

Machine Learning I

Mohamed Hussien

Introduction

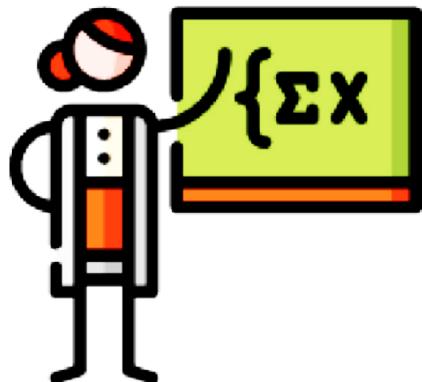
About Course

- Goal
- Scope
- Lecture Type
- Questions
- Feedback

Why These Sessions?

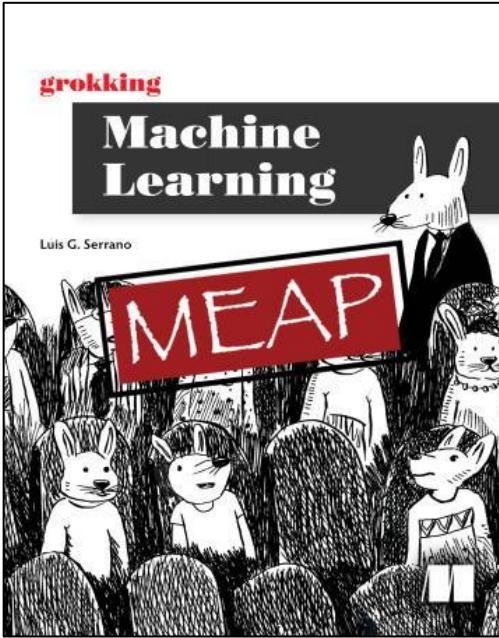


Music



Machine learning

References



Tools

colab

Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

What is Artificial Intelligence?

“

The set of all tasks in which a computer can make decisions.

What is Artificial Intelligence?

“

The set of all tasks in which a computer can make decisions.



What is Artificial Intelligence?

“

The set of all tasks in which a computer can make decisions.

Humans make decisions based on:

- Logic And Reasoning
- Experience



What is Machine Learning?

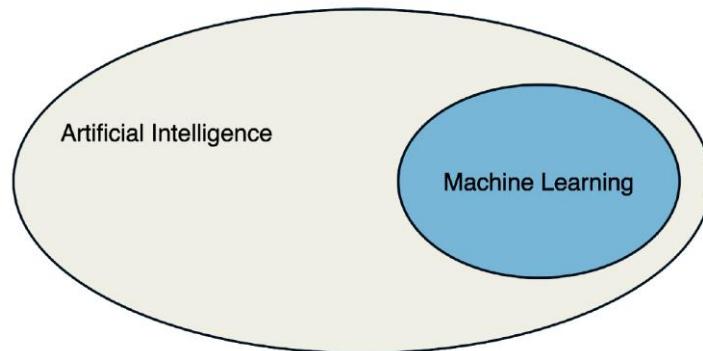
“

The set of all tasks in which a computer can make decisions based on **data**.

What is Machine Learning?

“

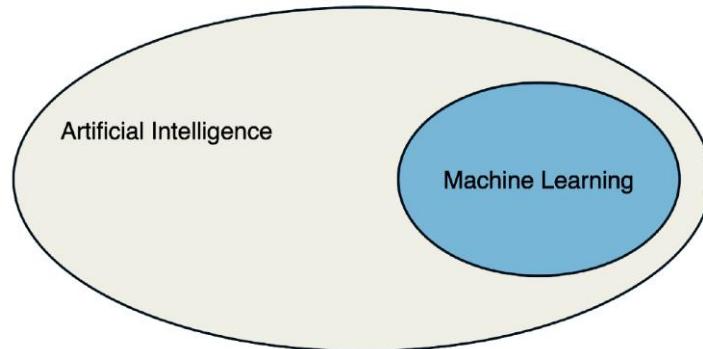
The set of all tasks in which a computer can make decisions based on **data**.



What is Machine Learning?

“

The set of all tasks in which a computer can make decisions based on **data**.



Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

How Do Humans Learn?

3 steps for humans to learn:

1- We **remember** past situations that were similar.

2- We **formulate** a general rule (Model).

3- We use this rule to **predict** what may happen in the future.

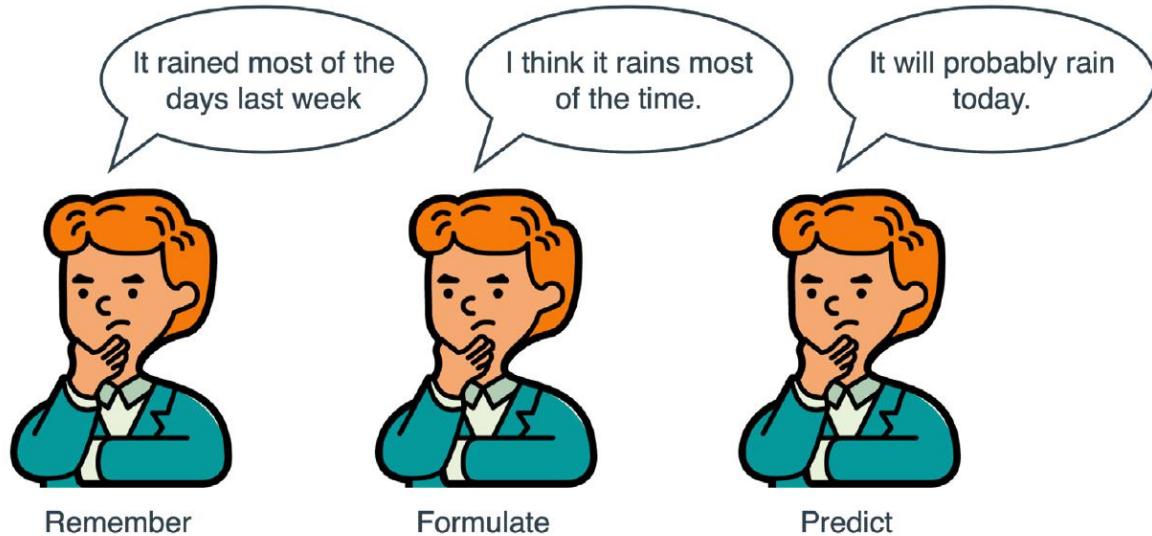
How Do Humans Learn?

