Week 1 Video Transcript

okay So in uh in this video I'll I'll give the lecture for week one which is

0:09

um which is populated with all this content here Uh and uh from that I I

0:17

crafted some goals uh obviously we can change this uh and

0:23

this is a draft version of week one Uh again all the the figures images and

0:30

slides they are very drafty Um but I hope that with this we can kind of uh

0:36

better differentiate the se the sections of week one and what could be taken to on

0:45

online learning um kind of self-study learning or not

0:50

So I'm going to start the course I I will going to teach as if I am giving a class in a in a in a uh in a classroom

1:00

with students Um and uh this will generate kind of a script uh usually

1:07

this online courses you have the script uh you know very

1:13

um flowing very well uh and here in here because it's a lecture uh you will see

1:21

lots of a and and and stopping points and comments but this is an

1:27

underlying material for uh a more a more kind of movie script

1:34

after Okay So let's uh let's start the the course

1:41

I'm going to start the I'm going to start the the

1:48

uh the lecture now Okay So just one second

1:58

Okay Okay Welcome to the course In this module we will

2:06

explore artificial intelligence what it really means how it learns

2:12

and by doing that we will face several examples of how AI is being used across

2:20

industries uh to create value We know that AI is everywhere

2:28

uh powering recommendation systems on Netflix or YouTube optimizing logistics

2:33

for Amazon or even in our cell phones uh it's it's for sure everywhere and

2:41

it's it's becoming vital of any modern business So what is artificial

2:50

intelligence Artificial intelligence is about getting computers to

2:56

do things that uh require human intelligence So understanding language

3:04

is an example reasoning is another

3:09

example For examp uh navigating the physical world learning predicting the

3:17

future All of these are underlying tasks that requires human intelligence and

3:24

that when we set computers to do it we're talking about artificial intelligence

3:32

AI is being uh is seen increasingly everywhere and

3:39

for sure is uh said to be the next phase of digital

3:45

transformation If we look at the um at the

3:52

history in the late 1990s the internet uh was the digital technology that

3:59

transformed businesses Then cloud computing came in Then mobile computing

4:05

in the late 2000s with the internet of things And now AI is the digital

4:15

technology that has the potential to transform businesses So companies

4:22

that were slow to react to the emergency of this uh digital technologies

4:29

uh were left behind uh and therefore from experience we know that companies

4:36

now need to react uh to this new emergence of

4:45

AI So what are the implications for

4:54

businesses First of all the f the the f uh the first thing is to know that even if

5:02

you're not in an industry that is IT related

5:08

um this this digital transformation will affect your business because AI is

5:16

everywhere and it's a general technology that is being used in several different

5:22

domains So managers need to understand this technology its

5:29

application and uh make changes to

5:35

uh to embark on this uh digital trend

5:44

These changes can be uh different business models uh changing technology infrastructure

5:52

organizational process uh processes and the culture of the business uh as

6:00

well So to let's start diving deep into

6:07

AI So AI is a fancy name This was my dog Sorry AI is a fancy name again uh to say

6:16

that a computer is doing what a human would be doing right There are uh

6:23

there's a taxonomy of AI uh and usually experts agree in this that there are two

6:29

different uh big groups or two different approaches to AI The first approach is

6:35

expert systems and the second approach is learning systems

6:42

So expert systems are AI So computers are doing what humans would do but they

6:49

are hardcoded So their knowledge comes from a human typing in the

6:56

rules One example of this is uh diagnostic systems in healthcare where

7:03

doctors create a set of rules from their knowledge they hardcode

7:09

that in the computer and then the computer kinds of gives the diagnos uh

7:15

the diagnosis What the characteristics of expert systems is that it relies on this

7:23

hard-coded rules uh written by domain experts But this uh has

7:29

some um some uh challenges and limitations So it struggles it it

7:37

struggles with uh un unexpected situations or ambiguity when these rules

7:44

are um when these rules are overlapping for

7:50

example So it is inspired in in the human intelligence right of the of

7:56

reasoning with logic Uh but it's hardcoded

8:02

uh there are more modern examples of rule-based uh AI systems for example uh

8:08

credit score engines that banks use uh uses or ellegibility for insurance

8:15

claims So we have predefined rules that determine the outcomes and we will code

8:23

these rules Sometimes we need the computer because to to aggregate all

8:30

these rules with a human mind would be very difficult Uh so we're not saying

8:38

that these systems are not good They're just limited in some aspects

8:46

Uh so for example an AI chess player and

8:52

uh uh could could be an expert system that has all the rules of chess inserted

8:59

into it So if something changes in the

9:06

program uh in the in the environment in the game that is not hardcoded the AI

9:11

will have problem problems with it Now the other branch of artificial

9:17

intelligence is the machine learning branch or learning

9:23

systems So this branch enables computers to learn from data and improve their

9:30

performance on tasks over time without being explicitly programmed

9:37

for the task So you want the computer to

9:43

uh to perform a task and you don't you won't explicitly

9:49

tell the computer how to do it You will just make the computer learn from first

9:57

principles Um so this branch of artificial

10:02

intelligence can generalize to new tasks and solve things that we don't know how

10:08

to solve We don't know the rules Uh and it was loosely inspired uh and and

10:16

validated by neuroscience Okay So uh we can we can give some examples

10:24

of machine learning uh mo of machine learning systems or learning systems

10:30

uh when a machine wants to predict whether a customer will purchase a

10:36

product right It can it can analyze the browsing behavior the time spent on the

10:42

product page and past transactions to estimate the likelihood

10:49

of of uh of purch of p of the client purchasing a product

10:56

So there is no rule there is no kind of specific hard-coded model that uses

11:03

browsing behavior uh time spent in in web pages and the transactions history

11:08

to estimate the the likelihood of conversions We actually need data to

11:16

learn So um there are several other examples

11:21

that we will come across but again we are enabling computers to learn from the

11:30

data uh and perform the task So in this case the computer will

11:37

be giving several and several and several examples of customers and it will extract from the data a knowledge

11:47

that kind of combines all of this uh features of browsing behavior uh time

11:55

spent on product pages and past transactions to estimate the likelihood

12:00

of a conversion So this will be the main focus of this

12:06

course So this course is mainly focused in machine learning Um and also uh in technical AI

12:14

literature sometimes you will see that expert systems are called symbolic AI

12:19

and learning systems are called machine learning or um statistic statistical

12:27

learning Okay So because we're focusing on AI uh on learning systems uh AI

12:35

learning systems or AI machine learning systems we need to understand

12:41

data I'm sorry data So let's talk a little bit more about

12:48

data So data are examples from which the AI

12:54

learns A data set is a collection of examples So here you can see this is a

13:01

collection of examples of a customer purchase history So you have the order

13:08

ID what the person purchased what category it was on uh the code of that

13:16

particular product the name of the seller and the purchase price So this is a data set It's a

13:23

collection of samples or examples of things that

13:30

happened and that the AI can learn from This is a numerical uh example of

13:38

data but data can be images uh it could be uh text any form of input

13:47

that can be processed by computers Okay Um another data another example

13:56

would be uh weather data So which includes temperature humidity and wind

14:04

Uh and uh it is used to predict the the forecast for tomorrow for example

14:11

So we are uh so as you can see here

14:20

uh the what the the in the examples or

14:25

data points they are called we can call it attribute uh we can call it instances So

14:34

instances so the first line the first row is one is one purchase u

14:41

example and it's an instance this is one one

14:46

data point of that purchase So every row in a data set is a

14:55

instance or an example or a a sample

15:02

point The the information that comes in the data

15:07

set in the columns are called attributes

15:13

or features Attributes or features

15:20

So the the category for example of the p uh of the product purchased is an

15:27

attribute or a feature of this particular data

15:35

set So different instances will have different values for each one of these

15:43

attributes Another example would be for uh for example the uh weather data

15:50

set where you know if it's cloudy if it's

15:57

windy the humidity uh or the the the air

16:03

pressure and this is for day one day two

16:10

day three So each day is an instance and the values of the

16:17

characteristics or attributes or features are appear in the columns for

16:24

each row So you can see here in this image

16:30

some of the attributes are numerical some of the attributes are categorical

16:36

or text For example um the the category

16:42

in on of the product uh uh purchased it could be text it could be

16:49

numbers it can be images uh data needs to be processed by

16:56

computers So in the end of the day this data will be transformed

17:02

into into something uh uh in in the form of an input that can be processed by a

17:10

computer Here is a is an example of an image An image can also be part of a

17:16

data set Uh for example this is the instance number one of a data set So

17:23

it's an image and an image is is composed of several um uh little squares

17:30

You can think about it as a a a a grid of very small um squares And each of

17:38

these squares have different intensities of red of green and then and of blue So

17:44

for example this first uh this image here is the first row of a data set

17:51

where it has several columns The the attribute R1 is the value or the

17:58

intensity of red of pixel one intensity of green of pixel one and intensity of

18:05

uh blue of pixel one then pixel 2 then pixel 3 and so on So you will have a a

18:12

data set with several columns and you could have a 100 images and you would

18:18

have 100 rows Okay Now uh data

18:26

sets are again the these collection of examples where machine learning will

18:32

will learn from Okay But not all data is useful right So you have to have high

18:39

quality well-prepared data sets so that your machine learning model can learn

18:46

Okay Just uh as an example a ride sharing app might collect data such as

18:53

pickup location drop off location type of the day uh time of the day driver

18:58

rating the fair uh you uh the fair used But if if your data is inconsistent if

19:06

you have several missing key variables um the AI won't learn right how

19:15

to for example uh predict the closest or the the best option of fair of the the

19:24

closest driver and the best option of of fair for you

19:30

Okay So continuing in uh uh talking about data the we are overwhelmed by data and

