0:01

welcome to the course In this module we'll explore artificial intelligence

0:06

What it really means how it learns and how it's being used across industries to

0:11

create value AI is everywhere powering recommendation systems on Netflix and

0:17

YouTube optimizing logistics for Amazon and embedded in our smartphones

0:23

It can become essential to modern business So what is artificial intelligence artificial intelligence is

0:30

about getting computers to do things that require human intelligence

0:36

Understanding language reasoning navigating the physical world learning

0:41

and predicting the future are all tasks that require human intelligence

0:46

So when we program computers to perform them we call it artificial intelligence

0:53

AI is increasingly present in our lives and it is considered the next phase of

0:59

digital transformation In the late 1990s the internet transformed businesses Then came cloud

1:07

computing followed by mobile computing and the internet of things in the 2000s

1:13

Now AI is the emerging digital technology posed to transform businesses

1:21

Companies that were slow to respond to past digital transformations were left

1:27

behind We know then from experience that businesses must act quickly to respond

1:34

to respond to AI What are the implications for businesses

1:40

first even if you're not in a an IT related industry this transformation

1:46

will still affect your business AI is a generalpurpose technology being applied

1:51

across many domains Managers need to understand the technology its

1:56

applications and make the necessary changes to embrace this digital shift

2:02

These changes may involve a new business model updated technology and infrastructure revised organizational

2:10

processes and changes in the company's culture

2:15

Let's start diving deep deeper into AI AI is a term used to describe computers

2:22

performing tasks that that humans typically do There is a common taxonomy of AI and

2:29

experts usually agree on two major categories Expert systems and learning

2:34

systems Expert systems involve computers doing what humans would do but with

2:41

hard-coded rules The knowledge in these systems is entered manually by humans

2:47

For example in health care doctors create diagnostic rules based on their expertise and this and these are

2:55

programmed into a computational system The computer then uses the rules to

3:02

provide diagnosis A key characteristic of expert systems

3:07

is their reliance on fixed hard-coded logic written by domain experts

3:13

However the systems face challenges when dealing with unexpected situations or

3:19

overlapping rules While inspired by human reasoning and logic they are

3:25

limited in flexibility Modern examples of rule-based AI systems

3:31

include credit scoring engines used by banks or systems that assess eligibility

3:37

for insurance claims The systems rely on predefined rules to determine outcomes

3:45

Computers are essential in this context because aggregating and applying all the rules manually would be too complex

3:54

Although effective the systems have limitations For instance an AI chess player could

4:00

function as an expert system by using the rules of chess But if the game's

4:05

environment changes in a way that wasn't anticipated or programmed the system

4:10

will fail to adapt The other branch of artificial

4:16

intelligence is machine learning or learning systems The systems allow computers

4:22

uh to learn from data and improve their performance over time without being explicitly programmed for each task

4:31

Instead of coding step-by-step instructions the machine learns from examples and patterns in the data This

4:39

approach allows the system to generalize to new tasks and solve problems for

4:45

which no clear rules exist It is somewhat inspired by neuroscience

4:51

and has been validated in practice An example of learning system is a machine

4:56

that predicts whether a customer will purchase a product It can analyze

5:02

browsing behavior uh type spend on a product page and past transactions to

5:08

estimate the likelihood of purchase There are no hard-coded rules for this

5:14

Instead the system learns from the data This process involves feeding the the

5:21

machine many examples of customer behavior From that data it learns to

5:26

combine features or characteristics such as browsing patterns time on site and

5:32

transaction history to predict

5:37

purchases This course will prim primarily focus on

5:42

machine learning In technical literature you may also encounter the terms

5:47

symbolic AI for expert systems and statistical learning for machine learning

5:59

Okay because we're focusing on AI learning systems or machine learning systems we need to understand data Data

6:07

refers to the examples from which AI learns A data set is a collection of examples

6:14

For instance this could be a collection of customer purchase history the records

6:20

of a customer purchase history You might have fields like order ID product

6:25

purchased category product code seller name and purchase price This is a data

6:32

set a structured collection of records representing real events that the AI can

6:38

learn from While this is a numerical example data can also be images text or

6:46

any form of input that computers can process Another example is weather data which

6:53

might include temperature humidity and wind speed This type of data could be

6:58

used to predict the weather forecast for the next day Each row in a data set represents an