3 steps for humans to learn:

1- We **remember** past situations that were similar.

2- We **formulate** a general rule (Model).

3- We use this rule to **predict** what may happen in the future.



Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

What is Model?

“

A set of rules that represent our data and can be used to make predictions.

$$f(\text{dog}) \rightarrow \text{dog}$$

$$f(\text{cat}) \rightarrow \text{cat}$$

Example: Spam or Ham?!

- Bob likes to send us a lot of emails.
- In particular, a lot of his emails are spam.
- It is Saturday, and we just got a notification of an email from Bob.
- Can we guess if it is spam or ham without looking at the email?

Let's use the **remember-formulate-predict** method.



Example: Spam or Ham?!

Model 1:

Remember

4 of the last 10 emails from Bob were spam.

Formulate(Model)

40% of Bob's emails are spam.

Should it be true?!

Predict

This new email is Ham.

Example: Spam or Ham?!

Model 1:

Remember

4 of the last 10 emails from Bob were spam.

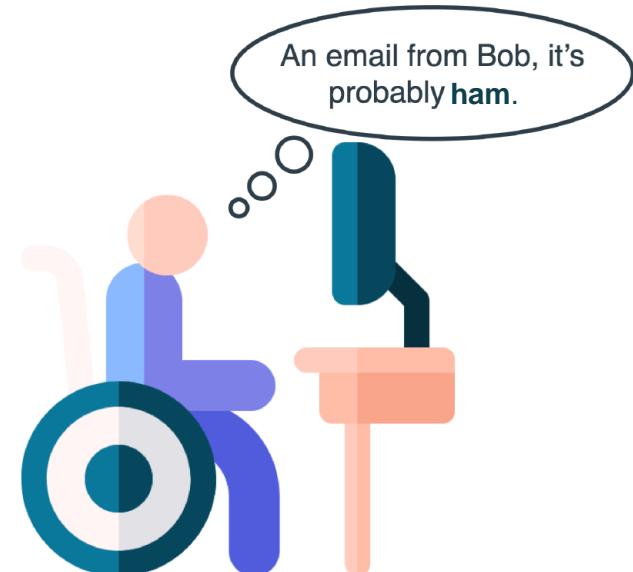
Formulate(Model)

40% of Bob's emails are spam.

Should it be true?!

Predict

This new email is Ham.



Example: Spam or Ham?!

Model 1:

Remember

4 of the last 10 emails from Bob were spam.

Formulate(Model)

40% of Bob's emails are spam.

Should it be true?!

Predict

This new email is Ham.

Can we have a better model?!



Example: Spam or Ham?!

Model 2 (Email Day):

Remember

Mo	Tu	Sa	Su	Su	We	Fr	Sa	Tu	Th
H	H	S	S	S	H	H	S	H	H

Formulate(Model)

Every email that Bob sends during the week is ham,
and during the weekend is spam.

Predict

Today is Saturday, this email is spam.

Example: Spam or Ham?!

Model 2 (Email Day):

Remember

Mo	Tu	Sa	Su	Su	We	Fr	Sa	Tu	Th
H	H	S	S	S	H	H	S	H	H

Formulate(Model)

Every email that Bob sends during the week is ham, and during the weekend is spam.

Predict

Today is Saturday, this email is spam.



Example: Spam or Ham?!

Model 3 (Email Size):

Remember

1KB 2KB 16KB 20KB 18KB 3KB 5KB 25KB 1KB 3KB

H H S S S H H S H H

Formulate(Model)

Any email larger of size 10KB or more is spam,
and any email of size less than 10KB is ham.

Predict

The email size is 19 KB, this email is spam.

Example: Spam or Ham?!

Model 3 (Email Size):

Remember

1KB 2KB 16KB 20KB 18KB 3KB 5KB 25KB 1KB 3KB

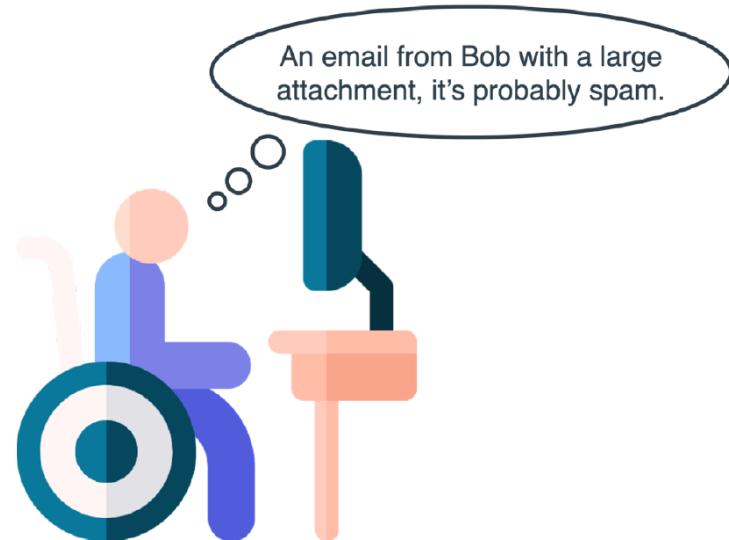
H H S S S H H S H H

Formulate(Model)

Any email larger of size 10KB or more is spam,
and any email of size less than 10KB is ham.

Predict

The email size is 19 KB, this email is spam.



Example: Spam or Ham?!

Model 3 (Email Size):

Remember

1KB 2KB 16KB 20KB 18KB 3KB 5KB 25KB 1KB 3KB

H H S S S H H S H H

Formulate(Model)

Any email larger of size 10KB or more is spam,
and any email of size less than 10KB is ham.

Predict

The email size is 19 KB, this email is spam.

What we call email size and email day? **Features**



What is Feature?

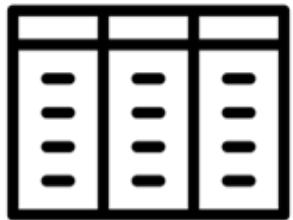
“

Any property or characteristic of the data that the model can use to make predictions.