19:40

uh the true value of data lies in what we can learn from it Right So uh the

19:49

goal is to take raw data and transform it into something

19:55

more useful uh information for example uh or

20:01

predictions predictions about what might happen next uh and that can and and that

20:07

can be used in the real world So more useful information you could um you

20:13

could have um uh some kind of important information

20:20

informing your decision making making predictions Predictions about something

20:26

that you want uh to uh to forecast for example or predicting something that's

20:33

going to happen next based on the information stored

20:41

uh uh sitting behind this raw data

20:47

So um the this a large amount of

20:52

data actually uh makes us uh kind of

20:58

find this um important informations this transformations and uh of of raw data

21:06

into information predictions uh through a iterative and exploratory

21:12

process So we have the data we start explore exploring this data We find uh

21:19

important correlations between attributes for example

21:24

uh or attributes that depends on each other

21:30

and this will actually uh will serve for an actionable insight

21:38

you can have an actionable insight and this will um change the course of your

21:46

decisions for example in businesses So uh let me give you an

21:53

example So you you go to a supermarket checkout Uh you have a loyalty card and

22:00

they'll give you um uh uh they'll give you some kind of

22:06

coupons and because they have your uh loyalty um they have your name address

22:13

and they have access to all sorts of demographic data about you and people

22:18

like you right So these coupons are individually

22:24

crafted uh for you and you will get

22:30

bargains and um they will sell more

22:35

So actually they know what you've bought They know your demographic information

22:42

your profile and they'll they'll take this data and they'll analyze it They'll

22:49

include uh this this data from maybe

22:54

thousands of millions of people just like you They'll do experiments to find

23:00

um given what you've bought today and your profile

23:05

uh what should be the next the next coupon to send you by email So it's kind

23:12

of a a mechanism for individual prices and it's it's very beneficial You get

23:19

the bargain and the super market sells more

23:25

So um this this process of getting the

23:30

data exploring the data finding important um uh important information or

23:38

useful information in the data and then do an actionable having an actionable

23:44

insight where you can inform decision making is part of what we call it's part

23:51

of a bigger um area uh of of a bigger

23:57

field called data science So data science is uh not only

24:05

transforming this raw data into um um into insights and into information

24:14

but it's also the um the whole process

24:19

So the process of storing the data of uh

24:24

cleaning the uh of acquiring the data of cleaning the data of storing the data of

24:30

then analyzing the data uh and and trying to to figure out um trans the uh

24:38

and trying to figure out um good information and insights from the data

24:45

It also encompasses the the the the uh building for example

24:53

predictions prediction models of what what can happen next Um so I'll give you

25:02

an example Imagine a telecom company Uh they they have they are they want to um

25:10

they want to try to to um learn

25:16

to oh god they they want to try to understand

25:22

um when customers cancel their um their plans

25:30

and and they want also to recommend given this insight that they are kind of

25:37

uh wanting and they're kind of looking for recommend retention strategies So

25:42

the data science process will collect clean store the data It will use what we

25:49

call data analytics and data mining to find this uh kind of hidden uh

25:57

information and correlations between all the attributes and features finding the

26:05

most important ones right So maybe the um the the clients that had um a

26:13

decrease in their uh in their calls the number of calls they will be

26:20

more prone uh to cancel their plans So

26:25

you have a bunch of data sets raw data and you with data analytics and data

26:30

mining you will kind of uh extract important information important

26:38

correlations important features of the data um that regards the cancellation of

26:46

plans and then you can build machine learning models AI models to predict

26:54

uh the the what's the probability of cancellations given the number of um

27:02

phone calls for example and also you can get this and transform this into a

27:09

recommendation and uh escalate this to the business management

27:15

uh to kind of change policies and strategies for

27:20

retention So in all this process uh you can generate reports and data

27:26

visualization These are very important skills for this um this

27:33

exploration uh uh of of the data science

27:38

um field inside businesses

27:44

Um it's it's so data mining just to uh because you come you will come across

27:50

these terms uh in uh uh in AI sometimes you hear about

27:56

data mining sometimes you hear about data analytics uh and data science

28:02

um think about u uh data mining and

28:08

analytics as analyzing the data the the current

28:14

data uh trying to understand why the data is that way um through statistics

28:23

through histograms um sometimes through machine learning models

28:30

Data mining is also uh the this process of finding important features important

28:39

correlations important information that um that is related to whatever your

28:47

aspect you're analyzing Um so in the case of our telecom company

28:54

uh the data analytics would be opening this data understanding each column of

29:01

the data each attribute working on um cleaning this data doing

29:08

um histograms and understanding why it's that way getting it prepared to then

29:15

mine this data in terms of finding uh interesting features that relates to

29:22

cancellations and then uh use machine learning and then after that you can use

29:28

machine learning models AI models to to do even more right predictions and and

29:36

recommendations so all of this cycle is the data science it's kind of

29:41

encompasses data data mining and data analytics I also I always think that

29:47

it's it's very good to make a um it's very good to make a uh uh

29:54

parallel with a chef that is working in a kitchen So the analytics and data

30:02

mining is like analyzing the fridge what you have what ingredients you have and

30:08

why you have that And uh the data mining is kind of

30:13

picking the the ingredients identifying the most promising elements

30:21

right So for example one uh tomatoes uh goes very well with

30:27

basil And so you would figure out how they might work together and the best

30:33

ones And the data science is the full culinary journey It covers everything

30:41

from choosing where to get the ingredients to cook uh uh and plating

30:47

the dish Um it it even u uh

30:53

incompasses predictions on how well it suits the dinner's preferences

30:58

So data mining and analytics equips you with the raw materials and insights and

31:05

data science is all this this whole process of bringing together uh the

31:11

experience of of leveraging data Now these distinctions are are not always

31:19

rigid right So they're they're used really loosely They they they um overlap

31:26

right So analytics overlap with with uh with data mining concepts and and they

31:32

use the same tools in data sciences Uh they data scientists uh often rely on

31:38

anal analytics um to validate their models Uh they use similar tools and

31:46

they have similar skills Um so there are significant a significant overlap but

31:53

it's important to kind of make uh this uh this uh distinctions because you will

31:59

come across this uh terminologies So

32:07

uh here is a a slide that I'm not sure if I'm I'm going to uh insert It's about

32:14

uh types of analytics I'm not sure if it it's good or not So in in the next uh

32:20

kind of interaction of the script I will decide if this will be in the u in the

32:26

script or not just have to think a little bit more about it

32:34

So so um going going for moving forward

32:40

um the the the value of data is learning from it and and it okay learning is fine

32:48

but you actually have to act on them So for for for businesses um it's very

32:54

important that managers know how to interpret findings and ask the right

33:00

questions right so that they can apply the insights and the recommendations and

33:06

and understand the predictions of all this data science process to real

33:12

business problems So it is very important to managers to to be

33:18

um to be digitally savvy w with data literacy and also have the domain

33:25

expertise of their uh of their industry or of their um niche

33:32

Um so not not every uh pattern in data for example is meaningful or uh

33:40

correlations or uh uh kind of uh insights Um so we have

33:49

to to uh analyze this uh managers have to understand

33:56

uh the the power and the limitations of

34:01

data analysis Sometimes data analysis comes with some um of uh with some uh

34:10

uh correlations um and relationships between attributes

34:17

uh that are totally nonsense right And it might it might be just a sporious

34:24

thing So and sometimes data the data scientists that are working on it they

34:30

don't have the expertise on the domain to kind of uh trash that out and say oh

34:37

this is not good So managers have to understand uh this process of data

34:44

science to um to understand the limitations of data

34:50

and then um have this scrutiny to to select

34:58

what's coming from the data analytics Um often the companies have uh data

35:05

scientists uh a a chief data scientist or a data architect and they uh this

35:12

professionals they they manage the the data flow they will they will design the infrastructure and we will talk about