7:05

individual instance or example Here one row is representing a single

7:13

purchase event These are sometimes also referred to as data points or samples

7:20

The information in each column is called an attribute or feature For example the

7:26

product category is a feature of the data set Different instances will have

7:31

different values for each feature In a weather data set each row might

7:38

represent one day of record Columns would include attributes like cloud

7:46

cover cover wind humidity and air pressure

7:54

Some attributes are numerical such as purchase price or temperature Others are

8:02

categorical such as product category or weather condition For example cloudy or

8:08

sunny Attributes can also be text or images Regardless of format all data must

8:15

eventually be transformed into a numerical form that a computer can

8:20

process An image can also be an instance in a

8:26

data set Some attributes are numerical such as purchase price or temperature Others are

8:34

categorical such as product category or weather condition For example cloudy or

8:40

sunny But don't worry we will learn more about data types in future lessons

8:47

Regardless of the type of the data in the end all data must must eventually be

8:52

transformed into a numerical form that a computer can process An image for example can also be an

9:00

instance in a data set Images consist of small squares the pixels and each pixel

9:08

contains intensity values for red green and blue

9:14

The data set would then include columns for the intensity of each color channel

9:21

for each pixel If you have a 100 images you would have 100 rows in the data set

9:28

In summary data sets are collections of examples from which machine learning models learn However not all data is

9:36

useful You need highquality well-prepared data sets for effective learning

9:42

For example a ride sharing app might collect data such as pickup location

9:48

drop off location time of day driver rating and fair amount

9:53

But if the data is inconsistent or contains many missing values the AI

9:58

model will struggle to learn In this case it won't be able to accurately predict the best driver or optimal fair

10:05

for a user Okay So continuing

10:15

our discussion on data we are surrounded by it and its true value lies in what we

10:21

can learn from it The goal is to take raw data and transform it into something

10:26

more useful such as information or predictions that can be applied in real

10:33

world scenarios Useful information can help support decisionmaking and

10:38

predictions can forecast what might happen next based on patterns in this

10:43

data Large volumes of data allow us to uncover valuable insights through an

10:50

exploratory and iterative process We explore data identify correlations

10:56

between attributes and detect dependencies These discoveries can lead to actionable

11:03

insights which may influence business's decisions

11:09

Here's an example When you check out a supermarket and use a loyalty card you

11:15

receive personalized coupons These are generated based on your

11:21

demographic profile and purchase history The supermarketer supermarket aggregates

11:27

this data along with data from millions of other customers and runs experiments

11:32

to determine what coupons to offer next It's a system of individualized pricing

11:39

You receive tailored barbins and the store increases sales

11:48

This process of gathering and analyzing data generating insights and making decisions is part of a broader field

11:55

called data science Data science involves more than just transforming raw data into insights It

12:03

includes the entire workflow acquiring cleaning storing and analyzing the data

12:11

to uncover meaningful information And in this it also includes building

12:19

models that predict future events Consider a telecom company that wants to

12:26

understand why customers cancel their plans and to develop strategies to retain them The data science process

12:34

begins with collecting cleaning and storing customer data Then

12:40

using data analytics and data mining analysts search for hidden patterns and

12:45

relationships such as a decreased such as a decrease in the number of calls

12:51

being a predictor of cancellation Once these patterns are identified

12:56

machine learning models can be developed to predict the likelihood of a customer churn The insights gained can inform

13:04

retention strategies and guide management decisions The output of this process often

13:11

includes reports and data visualizations which are key tools in the business

13:16

settings In this

13:22

you just heard me say data mining data analytics and data science You'll often

13:28

hear these terms Here is how we can distinguish them Data analytics focuses

13:34

in focuses on examining the current data to understand patterns trends and

13:40

distributions often through statistics charts and

13:45

basic analysis Data mining goes deeper by finding

13:50

meaningful features correlations and hidden patterns relevant to specific

13:55

questions like identifying which variables predict the cancellations

14:02

Data science encompasses both analytics and mining plus the full pipeline

14:08

gathering the data preparing it analyzing it building predictive models

14:13

and deploying those models Here is an analogy Imag imagine a a chef in a

14:21

kitchen Analytics and data mining involves checking what ingredients are in the fridge and deciding which are the

14:28

most promising Data science is the full culinary experience Choosing the ingredients

14:35

cooking the meal plating it and even predicting how much the d uh the the

14:41

diners will enjoy it While these distinctions exist in practice they often overlap Data

