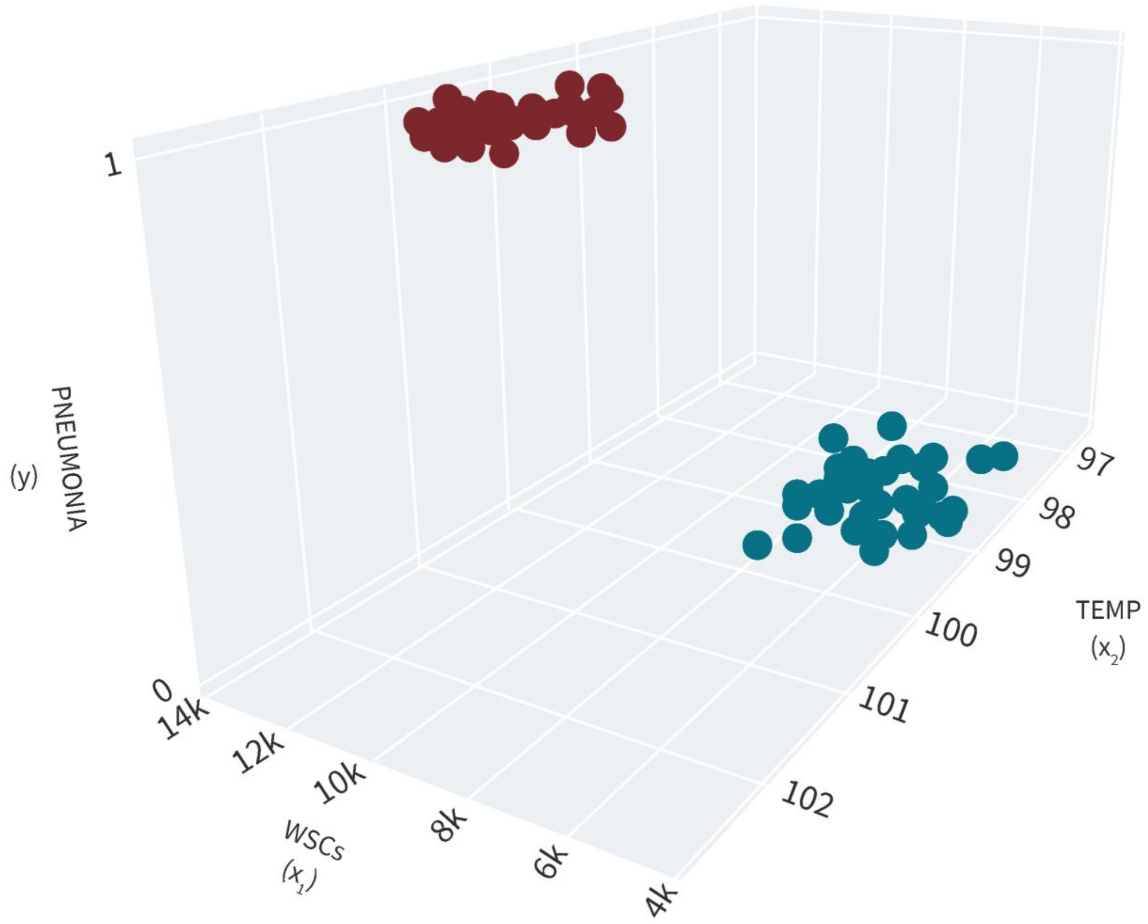


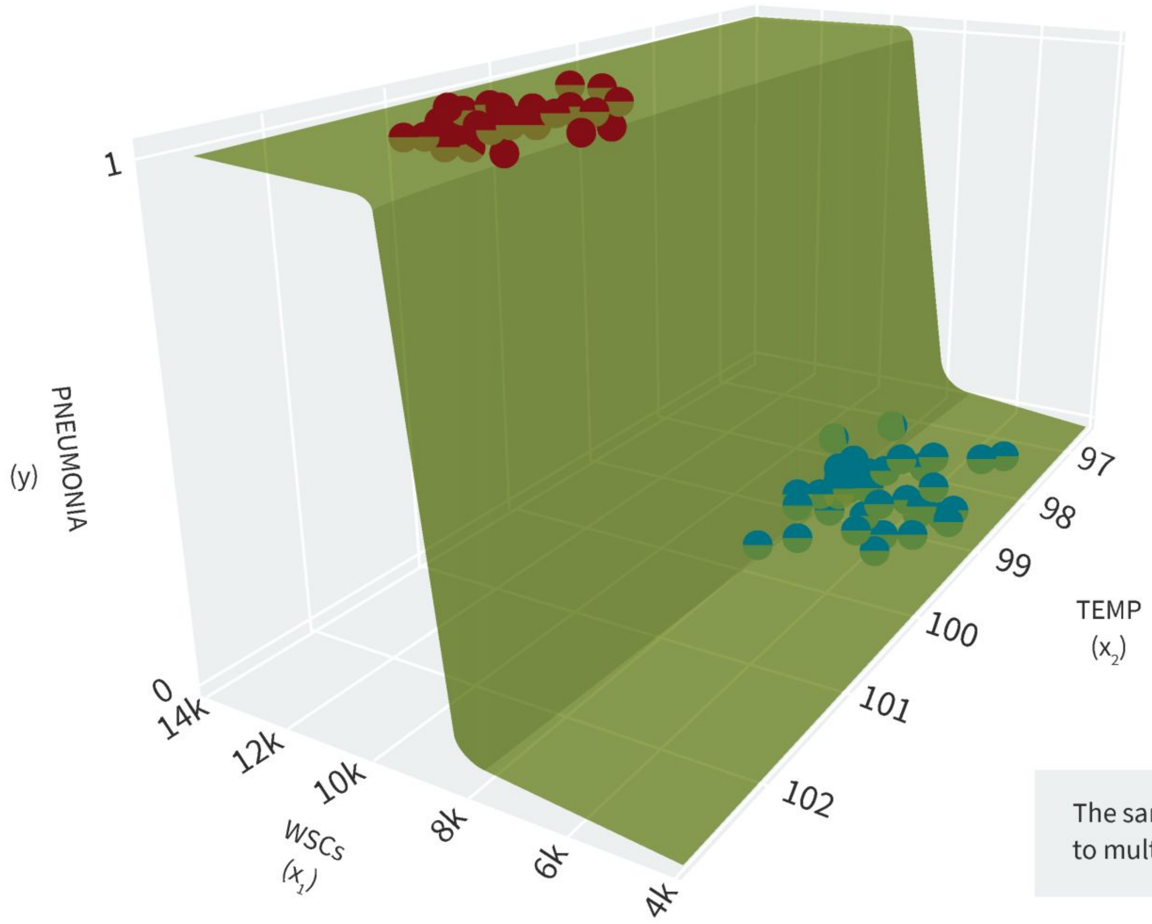
2 Features = Two-Dimensional

	WBCs	TEMP	ABNORMALITY
0	9797	100.181966	1
1	6562	97.604826	0
2	11449	101.491047	1
3	9250	97.778775	0
4	11520	100.401656	1
...	...	...	...
75	6357	98.461273	0
76	9139	98.748056	0
77	7675	98.614009	0
78	8515	98.181348	0
79	12679	100.529730	1