35:19

infrastructure in a uh uh just in a in a bit Um and they will also make choices

35:28

um make choices of um technologies they're they're going to

35:34

use for managing their data So managing their data meaning uh capturing storing

35:43

organizing the data We will talk about this as well And also they will choose

35:49

analytical tools um platforms frameworks for example if

35:56

they if they're if they're going to use uh third-party softwares with uh with

36:01

machine learning and statistical p analysis power or if they're going to

36:08

use for example Python libraries uh to help turn this uh complex data

36:15

sets into insights and then decision decisions So

36:22

um again uh we we want this insights to turn into action to create decisions

36:30

inside businesses that uh for sure and this this figure is overloaded uh I'm

36:38

not sure what this means matrix for example but but um for sure the data lit

36:43

literacy here uh and the commitment with with this digital um uh technology uh

36:52

has to has to exist to kind of um to kind of fuel the data uh driven

36:59

decisions a little bit more about what what

37:05

happened in in the world in this past uh uh years So um what happened is now we

37:15

have what we call big data right So we

37:20

uh in the past years data was generated um by having everything going digital

37:27

Our smartphones our refrigerator um

37:32

people and things that are connected uh uh all over the web and every time that

37:41

we connect something um we generate a piece of data Okay

37:47

So every time we check out an item at the supermarket every t time we swipe

37:54

our credit card every time we send an email uh or even if you uh even uh you

38:00

can consider the keystrokes on your keyboard every time we make a phone call

38:05

um or or walk through a sec a past a security camera all of this will

38:11

generate a little bit of data Okay So and because all of these devices all of

38:18

these equipments are connected and are tracking and logging all of this

38:24

information Now in the past years in the past years uh thanks to all

38:31

of this uh devices wearables social media and internet of things we have a

38:39

capac we have a uh amount of data a volume of data that is very big

38:47

And this kind of went uh in parallel or together with the computing capacity So

38:55

the the the data increased and the capacity to store data has also

39:01

increased So that's why we're seeing this boom in machine learning and that's

39:07

why we are in the golden era of AI where AI is trending because we can have data

39:15

to learn this uh models Okay So

39:22

um now before we had some data some companies stored a little bit of data

39:29

here and there from their operations but after all this uh digital technologies

39:36

um we have actually what we call big data So we have data that has volume

39:42

that has um a a rich v variety of data It could

39:49

be uh structured data unstructured data multimedia it could be videos it could

39:56

be text um and you you we're talking about

40:02

terabytes or or pabytes of of data Um also this this data

40:12

um because of our comput computing capacity can uh we can exchange or

40:18

stream this datas very quickly and um because of the amount of data um there's

40:26

lots of inconsistencies missing values ambiguity ambig ambiguities

40:33

So we have everything now in a scale that we didn't have before Okay Uh don't

40:41

worry uh we're not this is not the focus of the course um of big data per se If

40:50

you want to know more about big data and its challenges in business infrastructure for example

40:57

or or its computational challenges you can check out the extra

41:03

resources and and for sure this data is uh is is more uh rich in terms of

41:10

information um but it's also more challenging to

41:17

um to manage Okay So speaking of managing all this data um we continue

41:25

talking about data First of all um in in what means to have all this data for

41:33

your machine learning models uh for your AI models to work Well it means that you

41:39

need to manage you have to collect the data So you have to collect you have to

41:47

clean the data you have to store you have then to extract knowledge

41:53

of the data and and and organize the data and

41:59

make the data data pertinent for extracting insights of that data Okay So

42:06

the initial step to accomplish this in a in a um in a a business is through what

42:14

we call database management systems So database management systems We also call

42:19

this only databases It it's it's common to just call this as databases So um

42:26

databases are structured collections of data So there the the data set sits

42:31

inside this databases where you can find this information and retrieval of

42:40

information is possible Companies use this information um uh to manage their

42:48

their companies use several tools to manage their databases such as SQL uh or

42:53

big data platforms uh like snowflake or da d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d d dab bricks or something like that

43:00

Um we the the the core aspect of this course is not on uh data data

43:08

infrastructure um or information systems infrastructure for businesses but if you

43:14

want to learn more we will leave uh a extra resource for you

43:20

So this databases um they they have different characteristics Uh they're

43:27

operation databases So for businesses so this operation

43:33

databases they're usually stored locally for each operation of the company of the

43:39

company and other more comprehensive databases

43:45

uh where you have um uh information from

43:50

several operations and more historical data and

43:57

and and and and aggregation of data from all your processes They are usually uh

44:06

stored in big data management systems in the cloud

44:11

and in what we call uh data warehouses So here is uh

44:19

warehouses warehouses uh for this uh this

44:25

databases So you have I will give you an example

44:30

Think about Airbnb Every booking review photo location

44:37

uh uh every information is part of their data right

44:42

So um and they will use this data For example

44:49

um if a customer logs in and see a past past bookings a operation database is

44:57

accessed right So it's it's the operations of uh recent bookings It's

45:03

part of of a faster um with faster access database uh that

45:11

usually is stored for that particular operation Now

45:17

Airbnb they have a centralized for sure they have a a a data warehouse where

45:23

they have a bunch of uh aggregate data on all its operation all historical

45:30

informations and this data can be used for the purpose of extracting

45:38

knowledge from this this history from this past informations

45:45

So these the uh for example if um Airbnb

45:50

wants to recommend um uh for example uh

45:56

listing and adjusts prices It will for sure create a machine

46:03

learning model for example that uses historical data not only from past

46:09

bookings but from uh from uh for example u um characteristics or profile of the

46:17

user to make this recommendations So it will it will need this kind of

46:25

aggregate um pieces of information So the ex so let's talk now

46:35

about machine learning models Let's dive dive into

46:40

uh AI right so the machine learning models that we're talking about that we

46:46

that is the um the theme of this course and now that we know um that

46:54

machine learning models are used inside this data science cycle this data

47:00

science field um where we're going to leverage data to have insights and to

47:07

create tools uh that accomplishes humanlike tasks Um we

47:16

need to kind of understand it a little bit more what it is and how it works

47:26

Machine learning models are computational models um mathematical

47:32

models that can be uh that that that sits inside computers that ultimately

47:39

performs a task We saw that uh AI perform uh human uh intelligence tasks

47:47

for so for example um classifying um something

47:56

predicting something reasoning So the the task the underlying task of a model

48:03

depends on what you need the model to do But a model will receive inputs and it will do

48:14

this task this prediction classification or reasoning and will

48:20

output an answer for you Okay Um and

48:25

because we are talking about machine learning models these models will learn

48:31

from data sets they will learn how to do this with data sets Okay So again we

48:40

will uh the the model will learn to recognize for example a pattern

48:48

um and and output and uh a a underlying

48:54

task It will uh learn how to predict something it will and then we'll output

49:01

this something Okay so in the end of the day and here

49:07

are some examples I could present several photos

49:12

of dogs and cats and here we could have uh a figure with that dogs and cats here

49:19

um I would present a figure with dogs and cats to a machine that would

49:25

classify uh if that particular image is a dog or

49:30

is a cat So this this human related task this human um intelligence task that

49:38

we're performing is what we call classification We could also uh be

49:45

wanting to predict the stock market So

49:50

you will have inputs as the actual values of the stock market and the

49:55

outputs will be the forecast of the stock markets or the inputs could be

50:03

the the characteristics of the day um

50:09

pressure uh air pressure humidity uh if it's windy or not and so on And you will

50:17

learn to predict uh to forecast the weather tomorrow Okay So that's what a machine

50:24

learning model do It learns to do this mapping from input to output So if we if

50:33

we understand the output that uh this output depends on the input So y is

50:41

actually a function of f The machine learning model is learning without any

50:48

hard coding how to approximate this mapping from the input to the output

50:56

So it's actually learning how to map the u for

51:03

example pixels of images and its several features into

51:12

uh a cat right so the the uh the output is cat or dog for example so there's a

51:19

mapping that we are learning okay now for a uh we could we could use

51:29

the another example is an email spam detector Okay So you could feed in uh

51:35

the uh features of an email uh for example number of URLs and if that

51:43

person um is in your contact list or not and your model will uh reason if that

51:51

input is a spam or a not spam Okay So in

51:56

the end of the day your model is learning this mapping is it's it's kind

52:03

of um learning how to transform from your outputs to a

52:10

uh from your inputs to the output that you want for this PAM email example for

52:17

uh the the learning of this uh model will be made from data from data sets

52:27

that you already have So let's say that you have several example emails of um so

52:35

email one 2 3 and and so on and here's the number of URLs if the person is your

52:41

in your contact or not um in your contact list and other features So what

52:48

we're going to do is we're going to use this this this information this kind of uh knowledge

52:56

that is sitting behind all of this sitting behind this data set and and

53:02

through learning we will uh approximate this functions so that when a new email

53:10

comes in and you give the number of URLs of this particular new email and if it

53:16

if that is in your contacts or not and all the other features this machine will

53:22

reason for you if this email is a spam or not a spam

53:29

Okay So in machine learning a model will will approximate this function It will

53:37

identify this relationships between input and output um to make decisions to

53:44

make predictions to make classifications um on new information and the way it

53:51

will learn will be on information that in past information that sits in data

54:00

sets Okay So let's continue studying machine learning We saw that machine

54:06

learning models are models mathematical representations or computer that sits in

54:13

computers So computer models that given some inputs will accomplish a

54:21

task The way the machine learning learns the type of learning under the AI under

54:28

the machine learning model is actually very important

54:35

There are four different types of learning and they are very related to

54:43

the task you want to to accomplish So for example supervised

54:49

learning is one of the first method meth methods we have actually supervised

54:55

unsupervised self-supervised and reinforcement learning Some experts leave

55:00

self-supervised out of this list and kind of um uh place them uh just just

55:08

beneath like inside the group of unsupervised learning But I like to

55:13

highlight that because this is the the learning method that that is um

55:19

um used in most large language models uh the models that are

55:27

um that are used in in chat GPT for

55:32

example So supervised learning is when we train a model we we make it learn

55:38

using data but the data is labeled So you have input output what we call input

55:47

output pairs So we are telling the machine what the correct answer answers are This is

55:55

like teaching a student with with the answer keys And unsupervised learning is the

56:03

opposite You don't have any label data You're not telling the machine what is

56:09

the correct answer The model has to figure out patterns or clusters in its

56:15

own Self-supervised learning is in between both So the model creates its own labels for the

56:22

data and reinforcement learning is when an AI system learns by interacting with

56:28

the environment So um the super the the the

56:34

types of machine learning will they're very related to the task you want to accomplish

56:41

So let's dive into um each one of these and uh learn that

56:49

learn how to identify this tasks For example here classation and regression

56:55

are tasks that you can accomplish using supervised learning Clustering anomaly

57:00

detection you can accomplish using unsupervised learning classification regression and natural