14:48

scientists use analytics to validate models and apply similar tools and techniques as analysts

14:55

Still understanding the terminology is important because you will encounter these terms frequently

15:03

Moving forward the value of data lies in learning from it and business must act

15:10

after that learning It is essential that managers know how to interpret findings

15:17

ask the right questions apply this insights and recommendations and understand the predictions made from

15:25

the data science process to solve real business problems

15:31

Managers need to be digital digitally savvy with both data literacy and domain

15:36

expertise in their industry Not every pattern or correlation in the

15:41

data is meaningful Some may be esperious Data scientists

15:47

while skilled in analysis and in analytics might not have the necessary

15:52

domain knowledge to filter out irrelevant or misleading patterns

15:57

That's why managers must understand the powers and limitations of data analytics

16:02

to critically evaluate the outputs Organizations often employ data

16:08

scientists she chief data scientists or data architects These professionals manage data flow

16:15

design the infrastructure and make technological decisions related to

16:21

capturing storing organizing and analyzing data They also select

16:27

analytical tools and frameworks choosing between third-party software with built-in machine learning capabilities

16:34

or custom solutions The goal is to transform complex data

16:40

sets into actionable insights that inform decision-m

16:46

since it's critical to turn insights into actions The key takeaway is that

16:52

data literacy and commitment to digital transformation are crucial for driving these decisions

16:59

So let's step back and look at how we got here In recent years we've entered the age of big data As the world

17:06

digitized through smartphones appliances and internet connected devices huge

17:11

volumes of data began to be generated Every online transaction credit card

17:17

swipe email phone call or tap past the security camera for example produces a

17:24

small piece of data Thanks to wearable devices social media and the internet of

17:29

things we are we now have massive volumes of data

17:36

This explosion in data coincided with increase increases in computing power

17:43

enabling us to store process and analyze it efficiently This convergence explains

17:50

why we're now in a golden era of AI where machine learning thrives because

17:55

the required data is finally available Previously companies collected small

18:01

amounts of operational data Now the data companies can gather are big data

18:09

meaning that this data is characterized by volume There are massive data sets

18:15

measured in terabytes or pabytes Variety structured unstructured text images and

18:23

multimedia Velocity real time or high-speed data streaming and veracity Data often

18:31

includes inconsistence inconsistencies and missing values

18:38

Managing data at this scale is complex and while this course doesn't focus

18:43

specifically on big data infrastructure you'll find additional resources provided if you'd like to explore this

18:49

further Let's talk about managing data for machine learning and AI models Managing

18:56

data involves collecting cleaning storing organizing and transforming it

19:01

into formats that support insight generation The foundational step in this process is

19:08

using data management systems DBMS commonly referred to simply as databases

19:15

The systems store structured collection of data allowing efficient retrieval and

19:20

management Businesses use a variety of tools to manage databases such as SQL or

19:27

cloud-based big data platforms like Snowflake or Datab Bricks

19:32

Although this course doesn't focus on data infrastructure you'll find optimal materials optional materials to explore

19:39

this area further Databases vary in type You have operational databases where you

19:46

can store realtime data from daily business operations The these are often local and optimized

19:53

for quick access And you can have data warehouses that store aggregated and

19:59

historical data typically in the cloud for large-scale analytics and machine learning applications

20:06

For example take Airbnb Every booking review photo and location generates data

20:12

If a customer logs into view recent bookings the system access an

20:18

operational database for real-time information At the same time Airbnb maintains

20:24

centralized data warehouses containing historical and aggregated data from

20:29

across its platform This data supports insight generation and predictive

20:35

modeling Airbnb might use machine learning models to recommend listings or adjust prices

20:43

The model would be trained on historical data such as past bookings and user

20:48

profiles requiring access to large integrated data sets stored in data

20:54

warehouses Let's now talk about machine learning

21:00

models a central topic in this course Now that we understand machine learning's role within the broader data

21:06

science cycle where data is leveraged to generate insights and build tools that

21:12

perform human-like tasks we can look more closely at what machine learning models are and how they work

21:20

Machine learning models are computation and mathematical models that sit inside computers and per perform specific tasks