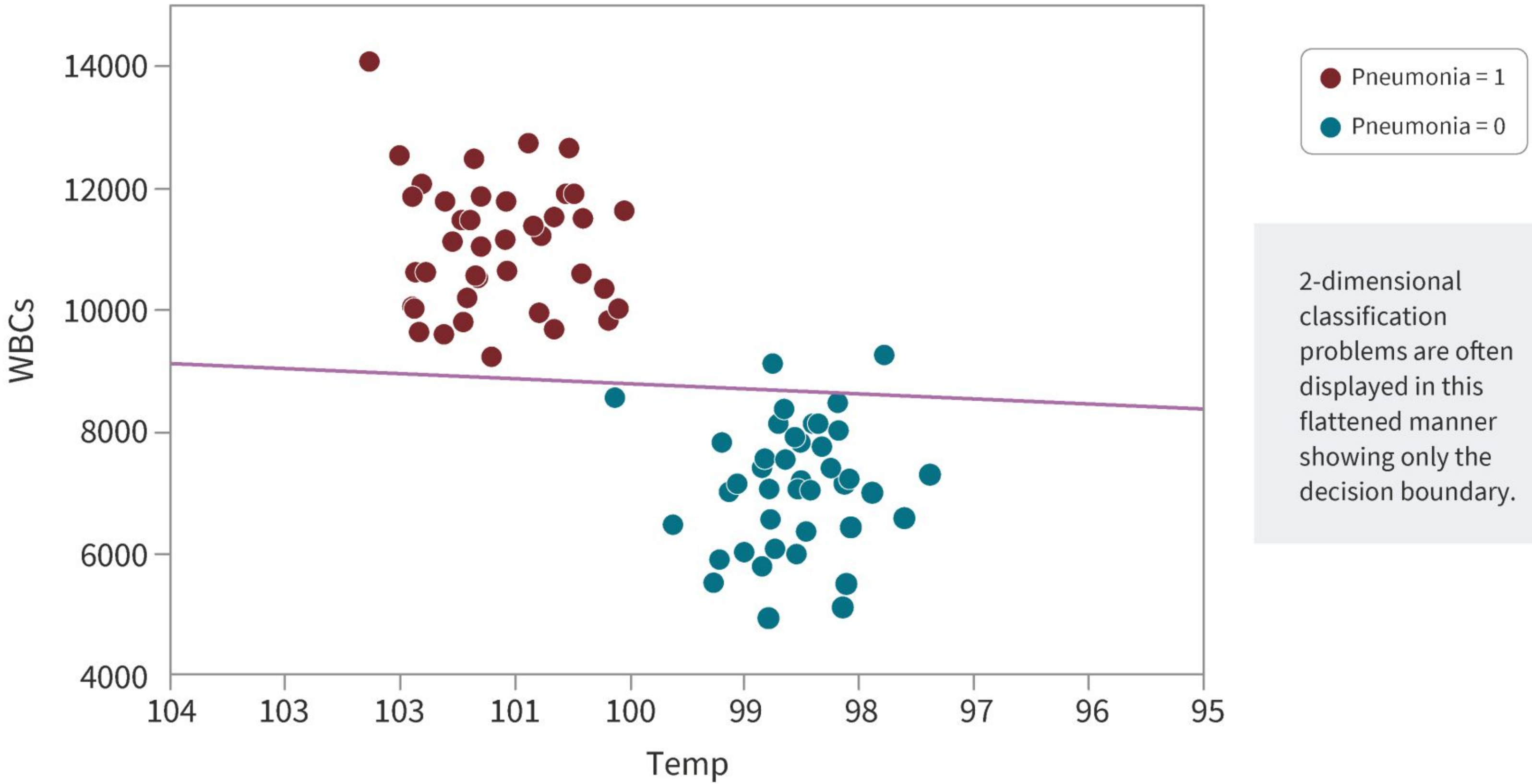
2 Features = Two-Dimensional

	WBCs	TEMP	ABNORMALITY
0	9797	100.181966	1
1	6562	97.604826	0
2	11449	101.491047	1
3	9250	97.778775	0
4	11520	100.401656	1
...	...	...	...
75	6357	98.461273	0
76	9139	98.748056	0
77	7675	98.614009	0
78	8515	98.181348	0
79	12679	100.529730	1
	x_1	x_2	y


1
0
PNEUMONIA
(y)
14k
12k
10k
8k
6k
4k
WSCs
(x_1)
97
98
99
100
101
102
TEMP
(x_2)



The same rules apply directly to multi-dimensional datasets.



2-dimensional classification problems are often displayed in this flattened manner showing only the decision boundary.