57:07

language processing tasks Um which is which is summarization

57:14

translation you can accomplish using self-supervised learning And if you want to optimize

57:24

um you can accomplish optimization using uh models that learn via reinforcement

57:32

learning So let's start talking about supervised learning Okay So

57:40

supervised learning So supervised learning is a

57:47

foundational approach uh where models learn from what we call labeled examples

57:54

So it requires labeled data uh what we call input and output data

58:02

Okay So uh the the data comes with a label

58:10

meaning the true target or outcome is

58:16

provided for the machine right as anam

58:22

as examples So let's say here you want to train your

58:28

your machine learning to predict um whether photographs of uh Alan Touring

58:37

uh whether photographs are are of Alan Touring or not So the way you do it is

58:43

you collect several several data several photos of Alan Touring and you also

58:51

provide a label saying yes this first photo is Alan Turing the second photo is

58:57

also Alan Touring and the third photo is also Alan Touring And the

59:03

task is um for example in this case is the

59:09

classification task where giving a new instance a new unseen data a new photo

59:18

your machine will take that input which is the photo and then it will predict in

59:26

this in this um case classify if that photo is of Alan Touring or

59:34

not So in the end of the day we use several data to train the

59:43

same thing as the learning process So you train your data set your sorry your

59:51

model and this model is deployed so that you can use and test this model in

59:59

unseen data So let me show you a very uh

1:00:06

curious video of a real world application of a classification task So

1:00:15

let me open this website You can see a video and this is a video of um of a a

1:00:24

Tesla auto driving system and you can see that what it's seeing is a truck

1:00:31

What is happening is there's a truck on in front of the car uh that holds uh

1:00:37

transporting um traffic lights and the the Tesla auto

1:00:45

driving system have learned through several examples and

1:00:50

photographs of of um of traffic lights

1:00:56

to identify that as a traffic light And so this is crazy because what you see in

1:01:02

the Tesla monitor is that like there are several traffic lights coming towards

1:01:07

you but it's because it learned to classify that particular thing as

1:01:16

traffic light So somebody showed several traffic lights and said look this is a

1:01:21

traffic light this is a traffic light and insisted uh showing lots of traffic

1:01:27

lights so that your model now when it sees a traffic light will output a

1:01:34

traffic light for you So it's it's doing what it's supposed to do um uh in this

1:01:40

real world system Another application would be for example recognizing tumors in X-ray So

1:01:48

you you you the input for your model is now X-ray images and the output of your

1:01:56

model is uh cancer or not cancer

1:02:03

So and you can also just predict a simple um thing as if it's going to rain

1:02:12

or not rain given um this um u inputs for example humidity

1:02:21

pressure and uh uh and and the and the and the characteristics of

1:02:28

the day if it's windy or not for example So let's um let's go a little

1:02:34

bit further in this concepts So when I say it requires training data and output

1:02:41

data I'm telling you that the the data set used to train So the data set used

1:02:50

for learning has to have columns For example

1:02:56

here we have days where humidity were was 1.2 For example this first row

1:03:03

humidi humidity was 1 point uh 1.2 2 and the pressure the air pressure was 1 um

1:03:11

sorry one and but it also has to have a

1:03:17

column where you're giving the label you're giving the target or you're

1:03:22

giving the answer you're kind of teaching the machine learning you're

1:03:28

showing a data where the where the machine learning can know the correct answer so for this day here day one

1:03:38

um given this values of humidity and pressure it rained This is a historical

1:03:45

data set So this is past data Day two the humidity was the same

1:03:52

but the pressure changed and there was no rain and so on So you have several data

1:03:59

data points okay in your data set So what are you going to do is you are

1:04:08

going to um

1:04:15

to train the the model to predict if

1:04:24

the if the if within this input so your model will receive an input's value of

1:04:31

humidity and pressure and it will label rain or not rain

1:04:39

Okay And this is a classification task So the prediction when we're talking

1:04:44

about classification task it means that the prediction is

1:04:51

discrete right Like it's it's a class It's really a class It could be one or

1:04:57

zero It could be rain no rain It could be um

1:05:03

um if you're trying to predict the the profile of a consumer consumer A B or C

1:05:10

So there are distinct classes or discrete classes So this this is what we

1:05:16

call a classification a classification uh task

1:05:22

Now the the goal as we said in in earlier uh this this week is that the

1:05:30

model uh it it needs to work in unseen data So if a new day comes in you will

1:05:38

deploy the the um your model So you don't have the

1:05:45

answer You don't have right um the answer It's it's it's it's it's not

1:05:52

something that went you don't you didn't store that in your database This is actually happening now So you're you're

1:05:58

testing or deploying your model in this in this new instance and that's where I

1:06:05

want it to to to output rain or no rain

1:06:11

So we have part we have data that needs to be used for

1:06:16

training Uh so we're teaching the model and some the unseen data will be uh used

1:06:26

to deploy the model Okay And and we will talk about training and testing uh with

1:06:32

a little bit more details um soon

1:06:37

Another example is for example uh if you want to predict the marital status of a

1:06:44

person giving the age and the income So the ML algorithm learns to map right So

1:06:52

from the features from from the input the or what we call um

1:06:59

attributes um so oops so from the attributes attributes

1:07:06

um will it will learn to predict this target So when a new person come comes

1:07:13

in right so an unseen person comes in that was not used to train my model So

1:07:20

for example Louisa my age and my income

1:07:25

and if I am married or not So what I want is my model to exactly output this

1:07:34

Okay Now um this this kinds of data sets they

1:07:42

need to have the label and how can we get this data sets

1:07:47

with the label Well you you have to have this in information for example here you have

1:07:54

personal information that you can be retrieving from a bank data set or

1:07:59

whatever Um but if you're talking about a model a

1:08:06

machine learning classification model for images

1:08:13

um you will need of for example images of a dogs of of dogs and cats You will

1:08:19

have to have a a human labeling that photos

1:08:26

right So um you will have to make think about a spreadsheet where you

1:08:34

have all the photos from Alan Touring Somebody will have to check if that photos are from really from Alan Touring

1:08:41

and will create a column that is uh the kind of the output or the target of Alan

1:08:48

Touring or not Alan Touring Alan Touring or not Alan Touring So this is called

1:08:54

labeling and it's a process that it's difficult it's expensive Um obviously

1:09:01

for data sets like this one this comes for granted but for other data sets labeling is a very um tedious and

1:09:08

expensive process and it's actually a major bottleneck uh in uh um um in supervised

1:09:18

learning Um so um uh what is this I remember So

1:09:27

um because of that um you can imagine that

1:09:34

uh how how smart Google is for example or or other uh big tech companies where

1:09:41

they ask you to check it to verify that you're a human but checking selecting

1:09:47

images of buses for example So you're labeling for free um this kind uh data

1:09:55

for for Google for example Okay So let's think about

1:10:03

um what a model a classification of a

1:10:09

model is Okay So oops

1:10:14

sorry Let me go back So a classification model for the

1:10:23

rain no rain example is this giving if if we were to plot that table

1:10:30

right So this table here let me go back really quick of humidity pressure and

1:10:36

the label If we were to plot this points So for example data point one right is

1:10:45

1.2 and one and it's rain So it's let's say that it's it's here So this is data

1:10:52

0.1 It has a value uh 1 1.2 of humidity and a value of um so

1:11:02

it has a value of 1.2 two of humidity one of pressure So it lies here and it

1:11:10

rains So in this in this um uh picture I'm depicting rain as

1:11:20

um red points and no rain as blue points

1:11:26

So what what we want is a

1:11:31

classification is a model will draw a line It will so you want a

1:11:37

classification task So in this case this line is is a is my model because this

1:11:44

this line here is actually capable of separate this data points Okay So a

1:11:52

simple um a simple line in this case is a classification model

1:12:00

So after I learned how to separate after I kind of so the learning process is how

1:12:08

to position this this line in this

1:12:15

um in this instance space right or in the input space So this is the space of

1:12:22

inputs and the model is positioned here where it learned

1:12:30

uh to separate So um the the goal is to when a new data

1:12:38

point comes in for example this is the new the unseen data

1:12:44

point the the model will be able you will kind of ask the model okay so is

1:12:51

this unseen data point rain or no rain so it will get the

1:12:56

coordinates and because it's just a line it will be sitting underneath the line and the and the model will be able to

1:13:04

say okay underneath the line is no rain it's blue it's no rain so this

1:13:12

classificator here uh this sorry classifier here oh I'm killing the

1:13:17

English language uh is just um what is above the line is rain what is below the

1:13:25

line is uh no rain so in the end of the day this classifier will and output and

1:13:33

reason and predict your class The the the the training mechanism

1:13:41

of h how do we kind of learn this line We will talk and discuss it later but

1:13:48

just so that you can visualize this Um this is a process that usually starts

1:13:56

random So you start with a random classifier and as it learns from the

1:14:01

data that is being made of that is made available to uh the training process um

1:14:09

this this classifier kind of gets better and better and better in in predicting

1:14:16

in in separating class A from B or rain

1:14:21

and no rain So the this is the learning this is uh representing the

1:14:28

learning mechanism and how the learning mechanism is actually shaping the the

1:14:35

classification the the classifier and obviously that several

1:14:41

classifiers would work right um you can see classifier number one number two and

1:14:47

number three the three of them would would correctly classify the training

1:14:54

instances um and they would um also be fairly

1:15:02

reasonable in classifying a new data point So let's say that the new data point falls here um all three of them