21:28

As we've discussed AI enables computers to carry out tasks that typically

21:33

requires human intelligence such as classification prediction and

21:38

reasoning A model receives inputs performs one of

21:43

the tasks and produces an output This models learn how to perform tasks from

21:50

data They learn to recognize patterns and use them to produce outputs such as

21:57

predictions or classifications For example if we present images of dogs

22:02

and cats a machine learning model can learn to classify whether a new image is

22:08

of a dog or of a cat This is a classification task Another example is

22:15

predicting the stock market The model will take historical data as input and

22:20

produce a forecast as output Similarly weather prediction involves input

22:26

variables such as air pressure humidity and wind conditions to forecast

22:32

tomorrow's weather In each of these cases the machine learning model learns

22:37

a mapping from inputs to outputs Mathematically we could say that the

22:42

output y is a function fx of the input x where f is the machine learning model

22:51

The model is then approximating this function this f ofx using data without

22:59

requiring hard-coded instructions Consider another example an email spam

23:06

detector The model might take features like the number of URLs in the email

23:12

whether the sender is in your contact list or not and other attributes It will

23:18

then classify a new email as PAN or not span because it learned from several

23:27

email samples To train this model we need labeled data

23:36

in this particular example So we need examples of past emails with features

23:44

or attributes and their corresponding labels if the email is span or not span

23:51

The model uses these examples to learn pattern and relationship so that when it

23:56

encounters a new email and unseen data it can accurately classify it based on

24:02

learned rules So let's summarize A machine learning

24:07

model learns relationships between inputs and outputs from data It also makes predictions classification or

24:14

decisions And it used past examples to generalize to new situations

24:22

Now let's talk about the types of learning in machine learning These

24:27

define how the model learns from data and are four main types supervised

24:34

learning unsupervised learning self-supervised learning and

24:39

reinforcement learning It is important for us to know how to differentiate this

24:44

learning because the learning has to do with the task that we need to accomplish

24:50

So supervised learning the model is trained on labeled data

24:56

So input output pairs where the correct answers are known This is like teaching

25:02

with an answer key Tasks such as classification and regression typically

25:10

uses supervised learning Unsupervised learning The model is

25:16

trained on unlabeled data The model must find patterns or groupings in the data on its own

25:24

Examples of tasks uh in unsupervised learning is clustering and anomaly

25:31

detection Self-supervised learning This is a hybrid approach The model generates its

25:38

own label from unlabeled raw data This technique is commonly used in large

25:44

language models like those uh models behind chat GPT and support tasks like

25:51

summarization and translation and other natural language processing tasks

25:58

Reinforcement learning The model learns by interacting with an environment It

26:04

receives feedback in the form of rewards or penalties and uses this feedback to

26:10

improve its performance over time This approach is often used for optimization

26:16

tasks and it's also used in self-supervised learning in to optimize

26:24

um this large language models

26:29

Each type of learning is suited to different tasks Supervised learning

26:35

usually is used for classification and regression Unsupervised learning

26:42

usually used for clustering and anomaly detection and self-supervised learning

26:48

used for language tasks such as summarization translations and finally

26:55

reinforcement learning used for optimizing and decision making in

27:01

dynamic environments We are going to see all of these

27:06

learnings in a little bit more detail So let's take a closer look at supervised

27:13

learning Let's talk about supervised learning

27:23

Supervised learning is a foundational approach where models learn from labeled examples This means the data use uh used

27:31

includes both the input variable and the correct output or target The model is

27:37

shown the right answers during training For example suppose we want to train a

27:43

machine learning model that recognizes whether photographs are of Alan Touring

27:49

We collect a data set of images and for each one we provide a a label Alan

27:56

Touring or not Alan Touring or correct and wrong The model learns to classify a

28:04

new unseen photo as either Alan Touring or not based on all of this

28:11

data set This is known as a classification task

28:18

So the goal is that the model is to be trained on this label data set and then

28:24

deployed to make predictions on new unseen data A real world example is Tesla

28:31

self-driving system In this video the system mclassifies a

28:37

truck transporting traffic lights as actual traffic lights This occurred

28:42

because the model had learned from many examples of photographs of traffic

28:48

lights to recognize the traffic lights When it saw multiple lights on the truck

28:54

it classified them accordingly This illustrates how machine learning models behaved based on their training