What is Feature?

“

Any property or characteristic of the data that the model can use to make predictions.

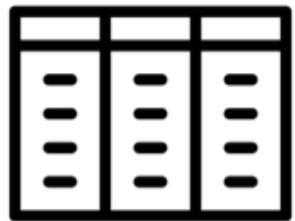


Tabular

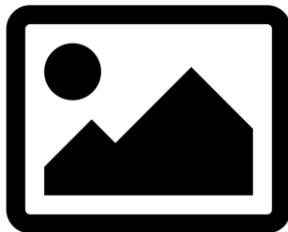
What is Feature?

“

Any property or characteristic of the data that the model can use to make predictions.



Tabular

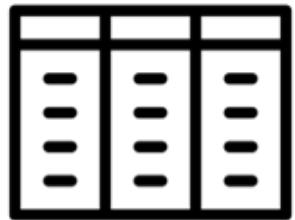


Image

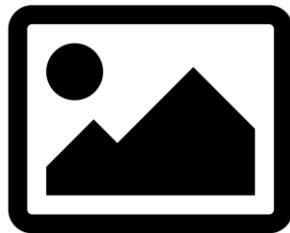
What is Feature?

“

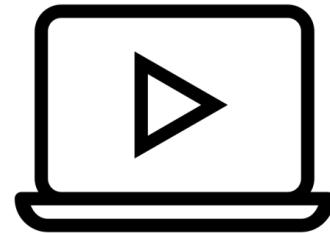
Any property or characteristic of the data that the model can use to make predictions.



Tabular



Image

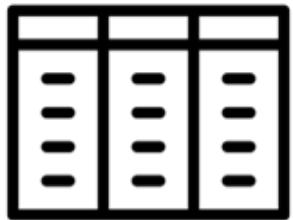


Video

What is Feature?

“

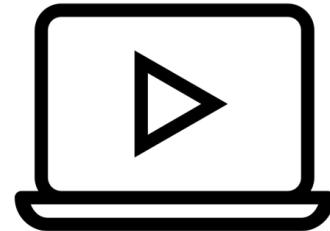
Any property or characteristic of the data that the model can use to make predictions.



Tabular



Image



Video

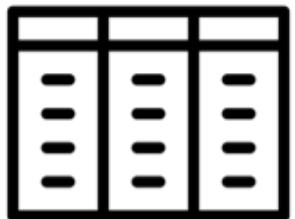


Text

What is Feature?

“

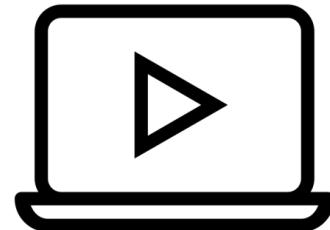
Any property or characteristic of the data that the model can use to make predictions.



Tabular



Image



Video



Text



Audio

Example: Spam or Ham?!

Using the 2 features:



Example: Spam or Ham?!

Using the 2 features:

Model 4

If an email is larger than 10KB or it is sent on the weekend, then it is classified as spam. Otherwise, it is classified as ham.

Example: Spam or Ham?!

Using the 2 features:

Model 4

If an email is larger than 10KB or it is sent on the weekend, then it is classified as spam. Otherwise, it is classified as ham.

Model 5

If the email is sent during the week, then it must be larger than 15KB to be classified as spam. If it is sent during the weekend, then it must be larger than 5KB to be classified as spam. Otherwise, it is classified as ham.

Example: Spam or Ham?!

Using the 2 features:

Model 4

If an email is larger than 10KB or it is sent on the weekend, then it is classified as spam. Otherwise, it is classified as ham.

Model 5

If the email is sent during the week, then it must be larger than 15KB to be classified as spam. If it is sent during the weekend, then it must be larger than 5KB to be classified as spam. Otherwise, it is classified as ham.

Model 6

Consider the number of the day, where Monday is 0, Tuesday is 1, ... Saturday is 5, and Sunday is 6. If we add the number of the day and the size of the email (in KB), and the result is 12 or more, then the email is classified as spam. Otherwise, it is classified as ham.

Example: Spam or Ham?!

Using the 2 features:

Model 4

If an email is larger than 10KB or it is sent on the weekend, then it is classified as spam. Otherwise, it is classified as ham.

Model 5

If the email is sent during the week, then it must be larger than 15KB to be classified as spam. If it is sent during the weekend, then it must be larger than 5KB to be classified as spam. Otherwise, it is classified as ham.

Model 6

Consider the number of the day, where Monday is 0, Tuesday is 1, ... Saturday is 5, and Sunday is 6. If we add the number of the day and the size of the email (in KB), and the result is 12 or more, then the email is classified as spam. Otherwise, it is classified as ham.

Could it be more complex?

Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

How Do Machines Learn?

3 steps for Machines to learn:

Remember

Look at a huge amount of data.

Formulate

Create models by going through many rules and formulas, and check which model fits the data best.

Predict

Use the model to make predictions about future data.

Machines can look at a huge amount of data and formulate a model quickly.

Example: Spam or Ham?!

More than 2 features:

Model 7

- If the email **has two or more spelling mistakes**, then it is classified as spam.
- Otherwise, if it has an attachment **larger than 20KB**, it is classified as spam.
- Otherwise, if **the sender is not in our contact list**, it is classified as spam.
- Otherwise, if it **has the words “buy” and “win”**, it is classified as spam.
- Otherwise, it is classified as ham.

Example: Spam or Ham?!

More than 2 features:

Model 7

- If the email **has two or more spelling mistakes**, then it is classified as spam.
- Otherwise, if it has an attachment **larger than 20KB**, it is classified as spam.
- Otherwise, if **the sender is not in our contact list**, it is classified as spam.
- Otherwise, if it **has the words “buy” and “win”**, it is classified as spam.
- Otherwise, it is classified as ham.

Model 8

If $(\text{size}) + 10 \times (\text{number of spelling mistakes}) - (\text{number of appearances of the word 'mom'}) + 4 \times (\text{number of appearances of the word 'buy'}) > 10$, then we classify the message as spam.
Otherwise we classify it as ham.

Example: Spam or Ham?!