1:15:10

would classify A If it falls here all three of them would classify B obviously

1:15:16

that um some of them will if if a data point for example falls here um

1:15:24

classifier number one would make a mistake because everything that's uh below the line would be B um but in in

1:15:34

uh in other on conversely if a

1:15:40

point B falls here um sorry a point A falls here uh the the third one would

1:15:48

make a mistake So and you would compare this classifiers and choose a better a better

1:15:54

one We will discuss several of these mechanisms next Okay So we saw the classification

1:16:02

task Now now uh just as a reminder that um a

1:16:10

a line in this case for uh two-dimensional problems and when I say

1:16:16

dimension I I am talking about the number of attributes So in this case we

1:16:21

have attribute H and attribute P So we can put in the in the graph um in a 2D

1:16:28

graph a line is a proper uh classifier but there are classifiers that are way

1:16:35

more complex right so you could have uh uh kind of arbitrary shapes that that

1:16:43

are way more complicated than this line so this example is a very uh basic

1:16:49

one let's talk about other predictive class a task So it's a prediction that

1:16:56

we're doing So it it falls under the umbrella of supervised learning Um but

1:17:03

it's called regression Okay So it also requires labeled data So

1:17:12

the input and the output data And um it it's it actually

1:17:21

um the model will learn how to um to

1:17:28

predict not a class not a discrete um not a discrete

1:17:36

thing or or group but a number So in

1:17:41

this case when we're talking about regression tasks we're talking about

1:17:47

predicting numerical values uh um

1:17:53

um sorry uh continuous values So numerical

1:17:59

continuous values it's not discrete It's not different classes Um it it could be

1:18:06

whatever number it is here It's not a predetermined class A B or C or one or

1:18:13

zero It could have any number any numerical continuous value So let's say

1:18:18

that we want to predict a um uh we want to map um the

1:18:29

um for a company We want to um we want to answer a question of how money spent

1:18:37

advertising predicts money earned in sales Right So let's say that we have

1:18:43

this data set and uh we we have um

1:18:50

uh the amount of uh money spent in light advertisement and the amount of money

1:18:57

spent in aggressive advertisement Um and we want to predict how much the

1:19:03

the the money spent in this advertisements predicts how we we're

1:19:09

going to earn in sales Okay So in this case uh what we want is to kind of fit

1:19:15

that function um where our model will receive an input

1:19:22

and we'll predict the so the um the money earned in sales Money

1:19:30

earned in sales is what we want to predict and the input is the um the

1:19:37

the money spent in advertising

1:19:42

could that could be the sum of both light and aggressive Okay So um what

1:19:49

what we're going to do what our what our uh model is doing is we're going to learn this function a function uh that

1:19:57

predicts y uh as as a function of of the input

1:20:05

Okay So uh we want to come with um we want to

1:20:11

come with with with f with this function and regression The most basic model for

1:20:19

regression is also a line Um it it's what we call linear regression Um so the

1:20:28

the goal is not separate between types of observations but to predict the number

1:20:36

of a particular uh a numerical number of the of a particular observation So uh

1:20:44

linear regression is the most common um model So given here the the money

1:20:53

spent in advertising if we if we were to to do a

1:20:59

graphic we would see the money in advertising and the money earned in

1:21:05

sales So if we put this points in the graph we have um this this is the input

1:21:13

what we call the input space or the instances space So you're kind of depicting all of this

1:21:20

um all of the points Okay So in in this case we have one 2 3 4 5 6 7 8 9 I'm

1:21:27

showing only three but we have nine points of this data set

1:21:33

um what what we're going to do for example linear regressor would kind of

1:21:39

would we would come up with this line with this linear equation

1:21:45

um that would would help us predict the

1:21:50

value So when a new data point comes in right where you

1:21:58

know oh I've I've I had 0.5 um um,000

1:22:06

uh spent in light advertisement and.5 so a total of one uh in aggressive

1:22:13

advertisement If you want to predict how much money you will earn in the sales

1:22:20

you can go and put this unseen uh point in uh to to the test with your model So

1:22:29

your model let's say that your your point your new point

1:22:35

um um comes comes here right Uh so an

1:22:42

amount of money of one okay and the

1:22:47

amount of adver the money earned it will be given by the model so oops by the

1:22:54

model by by the line right so by this function f okay so let's say oh it will

1:23:01

be $1.2 2 million Okay Or 1.3 So a linear regression is the most

1:23:10

um simple method of regression tasks But we

1:23:17

have several regression um tasks We have for example polomial

1:23:25

regression So some sometimes you can see that here the line is not capturing it's

1:23:31

capturing the trend but it's not capturing all the points You could for

1:23:37

example uh fit not a line but a polomial

1:23:43

like that Right So so something like I will have to erase this but uh you could you

1:23:52

could um come up with a with a model that is a polomial

1:23:58

uh fitting Okay And we will discuss this what types

1:24:04

of models uh that we have um much more in depth

1:24:10

um in in the next weeks

1:24:15

Okay So now let's uh so again the the how can I come up

1:24:23

with this line How can I come up with this polomial For example let's think

1:24:28

about the line during training doing the learning Uh I will have to find the

1:24:33

parameters of that line So the intercept of the line and the slope of the line that um um that that is able the best uh

1:24:44

the classifiers that are able to capture this this this numerical trend Okay So

1:24:52

again we will have to kind of during training we're kind of getting better

1:24:58

We're kind of adjusting adjusting the slope um adjusting the line adjusting

1:25:05

the slope adjusting the intercept uh up until a point where we will choose the

1:25:10

red one here because it is the best model Okay

1:25:17

to finalize uh supervised models Um supervised uh supervised

1:25:25

learning

1:25:32

models We're going to talk about forecasting Um forecasting also

1:25:39

um lands um resides on the superver uh supervised learning methods because it

1:25:47

requires labeled data So one column of your data set has to be the

1:25:53

target And um for examples are to predict um

1:26:00

um to predict the uh a value of a stock in a stock market or um a a

1:26:12

uh a trajectory or something that is kind of happening in the future

1:26:18

So the the way to do it is exactly the same in terms of um of uh getting um

1:26:27

offering to the model a data set of inputs attributes that are inputs and

1:26:34

attribute that is the target So for example here we have a curve and we have

1:26:40

the curve up until this point So um um this this curve or this

1:26:48

trajectory will be used for this for example let's say that this is the value of a stock uh this will be used for a

1:26:59

um for training the model and what we want ultimately is the output after that

1:27:07

So we're going to output the model after that So the the value of the stock after

1:27:13

this point in time So one way to do it is with the data that you have up until

1:27:20

this point you organize a data set in this following way Say that you have

1:27:25

several points um x1 x2

1:27:30

x3 x4 x5 x6

1:27:37

x7 x8 and so on So you will organize the table like this

1:27:45

A first instance of your data set will be x1

1:27:51

x2 and x3 So you're going to take this three points as input and you will predict the

1:28:02

fourth So the target will be oops this is uh this is x4

1:28:09

Okay As a second instance you are

1:28:14

um you will kind of select you won't use x1 anymore You will use the point x2 x3

1:28:22

and x4 All of these points are his uh you have they are sitting in your data

1:28:28

set to predict x5 So to predict x5 So

1:28:37

um and so on in the next round you won't use x2 you will actually use x3 x4 and

1:28:45

x5 to predict x6 right so that would be the third

1:28:50

instance so what is your uh model learning your model learning is learning

1:28:56

from uh uh inputs x1 um x2 x3 to input

1:29:03

x4 for x x of t x of t + 1 x of t + 2 to

1:29:11

input uh x t + 5 uh sorry oh my god let me put

1:29:18

this way um let me just rephrase this so your model will be learning and let me

1:29:25

use a a proper notation you want your your your um

1:29:32

um your model will be learning the next point in time giving the previous

1:29:39

point the pre one previous of that and two actually two previous of that So

1:29:46

think about the first line the first row here you're predicting x4 based on x1 x2

1:29:53

and x3 So that's what you're doing you're predicting and you're you're going to learn how to do that with all

1:30:01

these points that you have up until this um at this time After that you stop the

1:30:09

training and then you deploy your model to kind of forecast other uh points So

1:30:16

forecasting is very important and um we actually uh uh call it regressive

1:30:26

models because you need points in the past uh to kind of construct your

1:30:32

supervised learning data set Um and they have several implica uh

1:30:39

several um uses in um in economics for example

1:30:48

Okay So now what we're going to do we're going to talk about

1:30:53

um we're going to talk about unsupervised learning So different differently from the supervised learning

1:31:00

Unsupervised learning uh has data obviously because it learns from data but you don't have the target you don't

1:31:08

have any feedback of the target or the patterns uh that you want So

1:31:14

unsupervised learning is very it's related to the to the task of clustering

1:31:19

and with clustering you can do several things for example you uh with anomaly

1:31:25

detection So uh what what is clustering So clustering is kind of separating

1:31:32

things into sim similar groups Uh and that is what ultimately unsupervised

1:31:39

learning does Anomaly detection is just an application of kind of a clustering where um normal things are grouped into

1:31:47

similar um similar things are grouped in this normal group and something that is

1:31:54

an anomaly is kind of separated from that Um but so let's let's take a look at

1:32:02

what um what a what a a clustering means Let

1:32:07

me get rid of this because I was kind of practicing So here we have a data set where where we have uh name age and the