29:02

Another application is identifying tumors in X-ray images The image the

29:08

input is the image and the output is the classification cancer or not cancer

29:15

Similarly we might predict rain based on input variables like humidity air

29:21

pressure and wind conditions To train such a model the data set must

29:27

include both inputs and the target label For instance we might have a data set

29:32

with rows showing daily weather measurements humidity pressure and a

29:39

label indicating whether it rained The model learns the relationship between

29:45

the inputs and the output in this case rain or no rain

29:53

In classification tasks the prediction is discrete meaning the output belongs

29:59

to a set of distant categories rain or no rain

30:04

Class A or B one or zero So in a data set you have several instances with

30:11

several columns meaning that the columns are the features or the attributes And

30:17

one of that attributes is the target class It's what we call the label

30:24

The goal is for the model to work on new data When a new day comes in we provide

30:30

the input data and the model predicts where it will rain or not This unseen

30:37

data wasn't part of the training set So the model will be deployed to make

30:44

realtime predictions in unseen data

30:49

We will discuss more training and deploying this models later but the

30:55

principle remains use labeled data to teach the model and then use it on new

31:02

on new data Another example is predicting marital status based on age

31:07

and income The algorithm learns to map the inputs age and income to the label

31:15

or the output married or single In the case of supervised learning

31:23

sometimes data has the label the data has the labels for granted in

31:32

structured data sets Sometimes humans must label the data For example in the

31:38

case of the Alan Touring classification task a human should review the images of

31:43

Alan Touring and mark them as Alan Touring or not Alan Touring

31:49

Now let's revisit the rain example Suppose we have a data set with

31:57

two features humidity and pressure So two columns The third column is whether

32:05

it's going to rain or not So it's the label Suppose we plot humidity and pressure on

32:13

a 2D graph Each point represents a day

32:18

labeled rain or no rain And we're plotting this A simple classifier could

32:25

be just a line that separates the data point by class

32:30

This line is actually our model So it's a model of a line and if a new

32:36

data point falls above the line the model will classify it as rain and if it

32:42

falls below it will classify as no rain This line this classifier is learned

32:50

during training Initially the model starts with a random line as it and as it learns from the

32:57

training data it adjusts the line to better separate the classes

33:03

You could come up with other different types of classifiers not necessarily a

33:08

straight line and then you can compare this multiple classifiers

33:14

with their performance in unseen data

33:21

Now let's move to another supervised learning task Regression

33:26

Regression involves predicting a continuous numerical value rather than a

33:31

discrete class It also requires labeled data So input and output pairs

33:40

For instance a company may want to know how advertising spending uh predicts

33:46

sales The input might be dollars spent on uh

33:53

light and aggressive advertising and the output is the revenue generated

34:01

A regression will model or learn this mapping from the spending or how the

34:12

uh company spent to the output of

34:19

how the company is selling or the revenue generated

34:25

The simplest regression model is a linear regression which fits a line to

34:31

the data The model predicts sales based on input

34:36

advertising spend If a new point is introduced the model uses that line to

34:43

predict the outcome Sometimes a linear model is insufficient

34:49

Obviously a polomial regression for example might be better to capture the relationship if the data follows a curve

34:57

We'll explore different regression models more deeply in future sessions

35:02

To train a regression model the learning algorithm adjusts the line to best fit

35:08

the data It tunes parameters such as the slope of the line and the intercept of

35:15

the line to minimize the error

35:20

The prediction error is what we want to minimize when we are learning an

35:26

algorithm It's what we call the prediction error

35:34

We will talk more about training and learning algorithms in future lessons

35:40

Now let's consider forecasting Another task under supervised learning

35:45

Forecasting involves predicting future values on past data For instance we may

35:51

want to predict a stocks a stock price tomorrow using historical data

35:58

One method is to restructure the data into overlapping sequences

36:04

For example the price of the stock in day 1 2 and

36:09

three predict day four Then day two three and four predict day five

36:17

and day three four and five predict the day six and so on So we can come up with

36:24

a data set like that where we have three different columns of past

36:32

data of past information and the

36:39

target column which is the prediction of the

36:44

next day The model will learn how to predict the

36:50

next point based on a sequence of past values This is called a

36:59

forecasting supervised model task as it requires historical points to predict

37:06

future ones Forecasting is widely used in economics and other fields where time dependent

37:14

predictions are critical Now let's talk about unsupervised

37:20

learning Unlike supervised learning unsupervised unsupervised learning uses data that

37:27

does not include labeled targets or expected outcomes The model learns from

37:34

the structure and patterns in the data itself without being explicitly told what to predict

37:41

Unsupervised learning is closely associated with the task of clustering And one common application of clustering