Different paradigms of machine learning

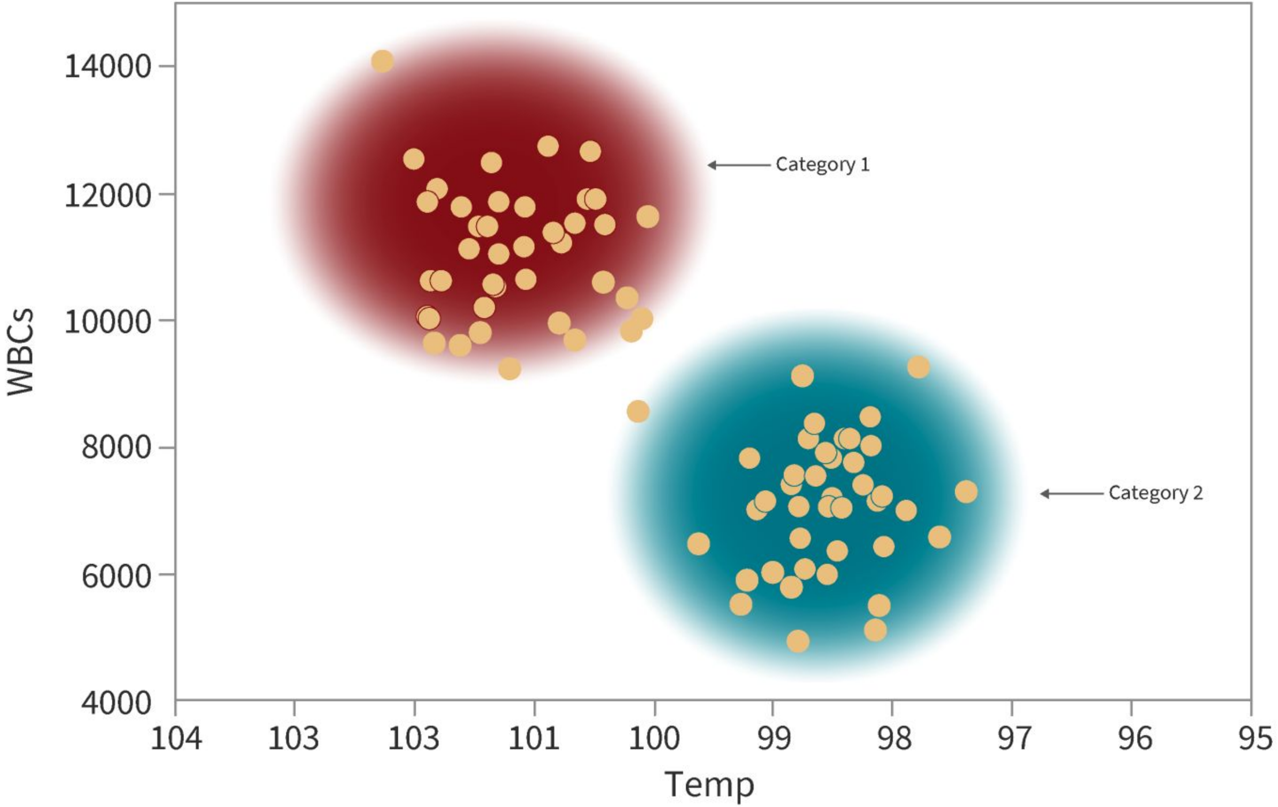
- Supervised learning
- Unsupervised learning
- Reinforcement learning