1:32:15

income I don't have a target I I I'm not trying to predict from the age the

1:32:22

income So it it's one might say oh this is the target I want to kind of do a

1:32:27

regression from age to income No that's not what I want to do I just want the a

1:32:33

machine learning model an AI to find characteristics of this

1:32:39

data similar characteristics on this data So for example what we can see in

1:32:46

if we plot this data where we have age in the x axis and income in the y- axis

1:32:52

is that there are there are people that have low uh age they're younger and have

1:33:02

low income There are people that are old and

1:33:07

have uh low income And for sure there are people that are old and have higher

1:33:14

income Okay So if you if you want to try to sell something this could be informing

1:33:22

your business that okay look target the the customers that are old Why Well

1:33:29

because younger people for sure don't have high income right But you didn't

1:33:35

tell your machine learning to explicitly do this you kind of you kind of inputed

1:33:41

the age the income So the the attributes or the or the inputs and the output of

1:33:48

your model is just a similarity grouping similarity

1:33:54

uh grouping Okay the way to train this

1:33:59

uh we will see this in deta a little bit more in details but your algorithm starts with like like points that that

1:34:07

uh are equivalent to the center of these groups that it's that that are um that

1:34:14

the model is finding And with all this data they

1:34:20

um you're not teaching it but it's kind of discovering So you start with kind of

1:34:27

different uh centrids and I'm I I will have to use different colors because of the groups have different colors So it's

1:34:34

a blue it's a green and it's a red or orange I don't know I think it's red

1:34:39

So you start with different kind of um representations for these groups and

1:34:47

while the machine learning is assessing the data the training data it will be it

1:34:52

will better position the centroidids and when it's all finished training you will

1:34:58

have one um centrid here representing the red

1:35:05

group a blue one there representing the um uh blue group and the green will kind

1:35:11

of fall here So then you uh you can see

1:35:17

that um you will have you can kind of measure the the distance of the of the

1:35:23

points of each row that are depicted as blue dots here to the centrids and

1:35:30

decide if it's green if it belongs to the green to the red and to the blue

1:35:37

Don't uh don't be scared if you're um you're not completely understanding what I'm talking about We're going to we're

1:35:44

going to take a look at this uh clustering mechanisms in more

1:35:50

detail So let's talk about reinforcement learning Reinforcement learning is the is um um also uh

1:36:00

um a branch of of machine learning and it we could say that it's it's a newer

1:36:07

branch of machine learning that has recently been um um accepted as a branch

1:36:13

of machine learning Um it gained momentum in the late 20 uh uh 2010 for

1:36:20

example Uh it's uh it's it's it's the basis of the success of deep mind uh

1:36:26

chess engine the alpho uh engine and um the alpha fold which is the the the

1:36:34

system that predicts protein proteins um structures um and that gave the Nobel

1:36:41

prize to um Danny Savis from from from Google

1:36:47

Um but the way it works it operates on a a

1:36:53

fundamentally it operates on a fundamentally different principle So instead of learning from uh labeled

1:37:01

examples supervised um kind of learning or finding patterns in an unlabeled data

1:37:09

such as uh unsupervised learning It learns from what we call interaction and

1:37:15

feedback Right So it reinforcement learning is is more like training a pat

1:37:21

and it is very related to the way humans um uh learn by experience Right So think

1:37:30

about a baby uh learning how to crawl it in first attempts are are failures Um

1:37:38

but at with practice and and and um and exploring the environment uh the

1:37:45

baby will will uh kind of figure it out So what we call is that the model learns

1:37:52

through trial and error Okay And uh the way we um the learn happens is that it

1:38:00

this this model gets rewarded for uh good decisions right and penalized for

1:38:09

uh poor decisions So for example the baby if um it tries to crawl uh with

1:38:16

only one hand it will it will be kind of penalized because it will the baby will

1:38:21

fall and hit his head his her head Um but as soon as the baby kind of

1:38:30

um get an action right of putting one arm after the other it will get a reward

1:38:37

because it will it will get to where it wants Okay He or she wants So

1:38:44

um it's it's a particularly uh powerful task and it it's usually um it's usually

1:38:53

used for optimization when you're so if you think about again the baby crawling

1:38:59

what the baby is doing is optimizing its decision to complete the the the task

1:39:05

Okay So in businesses you could think about maybe optimizing business

1:39:11

strategies uh when you don't have um uh

1:39:16

uh domain knowledge So lots of domain law knowledge in terms of rules It's a

1:39:23

very complex environment um it's um you don't have labeled data and you

1:39:32

don't have um you want to kind of know what is the

1:39:37

best strategy So this models can come up with an idea of what is good or a bad um

1:39:45

decision uh outcome um decision and and kind of a um strategy

1:39:53

Usually machine learning courses don't cover reinforcement learning as basic machine learning um like

1:40:01

um theory uh because it's still fairly uh

1:40:06

uh new and and and and niche but we will

1:40:12

talk a little bit more about it because it is part of large language models at some extent So here's the way it works

1:40:20

uh uh in reinforcement learning we learn the learning is uh held from experience

1:40:26

as a result uh of of exploration of what we call the agent or the model So your

1:40:33

model makes an action in an given environment and it will be uh penalized

1:40:41

or or will be rewarded So you can have a a positive reward or a negative reward

1:40:48

um and the the results of your actions will change the state So I'll give you

1:40:55

an example Uh let's say that this yellow agent this model wants to achieve a green square

1:41:03

here The environment is this um oh I'm not showing I didn't change

1:41:10

the slide Sorry So um here again it's the action environment a agent uh part

1:41:17

where I um described here and here's the example So so the yellow agent wants to

1:41:24

achieve the green square and it has to move in this lattice Um we don't we're not we're not

1:41:32

programming it uh we're not telling it the rules We're not hard coding it in

1:41:38

any way The yellow the yellow agent will learn

1:41:43

from exploring the environment The environment here is this puzzle is this

1:41:48

board is this lattice of squares and um and the actions are the movement So

1:41:55

for example let's say the uh oops So let's say that the agent tries to move

1:42:02

right because there's a wall there Uh well the the red squares means that um

1:42:10

it's it's a wall meaning that it it cannot do it cannot do this move It cannot move towards that direction So

1:42:17

the the agent will bump into this wall and will receive uh or a punishment or a

1:42:25

very small reward Right So this is an experience and he will keep that in a

1:42:31

data set and will learn from this experience that going to the right in

1:42:36

this particular state is not possible Okay Um so it took an action go to the

1:42:45

right in a particular state which is the first uh first position of the board to

1:42:52

the to the uh lower left and it got a reward and a new state The new state is

1:43:00

I didn't move at all I continue The new state is the same as before Then let's

1:43:06

say the uh the agent moved um um the agent moved

1:43:12

here to here right So it it went up and uh the action is going up The in

1:43:22

the new state is this new position So now I'm in the this this uh square here

1:43:30

and I wasn't rewarded at all So I keep exploring So I I kind of keep that

1:43:36

experience in my data set Now I for example let's say that I go to the right

1:43:44

and again this is my new state I wasn't rewarded But say that I try to go up now

1:43:50

So when I try to go up I won't get it I won't do it So why Because there's a

1:43:57

wall So if I try to go up okay my I will

1:44:03

receive a reward The the I won't change this the state will will kind of be the

1:44:08

same and I um um after my action of trying to go up

1:44:15

um but I will be penalized if I do if I try to go down uh

1:44:21

the same thing And so it it it keeps exploring

1:44:26

Okay Uh it keeps exploring up until a point Eventually it it will find a way

1:44:34

uh um to get to to the to the green point So it will find a trajectory that

1:44:41

takes him to the goal and once he hits the goal he gets a big reward Okay So

1:44:47

for example he came here um so he was exploring during learning and keeping

1:44:54

all of this data in a data set and he made these kinds of actions and then he

1:45:01

landed here Once he landed here he got a,000 points reward Okay So after all

1:45:09

this playing this one round of playing and getting to the goal you have several

1:45:14

data points of actions and its rewards and then you can use this data set to

1:45:20

train So the training mechanisms of reinforcement learning are very very

1:45:26

cons uh complex Okay we're not going to to go into details about reinforcement

1:45:32

learning but the concept must be very um but the concept it it's good to have the

1:45:37

concept clear of what it's it does uh because again it's part of larger language models at some uh for some

1:45:44

extent to some extent and um so what why are we talking about optimization Well

1:45:51

this agent learned through exploring um how to to win the game So it's kind

1:46:00

of doing uh if you keep exploring maybe you will find one way another way

1:46:06

another way and ultimately you can find better ways than others to

1:46:13

um to uh accomplish your goal It depends on how much exploration the agent does

1:46:18

But what we call this a um um an optimization kind of action right So one

1:46:26

example of reinforcement learning is the uh play playing video games So uh Google

1:46:33

has several papers on um agents and models that play it Atari games for

1:46:38

example or the game of Doom uh where they don't teach anything They just they

1:46:45

just kind of um release the agent with the model randomly exploring the

1:46:52

environment getting rewards which are the video game kind of points or lives

1:46:58

and after using this all this data from several runs the the the agent will kind

1:47:06

of optimize its actions to win the games Okay Now finally let's talk about

1:47:15

self-supervised learning Again this is uh as reinforcement learning it's it's

1:47:20

it's a very dense um technical um

1:47:28

um subject but it's important because it's part of the large language model So

1:47:34

self-super supervised learning is a kind of a

1:47:41

um mix between supervised and unsupervised learning So it's a kind of

1:47:46

a mix between both worlds and the the

1:47:51

way it the the tasks associated with this kinds of this kind of learning are

1:47:58

what we call natural language processing tasks So for example text summarization

1:48:04

translations uh questions and answers sentiment analysis which which is a classification