37:48

is anomaly detection In anomaly detection most data points

37:53

form coherent groups or clusters and anything that doesn't fit well into a

37:59

group is flagged as an anomaly So what does clustering mean clustering

38:05

involves grouping similar items together based on their attributes The machine

38:10

learning model identifies the natural groupings without prior label labels

38:16

This is the essence of unsupervised learning Let's consider a simple example

38:22

Suppose you have a data set containing individuals name age and income You're

38:27

not trying to predict income based on age Instead you're asking the model to

38:33

discover patterns or similarities in the data

38:38

If you plot age on the x-axis and income in the y-axis you may see that younger

38:45

individuals tend to have lower incomes Older individuals may have either low or

38:52

high income From this you might infer useful business insights For example for

38:58

instance targeting highincome older individuals uh for instance

39:04

targeting highincome older individuals as potential customers But this insight

39:09

wasn't programmed or labeled in the data The model found the grouping on its own

39:17

In the end this this results in groups of similar individuals similar data points The model doesn't

39:25

know what the group mean in human terms It simply organizes the data based on

39:30

similarity of its column If this process is still a bit unclear

39:36

don't worry We'll explore clustering mechanisms in greater detail later on

39:44

Now let's talk about reinforcement learning Reinforcement learning is a distinct branch of machine learning that

39:51

has gained traction in recent years It became specially perminent in the late

39:57

20 uh um 2010s through breakthroughs like Deep

40:03

Minds Alpha Go and Alpha Fold The the latter of which predicted

40:10

protein structures with such success it contributed

40:15

to a Nobel Prize awarded to Demi Hassabis at Google

40:20

Reinforcement learning operates on fundamentally different principles compared to supervised and unsupervised

40:26

learning Instead of learning from labeled data or pattern discovered in unlabeled data reinforcement learning is

40:34

based on interaction and feedback It mirrors how how humans and animals

40:42

learn from experience For example consider how a baby learns to crawl Early attempts may fail but through

40:48

exploration and trial and error the baby eventually discovers the correct sequence of movements

40:55

The baby receives positive reinforcement success like the parents clapping or

41:03

even not falling for their for its he

41:08

effective action and negative reinforcement failure or discomfort for

41:15

poor actions Similarly in reinforcement learning a model called an agent learns

41:21

to make decisions through trial and error in a given environment The agent receives rewards for good decisions and

41:29

penalties for poor ones This makes reinforcement learning especially powerful powerful for optimization

41:36

problems where the model is trying to discover the best strategy in complex or

41:42

uncertain environments where labels are not available Although most introductory machine

41:48

learning courses don't cover reinforcement learning in depth it is increasingly important especially in

41:55

modern AI systems like large language models which integrate components of it

42:02

So let's take a look of h how reinforcement learning works In reinforcement learning the model or the

42:09

agent interacts with the environment by taking actions Each actions leads to a new state and the agent receives a

42:16

reward positive or negative based on the outcome Over time the agent learns to

42:23

associate actions with better outcomes Here's an example Suppose a yellow agent

42:29

must reach a green square on a board The environment is a grid of squares and

42:34

some positions are block like blocks like walls Initially the agent doesn't

42:41

know anything It tries moving right but there's a wall so it receives a small or negative reward and remains in the small

42:48

position in the same position It records this experience of state

42:55

action new state and the reward Next it tries moving up It changes position and

43:02

adds that experience Over many such interactions the agent

43:07

builds a data set of experiences and begins to learn which actions lead to

43:13

better results Over multiple rounds of play the agent

43:19

collects many such experiences and can begin to optimize its strategy

43:25

Eventually after enough exploration the agent discovers a path to the goal the

43:31

green square Upon reaching the goal it receives a large reward for example a

43:36

thousand points a numerical reward This feedback reinforces the sequence of

43:43

actions that led to that success So over multiple rounds of play we're

43:52

not just finding a path but finding

43:57

optimizing a path to reach the goal more efficiently This process is why

44:04

reinforcement learning is often associated with optimization The agent not only not only learns how to succeed

44:14

but how to succeed better over time The more you explore the environment the

44:20

closer to an optimized action you will you will um obtain

44:29

A well-known application of reinforcement learning is in video games For instance Google have developed

44:36

agents that learn to play Atari games or complex 3D environments games like Doom

44:42

These agents are not explicitly programmed with rules Instead they interact with the game receive scores or