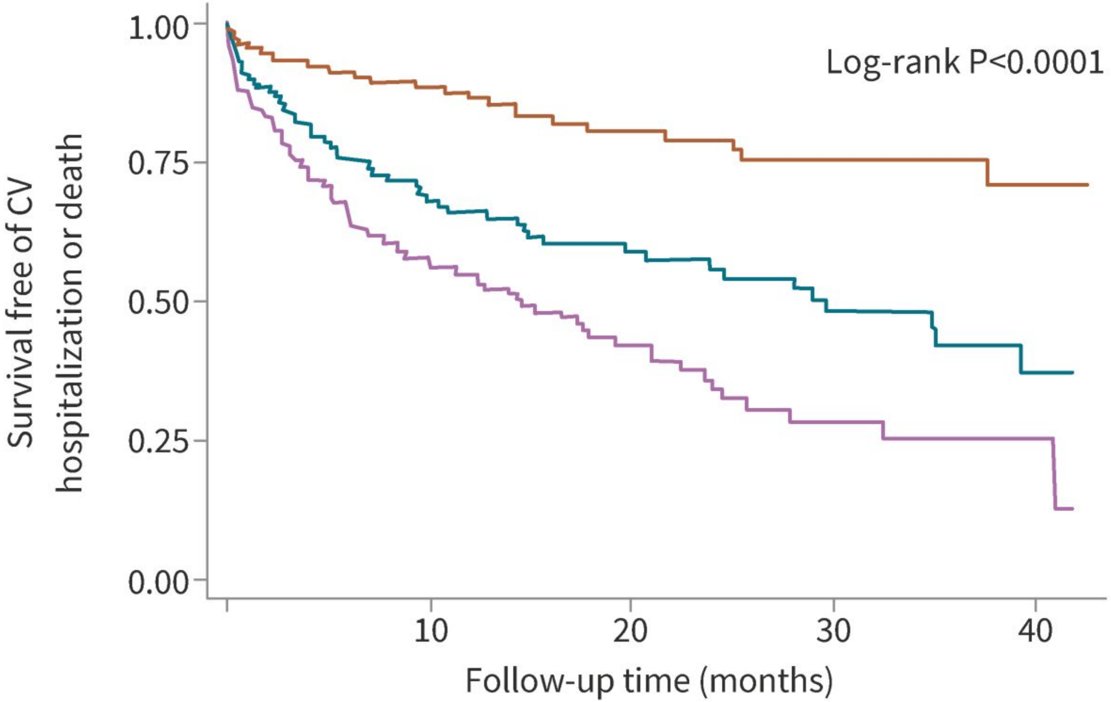
Different paradigms of machine learning

- Supervised learning
- **Unsupervised learning**
- **Reinforcement learning**

Unsupervised learning:
Finding patterns and underlying structure in **unlabeled** data as opposed to those with pre-determined labels

CLUSTERING UNLABELED DATA





Kaplan-Meier survival curves for the combined outcome of heart failure hospitalization, cardiovascular hospitalization, or death, stratified by the unsupervised clustering method of 397 patients using EMR data (67 clinical features)

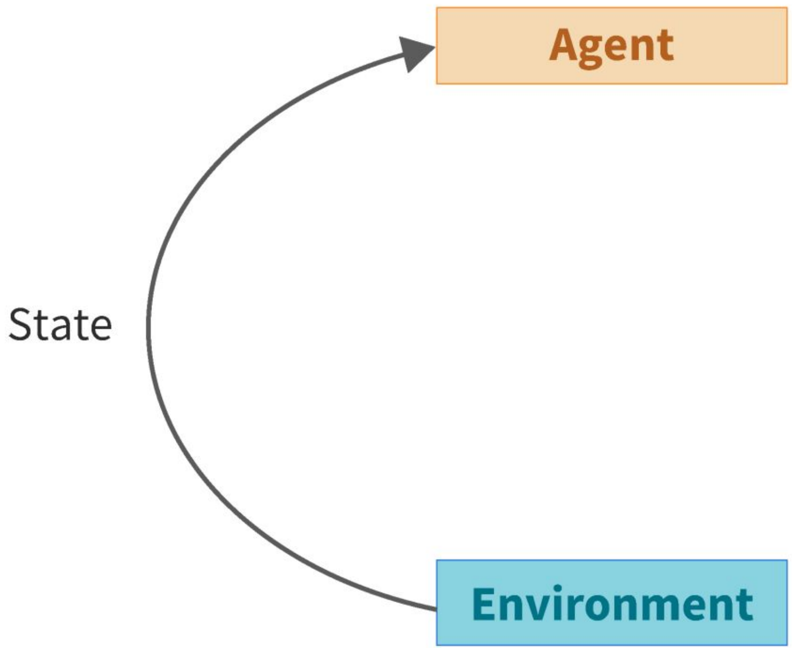
Reinforcement learning:
Paradigm of model as an “agent” interacting with an environment, continuously observing the current state of the environment and making decisions based on it