More than 2 features:

Model 7

- If the email **has two or more spelling mistakes**, then it is classified as spam.
- Otherwise, if it has an attachment **larger than 20KB**, it is classified as spam.
- Otherwise, if **the sender is not in our contact list**, it is classified as spam.
- Otherwise, if it **has the words “buy” and “win”**, it is classified as spam.
- Otherwise, it is classified as ham.

Model 8

If $(\text{size}) + 10 \times (\text{number of spelling mistakes}) - (\text{number of appearances of the word 'mom'}) + 4 \times (\text{number of appearances of the word 'buy'}) > 10$, then we classify the message as spam.
Otherwise we classify it as ham.

What is the best model?!

Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

How could these problems be solved?!

- Predicting housing prices based on their size, number of rooms, location, etc.
- Detecting spam and non-spam e-mails based on the words of the e-mail, the sender, etc.
- Recommending videos or movies to a user (for example in YouTube, IMDB, etc.).
- Playing games like chess or Go.

How could these problems be solved?!

- Predicting housing prices based on their size, number of rooms, location, etc.
- Detecting spam and non-spam e-mails based on the words of the e-mail, the sender, etc.
- Recommending videos or movies to a user (for example in YouTube, IMDB, etc.).
- Playing games like chess or Go.

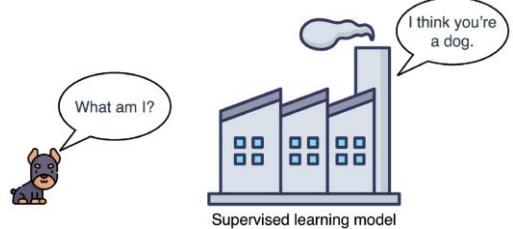
They are different, right?

Machine Learning Models Types



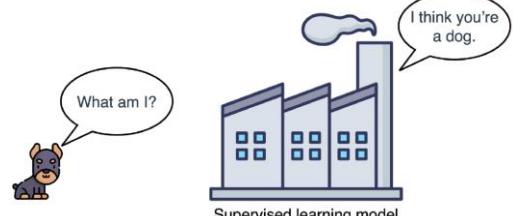
Machine Learning Models Types

Supervised Learning

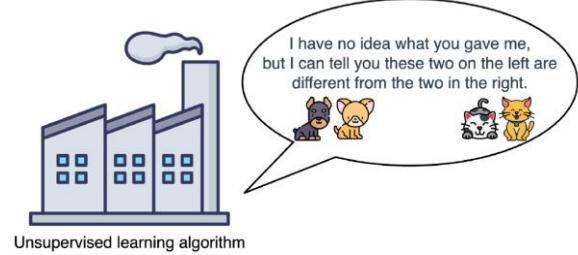


Machine Learning Models Types

Supervised Learning

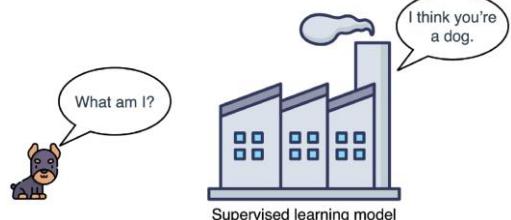


Unsupervised Learning

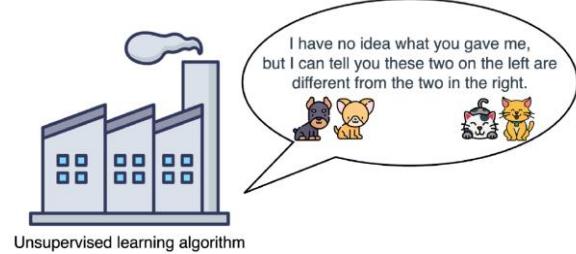


Machine Learning Models Types

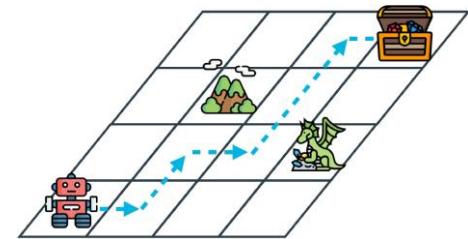
Supervised Learning



Unsupervised Learning

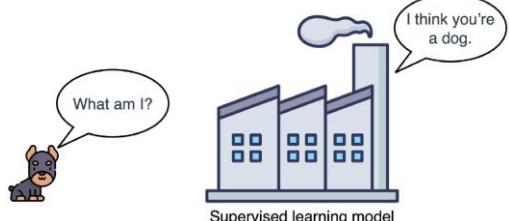


Reinforcement Learning

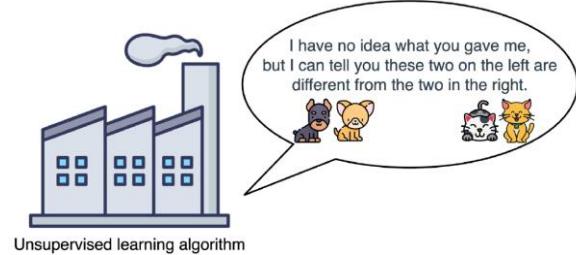


Machine Learning Models Types

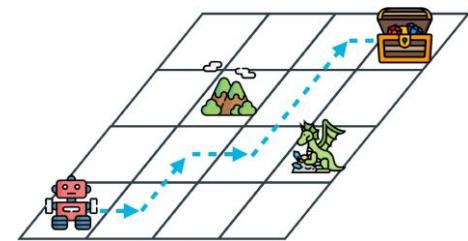
Supervised Learning



Unsupervised Learning



Reinforcement Learning



Let's talk first about the difference between **Data, Labels, Predictions, and Features...**

Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

What are Data, Labels, Predictions, Features?

What are Data, Labels, Predictions, Features?

“

Data is simply information.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

“

Predictions are the guess that the model makes.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

“

Predictions are the guess that the model makes.

“

Features are any property of the data that the model can use to make predictions.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

“

Predictions are the guess that the model makes.

“

Features are any property of the data that the model can use to make predictions.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

“

Predictions are the guess that the model makes.

“

Features are any property of the data that the model can use to make predictions.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

“

Predictions are the guess that the model makes.

“

Features are any property of the data that the model can use to make predictions.