44:49

penalties and gradually optimize their actions based on accumulated experiences

44:56

Although we won't explore reinforcement learning algorithms in depth in this course it is important to understand the

45:03

core concept learning through interactions and feedback to optimize outcomes

45:10

Now finally let's talk about self-supervised learning This is like

45:16

reinforcement learning a dense technical subject but it's important because it's part of large language models

45:24

Self-supervised learning is a kind of mix between supervised and unsupervised learning The tasks associated with this

45:31

kind of learning are what we call natural language processing tasks such as test summarization translation

45:38

question answering and sentiment analysis which is also a classification task task

45:45

This is the technology behind GPTs large language models Other tasks that use

45:52

self-supervised learning include image and video generation

46:01

Self-supervised learning has two main steps First take a large amount of

46:06

unsuper unsupervised data like images of cats and dogs without labels and

46:12

automatically and in the first step it automatically generates pseudo labels

46:18

You don't have that final column with the targets dogs and cats But the model

46:23

will create this labels for cat a pseudo label not necessarily cat but like

46:31

something different than the other something that is dogs In the end you're organizing the data so

46:38

that cats are closer to each other dogs are closer to each other and something like a helicopter is further away you

46:46

are creating better representations of unlabeled data That's the first step of

46:52

the algorithm Then in the second step you use

46:58

supervised learning to accomplish a task It could be regression classification

47:06

After organizing the data into better representations you use the pseudo labels to define a supervised data set

47:20

You can think of this same idea with words If you substitute the image examples with words the process is the

47:27

same Words like cat and dog will come closer together in their pseudo labels A

47:35

word like helicopter is far from them If you want to predict the next word as a

47:43

uh generative AI does or do translation or summarization you use this representation to do it

47:50

better So in the end the process is generate pseudo labels to get better

47:56

representations of data Use those features to build a supervised data set and accomplish a specific task

48:04

So we've now covered supervised unsupervised self-supervised and reinforcement learning methods

48:15

Now let's talk about the importance of effectively representing data This leads

48:20

us to the importance of having good features in your data Suppose you're

48:26

building a classifier or a regression model the features the attributes in

48:31

your data set or the columns must be relevant For example if you're building

48:36

an image uh an image classifier that detects whether an image contains a car and you

48:44

only use features like color and height the model will struggle

48:50

Many objects like cars and motorcycles can be the same height or color

48:56

So those features alone won't help the model distinguish them You need features

49:02

that are truly important This is where feature extraction comes in

49:08

In the car versus motorcycle example width might be a better feature because

49:14

motorcycles are typically narrower than cars

49:20

So your model will capture this more important aspect that is truly related

49:26

to the prediction of cars versus uh cars versus motorcycles

49:32

This there's an entire field of study dedicated to this and machine learning

49:38

protection use statistical or algorithmic meth methods to find the

49:43

most relevant features During data mining you can discover

49:48

which features matter most Having more data can also help reveal re

49:54

relationships that weren't previously obvious Algorithms can evaluate which

50:00

columns are more relevant to the task you're trying to accomplish In addition to feature extraction we

50:07

also talk about feature engineering This involves creating new features or

50:13

combining existing ones to make your model more effective

50:19

If you're tracking purchases for example and one of the columns is a date in the

50:25

format January 25th Friday you might want to split that into new features the

50:32

day of the week and whether it's a holiday So coming up with new new two new

50:40

columns one with the day of the week and the other if it is a holiday or not

50:49

These new features could be much more informative for your machine learning task Feature extraction and engineering

50:56

can make a real difference between a model that performs well and one that fails

51:03

This process is part of a broader step called data prep-processing

51:09

Before training a model you need to prepare your data properly Data

51:14

prep-processing includes doing data analytics

51:20

data mining feature engineering handling missing values removing

51:27

outliers filtering and transforming the data and

51:32

identifying important correlations

51:40

Outliers are data points that have values that

51:47

are too extreme and it comes in the way of your analysis

51:52

We will talk about outliers in the uh next weeks of the course

52:01

Feature engineering is especially important when working with unstructured data Structured data means each column

52:10

has a consistent meaning like name age income marital status

52:18

Each instance of the data set meaning each row fits

52:23

into that format Unstructured data

52:29

like a Twitter post or a book doesn't fit into neat columns It's still usable

52:36

right but you first need to engineer the structure you want

52:43

That might involve detecting uh uh uh structure uh uh summarizing

52:54

things classifying and then creating columns and other techniques This step is time