REINFORCEMENT LEARNING

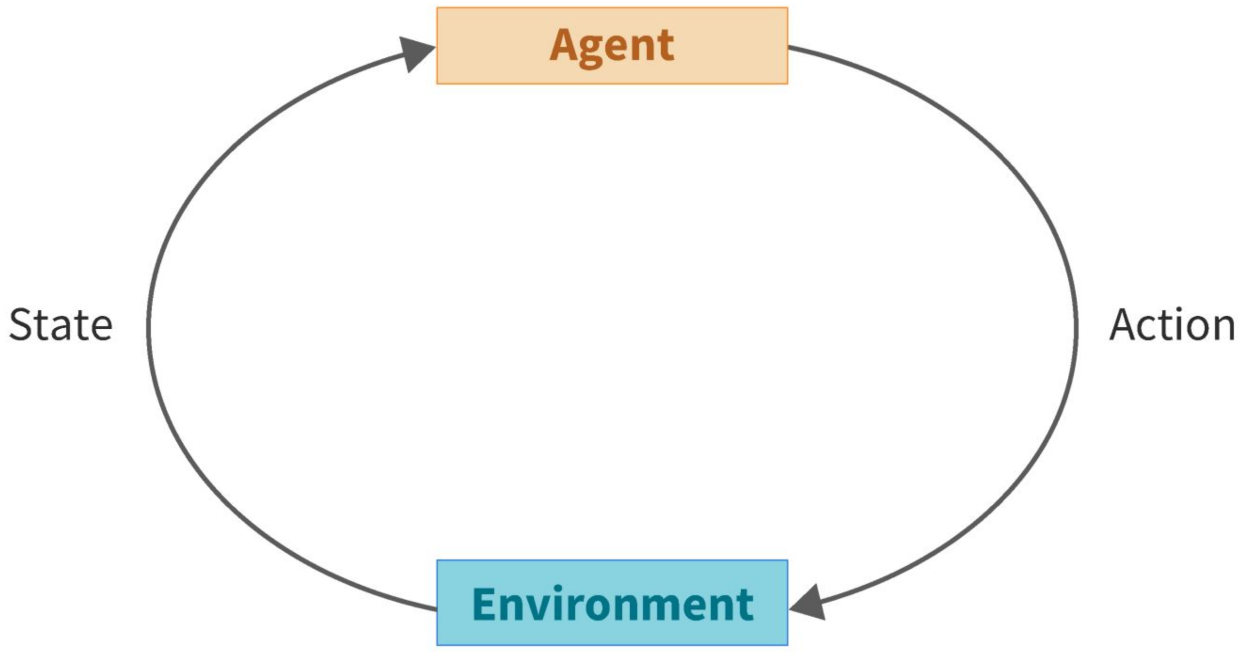
Agent

Environment

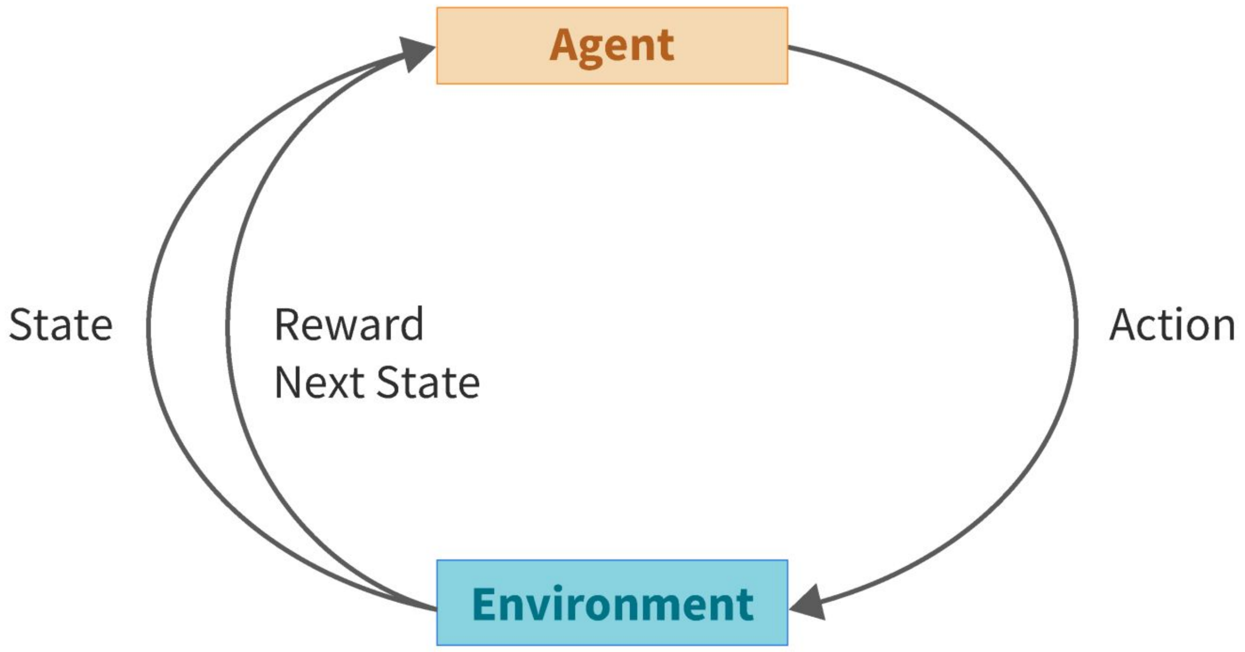
REINFORCEMENT LEARNING



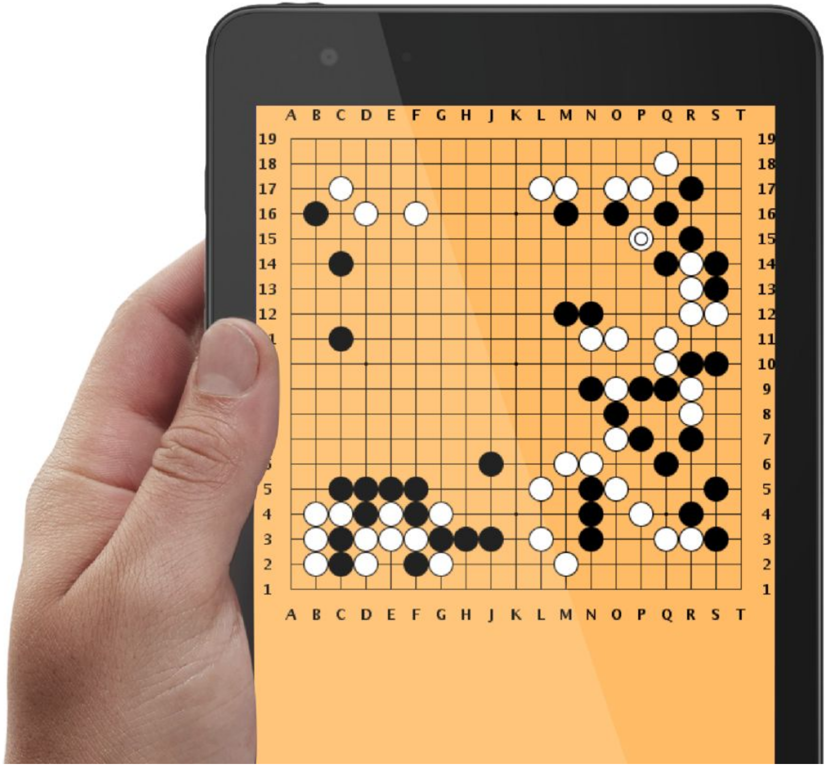
REINFORCEMENT LEARNING



REINFORCEMENT LEARNING



BEATING HUMANS ON THE GAME OF GO



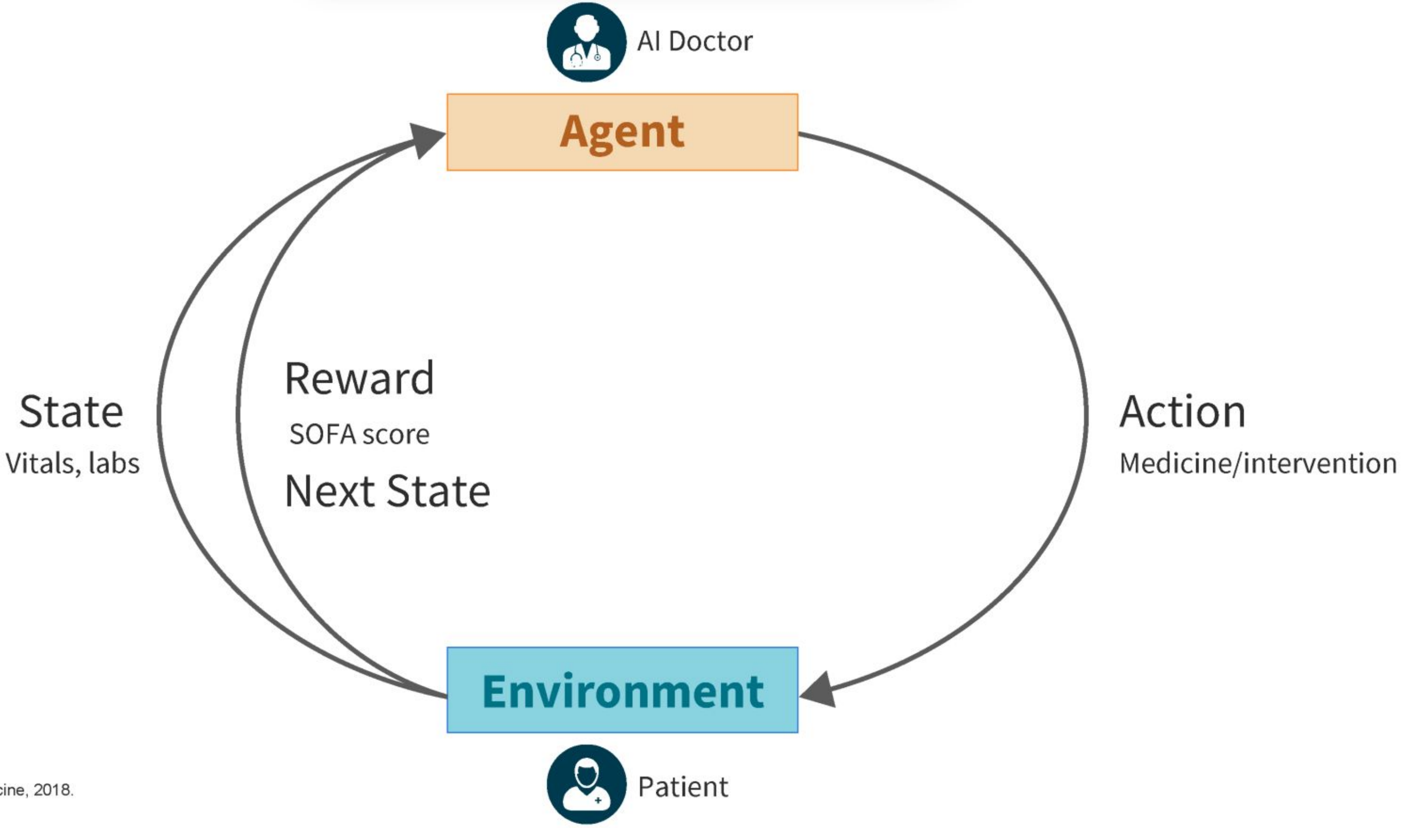
Objective : Win the game!

State: Position of all pieces

Action: Where to put the next piece down

Reward: 1 if win at the end of the game, 0 otherwise

REINFORCEMENT LEARNING



Komorowski et al. Nature Medicine, 2018.

Summary

Today we covered:

- The machine learning paradigm, and machine learning vs. traditional computer programming
- ML terminology and the ML training loop
- First ML models for regression (linear regression) and classification (logistic regression)
- Different machine learning paradigms: supervised learning, unsupervised learning, reinforcement learning

Coming up: Diving deeper into traditional methods for supervised learning, and then deep learning methods