What are Data, Labels, Predictions, Features?

“

Data is simply information.

“

Labels are values that we try to predict.

Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

“

Predictions are the guess that the model makes.

“

Features are any property of the data that the model can use to make predictions.

Data types



Data types

Labelled Data



Dog



Dog



Cat



Cat

Data types

Labelled Data



Labelled Data



Data types

Labelled Data



Dog Dog



Cat Cat

Labelled Data



18 lbs 14 lbs



12 lbs 9 lbs

Unlabelled Data



Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

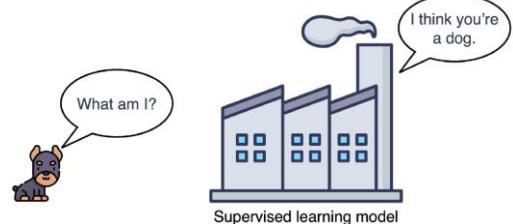
Supervised Learning

Unsupervised learning

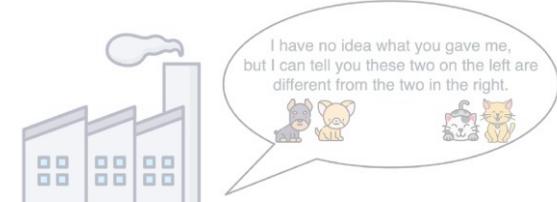
Reinforcement Learning

Machine Learning Models Types

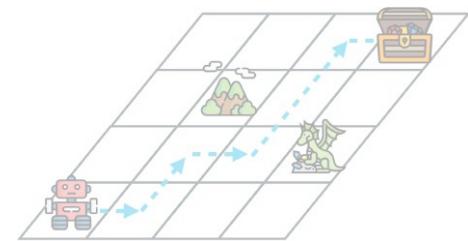
Supervised Learning



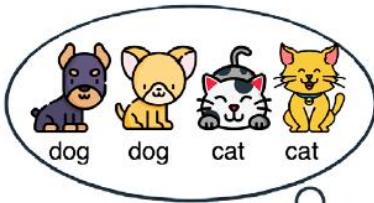
Unsupervised Learning



Reinforcement Learning



Supervised Learning



Supervised learning model

Remember



Supervised learning model

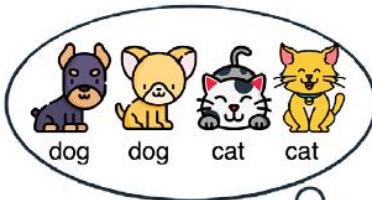
Formulate



Supervised learning model

Predict

Supervised Learning



Supervised learning model

Remember

What are supervised models types?