53:02

consuming and for sure challenging In businesses a lots of effort lots of

53:07

effort goes into this pre-processing phase Data analysts and scientists often

53:13

spend significant time preparing data before model training It's an iterative

53:19

detailed oriented process that relies heavily on expertise

53:24

Within pre-processing you'll often perform feature engineering feature

53:29

scaling dimensionality reduction data cleaning and data augmentation

53:37

Feature engineering we have already discussed Feature scaling ensures numerical features are on similar ranges

53:47

And we will talk more about feature scaling Dimensionality reduction simplifies

53:54

highdimensional data while retaining essential information

53:59

Just to clarify dimensions the dimension of a data set is the

54:04

number of features the data set has If you have a data set with two features

54:10

you have a two-dimensional data set In highdimensional data set means that

54:16

you have lots and lots of columns or attributes or features

54:22

and doing dimensionality reduction simplifies this highdimensional data set while

54:31

retaining essential information Data cleaning removes outliers

54:40

and correct inconsistencies and data augmentation might involve generating

54:45

synthetic data to supplement sparse data sets We'll explore synthetic data in

54:51

upcoming sessions Feature scaling and dimensionality redu reduction will also

54:56

be covered next week when we discuss model evaluation and factors that influence algorithms performance

55:06

Let's wrap up by discussing what happens after completing the data science cycle

55:11

Storing data cleaning data prep-processing the data feeding it up to the model Once the model is trained

55:19

the next step is deployment It is important to remember that multiple

55:25

models can solve the same task For example classification can be handled by random forests neural networks or

55:32

decision trees You'll need to select a model that performs well for your

55:39

specific task And in the end data science involves

55:45

lots of experimentation It is an experimentation science You may

55:50

train several different models and choose the best one based on how they perform

55:56

To do this we use evaluation metrics to compare models and

56:02

select the best one using a common test data set

56:09

We'll cover evaluation metrics in details for each type of task But let's

56:14

start with the basics ide with the basic idea of training and testing data

56:21

The learning process also known as training or fitting is the adjustment of

56:26

a model's parameter to best fit the data For example in a linear regression you

56:32

adjust the slope and intercept until the best fitting line is found

56:37

This is done using the training data which is typically past data

56:43

It may be labeled or unlabeled depending on your task The task data is a separate portion of

56:51

the data set that is held out specifically to evaluate the performance

56:57

of the model This data is not used during training

57:02

Holding out this data helps us assess how well the model generalizes to new

57:09

unseen data For example consider a data set used to predict whether it will rain

57:16

or not based on variables like humidity and pressure You might have several days

57:21

worth of data You can hold out some of it say 10% as test data You train your

57:29

model on the 90% of data you have from day 1 day 2 day three and so on and then

57:36

evaluate it using data from the 10% hold

57:41

out data You can evaluate several models in this

57:48

hold out data and because you know the answer you know the label or you know

57:57

what it should be you can then

58:02

assess which of the several models is better to your task

58:09

If you're using if you're performing classification for example a common

58:14

evaluation metric is accuracy which is the number of correct predictions

58:20

divided by the total number of predictions by comparing the model's

58:25

predictions on the test data to the actual outcomes that you have You can

58:30

calculate this metric This process allows you to evaluate and compare multiple models You might train

58:37

different neural networks a support vector machine and a decision tree then compare their performance on the same

58:44

test set using a consistent metric This testing process gives you insights into

58:50

how good each model is at generalizing The term generalization

58:56

is the term used to say that a model will apply its

59:03

knowledge to unseen data So this testing process gives you an

59:09

insight on how well the model will do that We'll explore the several metrics

59:15

and evaluation techniques in more detail in upcoming weeks To finish this first lecture here's a

59:23

summary of the AI building blocks So first of all we have to collect the data

59:30

then pre-process the data engineer or extract features

59:37

then train and evaluate models During training you adjust or fit the model by

59:45

finding the best parameters such as the slope or intercept in a linear model

59:51

After evaluation you select the best performing model and

59:56

deploy it for tasks like classification forecasting or summarization such as using chat GPT

1:00:04

or other language models Once deployed you monitor performance and repeat the

1:00:10

cycle You may continue collecting new data and retraining to improve the model

1:00:16

over time This creates a feedback loop that help continuously refine your

1:00:21

models