1:48:11

task So actually self-supervised learning can also do classification and

1:48:18

regression and do also NLP tasks Um so

1:48:24

NLP is is tasks that we accomplish uh with language right So summarizing texts

1:48:31

translating texts uh answering questions So the GPTs

1:48:36

uh large language models um so the GPTs which have the uh the name of their

1:48:43

models or large language models are uh based on self-supervised

1:48:49

learning Other tasks that use self-supervised learning is image and v

1:48:54

video generation So nowadays you can go to any large language model and you can

1:49:00

ask to generate an image or a video based on your words uh or a description

1:49:05

that you're giving it It for sure this um

1:49:10

um these models are using some kind of self-supervised learning Okay And so

1:49:18

when we say when we talk about self-supervised learning we're

1:49:24

talking about models that have two different stages The

1:49:31

um the first stage of the model is to

1:49:38

uh take unlabeled data Right So uh the the self-supervised

1:49:47

uh learning is very good for when you have immense amounts of of unlabeled

1:49:53

data For example text all the text in the worldwide web in the internet they

1:50:00

don't have label They're just words right So you you don't have label you could kind of do a supervised kind of

1:50:07

task if you want to classify something but you will have some labeler to um to

1:50:13

label them So self-supervised learning has two

1:50:19

different kinds of of of tasks uh uh two

1:50:24

uh two underlying steps The first step is to take this amount this great amount

1:50:30

of unsupervised data For example you have images of a cat of cats and dogs

1:50:36

but you don't have the label You just have the images You don't have that last column with the targets Um but somehow

1:50:45

the first part is um of the self

1:50:50

supervised learning is kind of generate automatically generate pseudo labels

1:50:58

Pseudo labels So you take unsupervised data and

1:51:03

you have a mechanism to uh call it pseudo label So in this first step the

1:51:10

mechanism will kind of create pseudo labels for all the cats and a different

1:51:15

pseudo label for for the dogs So in the end of the day uh if you think about a

1:51:21

data set of several images or uh what you're doing is you're you're you're

1:51:27

you're kind of putting cats closer to each other dogs closer to each

1:51:34

other and more distance from the cat And for example helicopter more even more

1:51:39

distance because it's not even an animal So you're kind of having good

1:51:45

representations of unlabeled data And then after that you're you're you are

1:51:54

um using a supervised learning technique to accomplish a task Okay So if let's

1:52:02

say that after doing pseudo labels and un un um so you don't have the labels I

1:52:08

have to take this out uh but you kind of separated them in in this representation

1:52:14

the better representations then you can come here and say okay given this pseudo

1:52:21

representations given the pseudo labels um um given this this uh better better

1:52:30

representation find an animal So then you will use a

1:52:36

supervised way So you have uh you will have a

1:52:41

target a target column and here you will have the representation or the

1:52:48

coordinates of the images Right So let's say they have x and y

1:52:53

values and you say well this one this one this one and this one So I don't

1:53:00

know all XY is positive Um they are

1:53:07

animals and animals We have four of them

1:53:12

there And then we have one that has a X that is

1:53:18

negative Uh and let me put five here This is not animal So you will kind of

1:53:26

generate a supervised data set and your classifier will be very it

1:53:35

it will make um um lots of difference if you have this representation of the data

1:53:44

uh more organized because then it's easier for your uh cla uh classifier to

1:53:50

see that if a if the xcoordinate is negative then it's not an animal Okay So

1:53:56

it makes the super uh the supervision task way more efficient and that's what

1:54:03

is um underneath the large language models

1:54:09

Okay Um don't think about you don't need to think this about images You if you

1:54:15

substitute everything that I told you with words um it's the same thing So imagine that

1:54:23

you're organizing words in a coordinate and you have cat cat dog dog word word

1:54:29

dog word cat word helicopter Uh the word dog and word cat will be will be closer

1:54:35

to each other and the word helicopter is uh distant apart So if you want to do a

1:54:41

prediction of the next possible word or a translation or a summarization

1:54:47

um you can use a this you can kind of come up with a supervised data set that

1:54:54

uses this uh this better representation

1:55:00

uh and do a more effective task So in the end of the day we have a uh step one

1:55:06

you you will generate pseudo labels in order to get better representations of

1:55:12

data better features that represents the data So the features are not uh they're

1:55:18

the data is not kind of not organized Now you're organizing and you're going to get this features to kind of come up

1:55:26

with a supervised data set to accomplish some kinds of task

1:55:34

Okay so we we talked about all of this um

1:55:41

supervised unsupervised self-supervised and reinforcement learning method uh

1:55:47

types Here is a is a summary for for you

1:55:52

Now as I said here in the um in the last part that we we want to we need to

1:55:59

represent our data effectively Okay So

1:56:05

this will uh leads us to the next item that is the importance of having good

1:56:14

features in your data Okay So let's say that

1:56:20

um uh you want to have a classifier or uh you want to have

1:56:29

um a regression model whatever um whatever kind of model that that you're

1:56:35

working with right so let's say that

1:56:42

um just one second that I lost my comments ments I lost it Where is

1:56:50

it I lost my comments Where is

1:56:57

it I don't know where my comments are Um okay So uh so

1:57:07

the Oh I found it Oops Oh

1:57:12

man Okay So let's say that you're building a a

1:57:19

classifier of uh for example images and the output is saying if it's a car or

1:57:25

not a car So it we have to have fe the features the attributes of this of this

1:57:34

data set they they they must be relevant For example if one of the attributes is

1:57:40

color it will be very hard for the classifi if if you're working only with

1:57:46

color and and uh and height of the of the of

1:57:54

the of the object in the image This will be very hard because like cars and

1:57:59

motorcycles have kind of the same height and they all are red or or or or gray or

1:58:07

black or green So you you need to have some features that are very important

1:58:14

Okay So we need to extract features and this is crucial for the the the the

1:58:21

success of the model Okay So the the model needs to kind of capture the important aspects of the

1:58:28

data that is related to the prediction it's doing right So again color and

1:58:33

height will won't do it So the the uh there's a field called feature a feature

1:58:42

extraction or feature selection that experts use machine learning algorithms

1:58:48

as well to kind of accomplish it they can use also statistical algorithms to

1:58:54

kind of see to find what what attributes are better So for example one

1:59:03

one attribute would be the width right So because motorcycles are kind of thin

1:59:11

and cars h they are they're they they have a larger width that would be

1:59:18

a a um a a feature that is more relevant to the classification uh model Okay Um

1:59:28

so during data mining you you kind of have to discover which ones are most

1:59:36

important kind of what features are most important and sometimes having lots of

1:59:41

data will make you will make you see uh things that um things that are weren't clear

1:59:51

before So you you never thought about that but using algorithms you can check

1:59:56

which column is more relevant to the task that you're

2:00:04

accomplishing Okay And uh just to uh uh wrap this part of feature extraction

2:00:12

um I I want to I want to talk about a uh what we call feature

2:00:20

engineering Sometimes uh it's also called it's it's also part of feature

2:00:26

extraction but um but you can also um hear this term as

2:00:33

feature engineering So what what is feature engineering So sometimes

2:00:40

um in the in this process of discovery and in the process of extracting

2:00:46

features that are relevant to the task it might be that you need to create new

2:00:54

features So new attributes or combine existing

2:01:01

raw attributes uh in order to to have something that is um

2:01:09

meaningful for your uh for your machine learning task So here's an example uh

2:01:17

maybe you are um tracking uh purchases and there's a

2:01:25

column uh in the data uh the is the date of the purchase Now if you're if you

2:01:33

want to kind of build a prediction of whether the sales are um

2:01:41

increasing or decreasing or even if you're not predicting you don't have like a a model of prediction but maybe

2:01:49

you just want to explain some events Maybe it's splitting this feature

2:01:57

into two new ones where one is the day of the week and the other one is a uh is

2:02:02

is a is a feature uh telling you if it's a holiday or not

2:02:10

So you you you created this new features and these new

2:02:17

features for sure would be important for

2:02:24

um for the success of the model And again this feature extraction and

2:02:29

feature engineering are aspects that are so important and they can be a difference between a um a model that is

2:02:37

successful and a model that um uh that

2:02:42

fails or uh has not does not perform

2:02:48

um uh well Usually um there's a step um um in uh

2:02:58

like before you could u use this data to train your machine learning models to

2:03:03

accomplish tasks uh that we call data prep-processing So

2:03:10

um the data prep-processing step is the the analytics data mining and feature

2:03:17

engineering steps So is when you understand what you have you deal with

2:03:22

mi missing values Um you deal with kind of uh applying filters to the data

2:03:29

removing outliers Um then finding which filters and

2:03:36

important correlations um inside the the the data um there are

2:03:42

inside inside the data You come up with good features if you need to you create

2:03:48

them and um so this this is called pre-processing

2:03:53

and feature engineering like creating new data is is is a very

2:03:59

um um it's very common when the data is not structured So here's a concept that

2:04:05

is important Uh a structured data is the one that that each column fits a purpose

2:04:12

So for example you have name age income and the uh marital status So each figure

2:04:22

of uh of of a row uh each figure will um

2:04:29

um different figures will kind of appear in this in the columns of each

2:04:35

row um and and they are the same for each instances Now there there is data

2:04:41

that is not structured They're not organized that way So for example a Twitter um um post

2:04:50

um or a a book that's data as well You can use that right for example to uh for

2:04:58

uh accomplishing natural language processing But the data is not a structure uh uh it's just a bunch of