Supervised learning model

Formulate



Supervised learning model

Predict

Data types

Labelled Data



Dog Dog



Cat Cat

Labelled Data



18 lbs 14 lbs



12 lbs 9 lbs

Unlabelled Data



Data types

Labelled Data

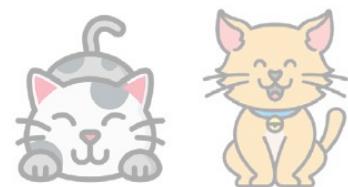


Categorical Data

Labelled Data



Unlabelled Data



Data types

Labelled Data



Dog



Cat

Categorical Data

Labelled Data



18 lbs

14 lbs



12 lbs

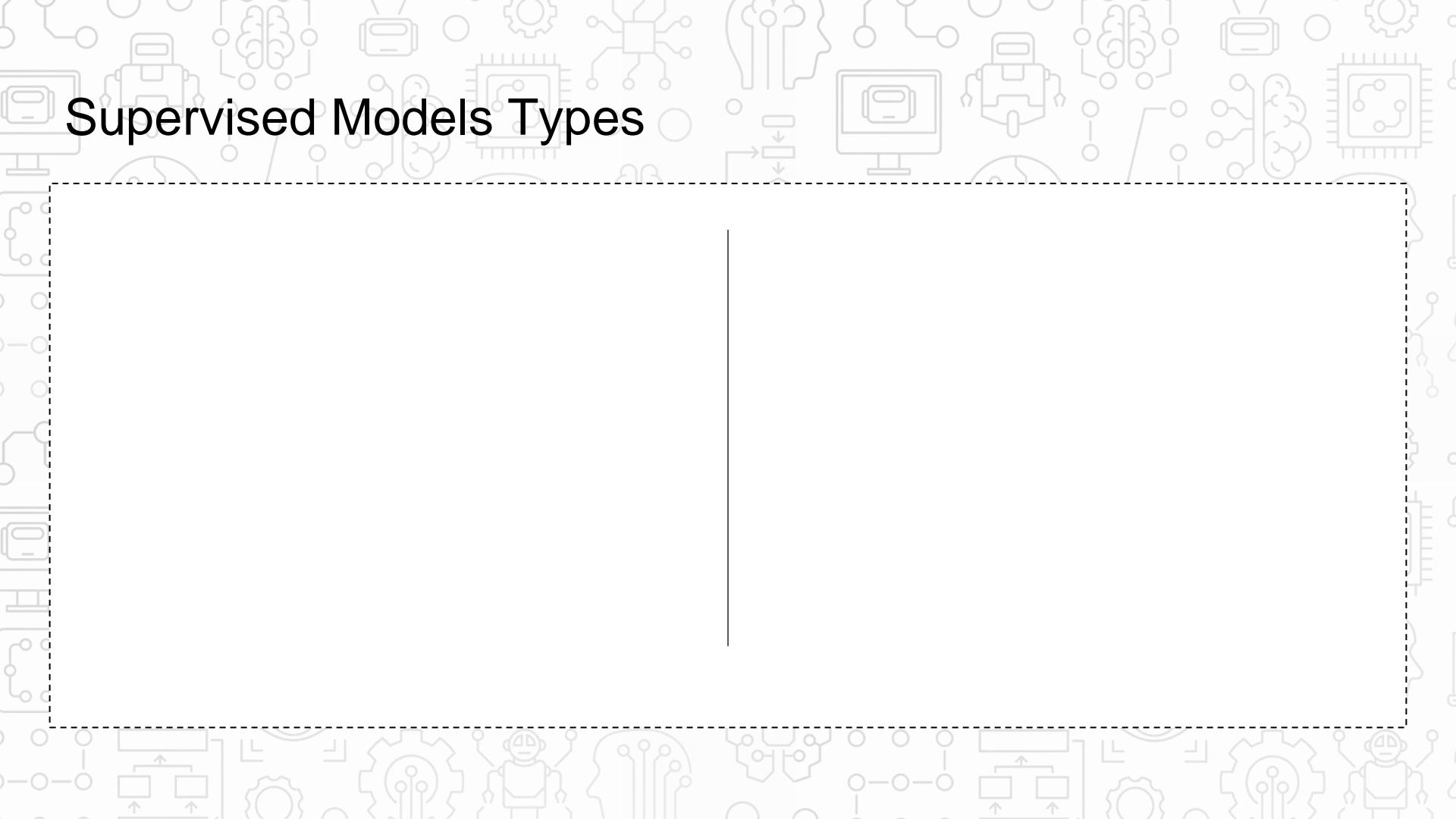
9 lbs

Numerical Data

Unlabelled Data



Supervised Models Types



Supervised Models Types

Classification Models

“

The types of models that predict **categorical** data. The output of a classification model is a **category**, or a **state**, such as the type of animal (cat or dog).

Example: Email spam detection model

Supervised Models Types

Classification Models

“

The types of models that predict **categorical** data. The output of a classification model is a **category**, or a **state**, such as the type of animal (cat or dog).

Example: Email spam detection model

Regression Models

“

The types of models that predict **numerical** data. The output of a regression model is a **number**, such as the weight of the animal.

Example: Housing prices model

Supervised Models Types

Classification Models

“

The types of models that predict **categorical** data. The output of a classification model is a **category**, or a **state**, such as the type of animal (cat or dog).

Example: Email spam detection model

Any other examples?

Regression Models

“

The types of models that predict **numerical** data. The output of a regression model is a **number**, such as the weight of the animal.

Example: Housing prices model

Supervised Models Types

Example

Supervised Models Types

Example



Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

Reinforcement Learning

Machine Learning Models Types

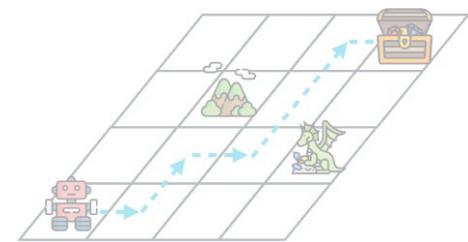
Supervised Learning



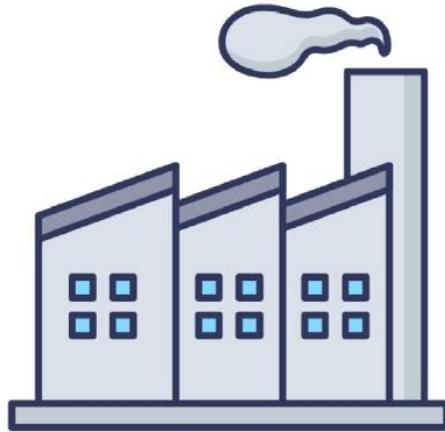
Unsupervised Learning



Reinforcement Learning



Unsupervised Learning

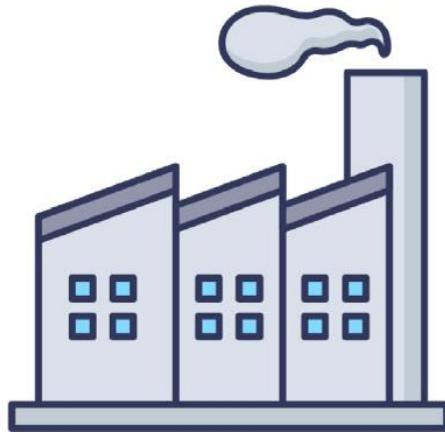


Unsupervised learning algorithm

I have no idea what you gave me,
but I can tell you these two on the left are
different from the two in the right.



Unsupervised Learning



Unsupervised learning algorithm

I have no idea what you gave me,
but I can tell you these two on the left are
different from the two in the right.



What are unsupervised models types?

Unsupervised Models Types

Unsupervised Models Types

Clustering Algorithms

“

The algorithms that group data into clusters based on similarity.

Unsupervised Models Types

Clustering Algorithms

“

The algorithms that group data into clusters based on similarity.

Dimensionality Reduction Algorithms

“

The algorithms that simplify our data and faithfully describe it with fewer features.

Unsupervised Models Types

Clustering Algorithms

“

The algorithms that group data into clusters based on similarity.

Dimensionality Reduction Algorithms

“

The algorithms that simplify our data and faithfully describe it with fewer features.

Generative Algorithms

“

The algorithms that are able to generate new data points that resemble the existing data.

Unsupervised Models Types

Clustering Algorithms

“

The algorithms that group data into clusters based on similarity.

Dimensionality Reduction Algorithms

“

The algorithms that simplify our data and faithfully describe it with fewer features.

Generative Algorithms

“

The algorithms that are able to generate new data points that resemble the existing data.

Let's take an example from each type...

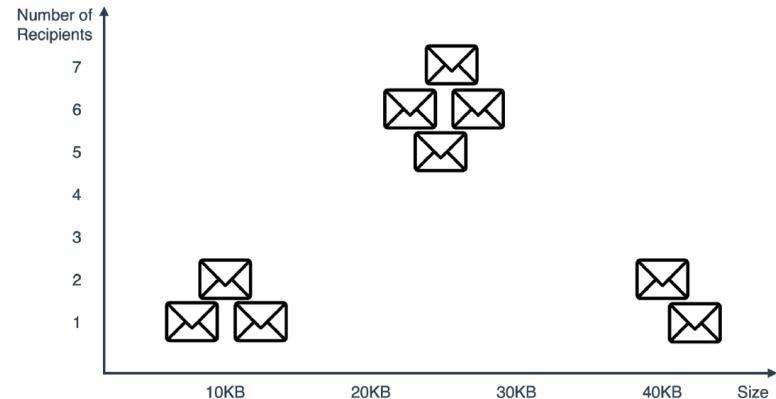
Clustering Algorithms

Clustering Algorithms

Email	Size	Recipients
1	8	1
2	12	1
3	43	1
4	10	2
5	40	2
6	25	5
7	23	6
8	28	6
9	26	7

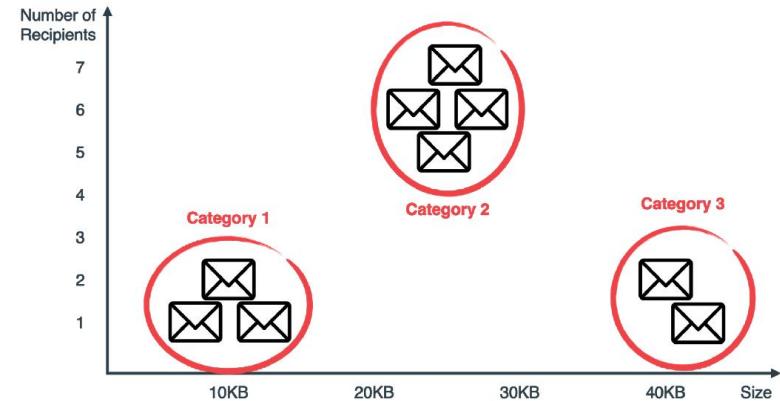
Clustering Algorithms

Email	Size	Recipients
1	8	1
2	12	1
3	43	1
4	10	2
5	40	2
6	25	5
7	23	6
8	28	6
9	26	7



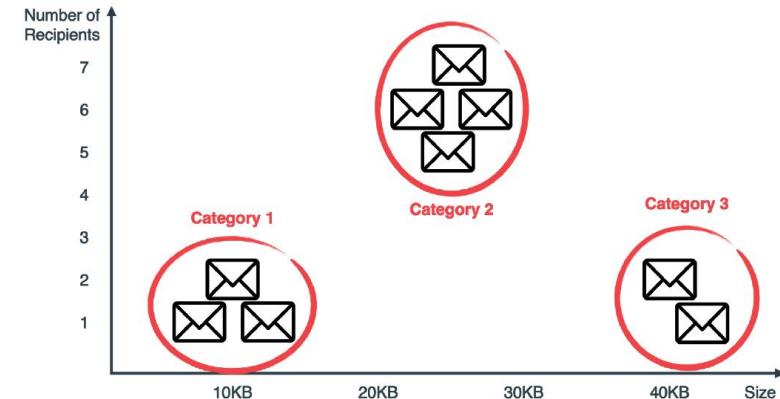
Clustering Algorithms

Email	Size	Recipients
1	8	1
2	12	1
3	43	1
4	10	2
5	40	2
6	25	5
7	23	6
8	28	6
9	26	7



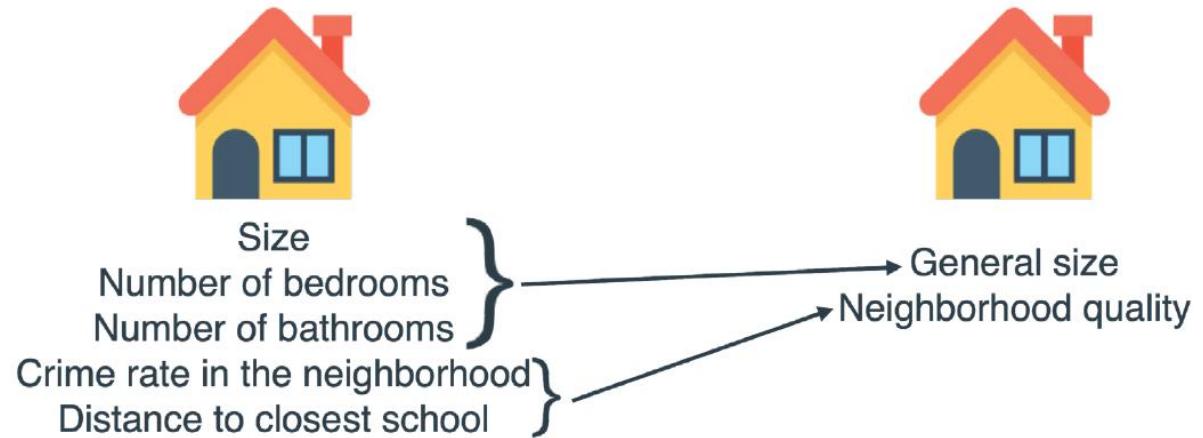
Clustering Algorithms

Email	Size	Recipients
1	8	1
2	12	1
3	43	1
4	10	2
5	40	2
6	25	5
7	23	6
8	28	6
9	26	7



Any other clustering problems?

Dimensionality Reduction Algorithms



Generative Algorithms



Let's check this link:

<https://generated.photos/face-generator>

Lecture Overview

About the course

What is machine learning

How do humans Learn?

Model and Features

How do machines learn?