2:05:06

words um that you cannot separate in in columns right So when you have this

2:05:13

kinds of data you will you have to kind of engineer and try to organize or

2:05:19

structure the data as much as possible This is a time consuming and challenging

2:05:24

process So in businesses this is something that takes time and often

2:05:30

requires uh data analysts data scientists to work this pre-processing

2:05:37

uh task right Uh so they spend a lot of time and is for sure an art Um uh it you

2:05:47

have to have you have to have like experienced ex uh experts that know what they're doing and kind of the ways they

2:05:54

can go through to um engineer good data and do a grow a great pre-processing

2:06:02

step inside the pre-processing step And um we have feature engineering that we

2:06:08

just talked about Sometimes you have to scale the features Uh sometimes you have

2:06:13

to do dimensionality reduction data cleaning data augmentation Some of them are kind of

2:06:19

straightforward Feature engineering we discussed Data cleaning is when you have you want to get rid of outliers or kind

2:06:25

of um um some some data that has uh you know like unreadable characters

2:06:33

something that is kind of um it needs a cleaning data augmentation you you deal

2:06:38

with missing values or the generation of new um instances Uh sometimes you have

2:06:46

to generate synthetic data uh to because you don't have enough data

2:06:52

and we will talk about synthetic data um uh in in the in the in the next

2:06:59

weeks Uh sometimes we will you have to do feature scaling and dimensionality

2:07:04

reduction feature scale and dimensionality reduction We will also see uh in the next week when we're

2:07:11

talking about uh evaluation and how we

2:07:18

uh what are the aspects that kind of affect the goodness of an algorithm So

2:07:24

that's where uh I'm going to wrap up Uh so

2:07:30

after when you after kind of um doing the data science cycle uh

2:07:37

storing data cleaning data do pre pre-processing um then you kind of feed this data to

2:07:45

your model then then the the model is trained and then you will deploy the model Uh we have to remember that

2:07:52

several models can tackle diff uh the same task So for example classification tasks can be accomplished by a random

2:08:00

forest model or neural networks decision trees So don't worry we'll all see it what that what kinds of models uh uh

2:08:08

what the these are Um but you have to select a model that is kind of good uh

2:08:16

that is performing well in your task So data science is very much about

2:08:23

experimentation you sometimes come up with several different models and you will choose the best one amongst them So

2:08:30

in the end of the day we will have to compare models using some kind of metric uh a

2:08:38

metric that corresponds to the goodness of a model and we will select the

2:08:44

performance uh we will select the best model giving the performance of that of

2:08:50

the models uh um against this uh metric and using a common test data set

2:08:59

So as I said uh said we're going to understand uh in more details the

2:09:04

metrics of each uh uh uh kind of the the

2:09:10

metrics used to evaluate models depending on their on their task the

2:09:15

task they're performing but I will

2:09:20

uh explain to you the training and testing data division so that we can uh

2:09:25

understand a little bit more uh the evaluation

2:09:31

process So the the training sometimes also so the learning sometimes known as

2:09:38

training and also sometimes known as fitting is that process of adjusting a

2:09:46

model's parameter right to find a best model So if you if you think about the

2:09:54

the the linear regression is trying to find uh like the line the slope of the

2:10:00

line and where it intercepts u until you have what we call a best fit

2:10:08

So this training data is the the the data that is sitting on your data set

2:10:14

past data Uh it can be labeled or unlabeled depending on your uh task the

2:10:21

task that you're accomplishing But the model is learning

2:10:26

from this training data Okay Now the uh

2:10:33

the test data this this similar test data is a test it's a collection of data

2:10:40

that you you separate you we call it you a hold out you

2:10:49

separate so that you could evaluate the goodness of your uh of of of your model

2:10:58

and even compare it with other with other models So um let

2:11:05

me let me uh just go back to the data set where we wanted to predict the um if

2:11:16

it's going to rain or not Um so rain or no rain for several uh

2:11:24

instances Okay Um

2:11:30

and we had humidity values and pressure values This was the

2:11:35

target Um so we had several values here and we could just go back to that um

2:11:43

figure Um and let's say so the way we do it is

2:11:51

we we hold out a a a quantity and an amount of data that we know a labeled

2:12:01

data that exists in our data set set We hold out and we call that a test data

2:12:07

Usually is a smaller portion of our our data set some uh usually 10% of the data

2:12:16

5% of the data of the data points of the instances But what you're going to do is

2:12:24

you're going to train your model here in in the in the uh kind of instances of

2:12:34

day one and day two And on day three and and day three and day four you're not

2:12:39

going to use for training you're going to hold out and you're going to

2:12:45

deploy your your trained model here as a test So for example if you're doing this

2:12:54

classification one one uh metric and we will see metrics um uh in details could

2:13:02

be the accuracy So the number of um correct predictions that you had over

2:13:11

the total number of predictions So what you're going to do is you're going to

2:13:17

test you're going to get this data

2:13:23

um marine and rain So you're going to you're going to get this hold out data

2:13:28

this hold out data in which you know what the the model should predict and

2:13:35

then you're going to calculate the goodness metric So the goodness metric is calculated exactly in this

2:13:44

test uh data set and then you can uh deploy you can you can kind of test this

2:13:52

um uh this data you can you can test or evaluate

2:13:57

uh several models So you can you can uh for example train different neural

2:14:02

networks uh you can train an SVM or a decision

2:14:07

tree So you can compare the performance of this four um different models using a

2:14:16

metric in this test data set because you know what

2:14:22

it should be outputting and so you can calculate the

2:14:28

goodness or how good your um algorithm is your model is But don't worry we will

2:14:36

see this in details um in the next week So um we will uh uh

2:14:43

pick up exactly in in this uh evaluation part of the course

2:14:51

So in the next weeks we will going we're we we will uh pick up here in

2:14:58

evaluation and we will open aspects of the training and testing so that we can

2:15:05

understand better the evaluation of models we will okay so I will leave uh

2:15:13

to finish I'll leave you here with a with a little bit of a of a a building block of AI So first you collect the

2:15:20

data we pre-process this data uh we we engineer features or extract features

2:15:28

Then we train and we evaluate models We

2:15:33

uh we will so you you you can train uh tune we will differentiate training and

2:15:40

tuning um as well But understand training and tuning like as kind of

2:15:46

fitting the model finding the the the best parameters of your model as we saw

2:15:53

with uh the intercept or the slope of a cur of a of a line And then you will

2:15:59

evaluate this models and you will choose a model to deploy right So you will

2:16:05

deploy you will do your classification task or your forecasting tax or task or

2:16:10

your summarization using chat GPT for example or other LLM that kind of won

2:16:16

the evaluation and then you just keep monitoring uh and you can always kind of

2:16:24

perform this as a cycle you can collect more data and you kind of can

2:16:31

uh do this uh in in a kind kind of feedback loop where you all you're

2:16:37

always getting your your your models better All right thank you so much and

2:16:44

uh this is week one Um just uh to kind of wrap up here as a

2:16:53

comment uh I'm not sure if I should already talk about what metrics uh like give more

2:17:02

details um regarding the metrics I have this slide here where like oh if you're

2:17:09

doing classification and I would define accuracy here um and if you're doing regression the

2:17:17

mean squared error or if I I mean it it is more specific uh to each model So I

2:17:23

was thinking about talking about this in the in the next

2:17:29

weeks So this is something Harshida for for Harshida now something that I would

2:17:34

love to have your feedback uh of what you think it would be good One thing that I noticed after the like I finished

2:17:42

and I have some so many comments to kind of uh um uh change here

2:17:48

but when I'm when I'm starting to explain the the tasks I am definitely

2:17:56

going to not use I'm going to use training and not test yet because I I

2:18:02

kind of I I never um defined this I would just I would just say that we're

2:18:08

deploying in unseen data right so I will go this way like

2:18:15

deploying in unseen data for all the types of uh of machine learnings um

2:18:23

uh everywhere I kind of um

2:18:29

show data sets right I will do that in that way

2:18:35

And after that when I come so so so then

2:18:41

when I'm talking about evaluation I can define what is the test

2:18:49

the training and the test and you know like the percentage of the data split

2:18:55

but again the way I did it uh so I have to to just be careful with the script

2:19:01

not to say test before this slide side um because then I'm going to explain the

2:19:09

the test set the the division and next the week two I will start in evaluation

2:19:17

and I will talk about the t the divisions on uh training validation and

2:19:23

test So so I'm I'm going to go in more depth and also explain the metrics but

2:19:29

the way I did right now I didn't explain how the training like the loss

2:19:35

function like how to adjust this uh models

2:19:41

um and what kind of metric we use

2:19:47

to kind of measure the goodness So Harshida if you think I should uh change

2:19:53

something here let me know Uh it's it's kind of different ways that we

2:20:00

could uh convey the message Let me know if it's understandable or if you think

2:20:06

we could we had to change that uh the video if you kind of um um it's

2:20:13

it's very long but I'm talking very slowly and kind of uh uh

2:20:20

brainstorming and also I think like if we uh with a script where you you're

2:20:26

reading you're not gagging or um it's it's more clean you have the images to

2:20:31

show like a professional thing I would say that the like the the video would

2:20:36

reduce a thought So it might be uh that you will give me a feedback like no tell

2:20:43

us a little bit more about the metrics at this point because it's not understandable Uh include something and

2:20:50

remove something So that's the kinds of uh feedback that I want from you

English (auto-generated)

All

For you

Recently uploaded