Machine Learning Models Types

Data Types

Supervised Learning

Unsupervised learning

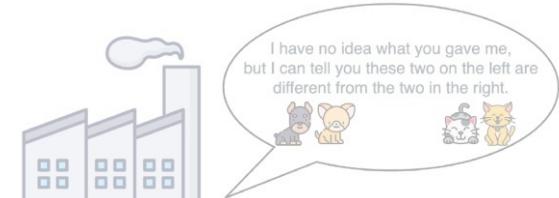
Reinforcement Learning

Machine Learning Models Types

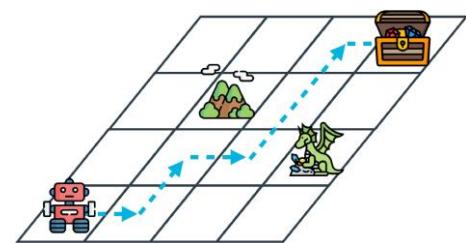
Supervised Learning



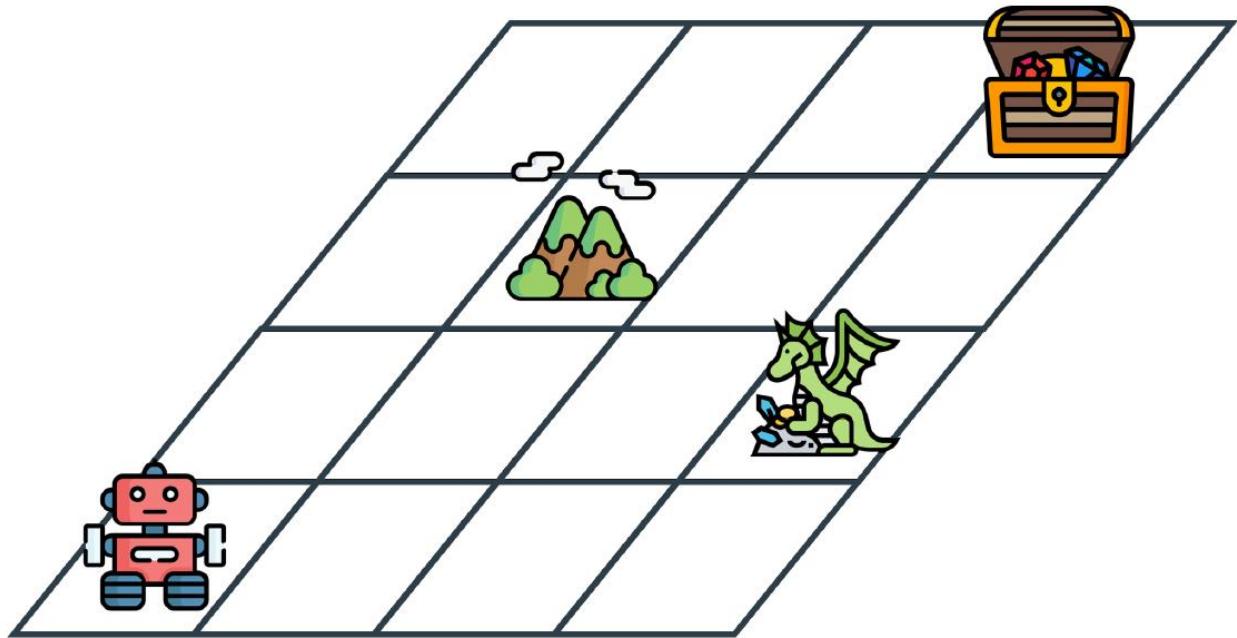
Unsupervised Learning



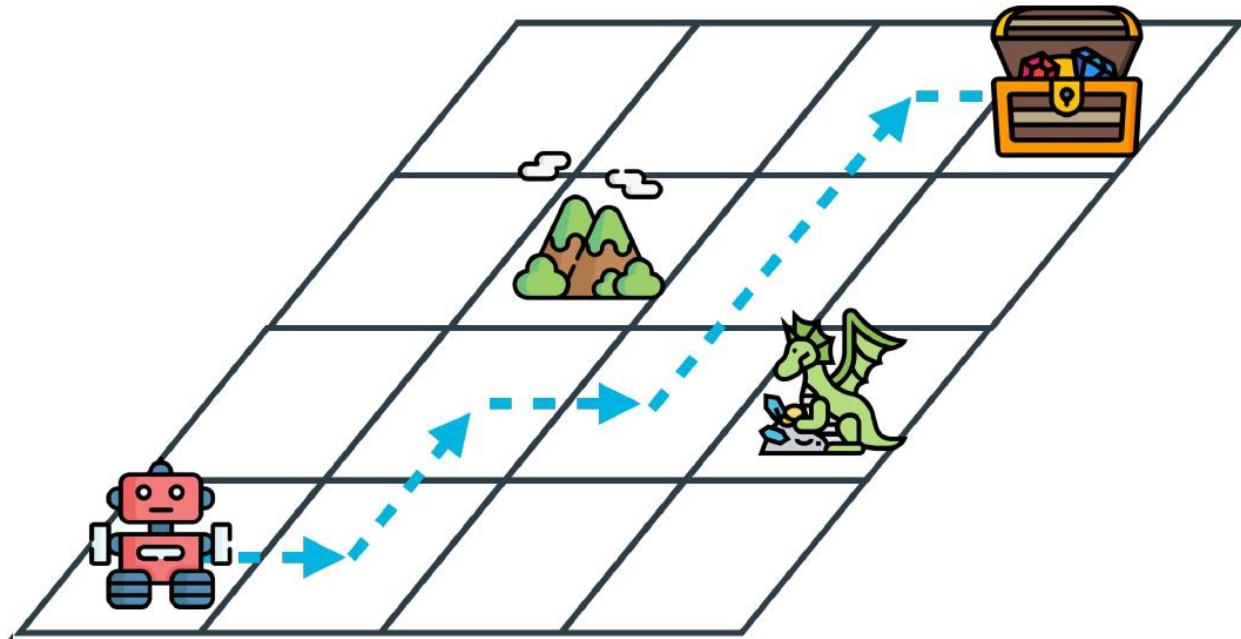
Reinforcement Learning



Reinforcement Learning



Reinforcement Learning



Reinforcement Learning

Example



Summary

- Machine learning like a melody
- Machine learning is a part of artificial intelligence
- Humans and Machines use Remember-Formulate-Predict framework to learn
- Data Types:
 - Labelled (Categorical)
 - Labelled (Numerical)
 - Unlabelled
- There is a difference between Data, Labels, Predictions, and Features
- machine learning models types:
 - Supervised learning
 - Classification Models
 - Regression Models
 - Unsupervised learning
 - Clustering Models
 - Dimensionality Reduction Models
 - Generative models
 - Reinforcement learning

Exercises

Exercise 1

For each of the following scenarios, state if it is an example of supervised or unsupervised learning.

Explain your answers. In cases of ambiguity, pick one and explain why you picked it.

- a) A recommendation system on a social network that recommends potential friends to a user.
- b) A system in a news site that divides the news into topics.
- c) The Google autocomplete feature for sentences.
- d) A recommendation system on an online retailer that recommends users what to buy based on their past history.
- e) A system in a credit card company that captures fraudulent transactions.

Exercises

Exercise 2

For each of the following applications of Machine Learning, would you use regression or classification

in order to solve it? Explain your answers. In cases of ambiguity, pick one and explain why you picked
it.

- a) An online store predicting how much money a user will spend on their site.
- b) Alexa decoding voice and turning it into text.
- c) Selling or buying stock from a particular company.
- d) YouTube recommending a video to a user.