**Week 1 - Machine Learning Intro**

0:00  
This video is going to give a brief introduction into the topic of machine learning.

0:06  
So the first question is, what is machine learning?

0:09  
And the definition of machine learning goes back as far as 1959, when Arthur Samuel here said that it's a field of study that gives computers the ability to learn without being explicitly programmed.

0:21  
And there were several iterations of neural networks and machine learning algorithms over the year, But it's only really in the last couple of years where it's become fruitful.

0:30  
In 1999, Whitney and Frank then said that learning is changing behaviour in a way that makes performance better.

0:38  
In the future, what we're going to look at or when we look at machine learning, what we're going to look at it as algorithms that automatically improve performance through experience.

0:50  
And often this means defining a model by hand.

0:52  
That is, you hand craft or hand code a neural network or some other kind of variant of that, a support vector machine or even a regression algorithm.

1:02  
And then we tune the data to fit its parameters.

1:06  
I'm sorry, we use the data to tune the parameters.

1:10  
So with a neural network, the neural network at the start, all of the parameters that configure the neural network, the strength of the connection between the neurons, between the elements in the network.

1:20  
Similarly, the slope of the line in a, in a regression model, we need to try and figure out what is the slope, what are these connection values?

1:29  
You know, how, how much, what is the width of the connection of the pipe between neurons.

1:37  
So all of these parameters, we need data to fit that.

1:40  
And then we need an algorithm that'll use that data that'll assume the parameters.

1:44  
And it's the same thing that happens in our brain on a daily basis.

1:47  
If we want to learn a new task, say we want to learn how to throw and catch a ball or something like that, we have to train it, give it lots and lots of examples of looking at doing it over and over again or learning the difference between two types of dogs.

1:59  
We need to see a few examples of each before we can, before our brain and the algorithm in our brain can tune the neurons that are there so that we can, so that we can recognise in the future.

2:13  
There are lots of problems that are difficult for humans but easy for computers.

2:16  
Computers have been around for quite a long time, you know, 40-50 years where we've, they've been able to outperform humans at these problems, close to 100 years probably.

2:25  
So calculating large arithmetic problems, probably going back as far as some of the work that Turing did on the Enigma code.

2:34  
But then there's a whole other set of problems that are very easy for humans but very, very difficult for computers.

2:40  
Say recognising a picture of a person from the side or detecting a face or something like that, to try and feed all that data in and form the pixels and to write calls.

2:51  
Let's say.

2:51  
If you wanted to write it in terms of if statements and for loops and functions, it'd be very, very difficult to write that software for a human to write that software to be able to from a set of pixels, recognise a person's face within the image.

3:04  
Like there would be if you tried to do it really just with if statements.

3:07  
You have to write so many if conditions, it'd just be impossible.

3:12  
Essentially machine learning then tries to leverage the things that computers are good at, so large arithmetic and repetitive problems to solve things that we are usually able to, that humans are usually able to do more easily.

3:26  
So we're still inside a neural network.

3:28  
There still is essentially only for loops and if statements and all of these things and multiplies and odds, but it's arranged in such a way that the model and the parameters are able to generalise and to do things that humans have typically been good at and machines haven't been good at.

3:45  
And we'll see examples of this over the next couple of slides and over the next few weeks indeed.

3:54  
So why do we want to give machines the ability to learn?

4:00  
So the real world is complex and it's difficult to handcraft solutions.

4:04  
So as I said on the previous slide, if you think how many if statements would be needed in just a traditional kind of procedural programming or traditional programming to detect a face or to detect, you know, the difference between an image with a cat or a dog and I should be able to discriminate between the two.

4:19  
It would just be really tricky and especially if you had to try and write this code yourself.

4:23  
So we need to be able to automate this to get the computers to learn rather than us having to teach the computers.

4:29  
Machine learning is the preferred, preferred framework for applications in many fields now.

4:33  
So computer vision, like I said, computer vision is very, very difficult traditionally for machines to do.

4:39  
They're getting better at it, in fact, much, much better.

4:41  
But it's something that humans naturally are able to do from a very young age.

4:45  
Natural language processing, again, we can say the same thing in many, many different ways.

4:49  
And again, it would be almost impossible to write a set of if statements to check all the different permutations and combinations in the way a person can say something very simple to actually be able to capture that and understand it.

5:02  
So again, neural networks and machine learning algorithms are being used more in this area.

5:06  
Speech recognition for similar reasons.

5:08  
Robotics, again, for similar reasons, because again, they will encounter so many different scenarios of, you know, if it's moving through even the lawn cutting robot robots that I have, you have to write an if statement for every single sensor combination it would encounter.

5:22  
It would again be next to impossible.

5:24  
While if you can get them to learn by themselves using machine learning, then it's much more efficient.

5:30  
Humans in within the terms of natural language processing.

5:33  
Humans can typically parse sentences, but computers in the traditional way of quoting them do not get the context.

5:41  
Humans are not consistent with grammar, not consistent with word order.

5:45  
Use different synonyms, such words to replace each other.

5:49  
While computers in the kind of traditional way of calling them, expect consistency, that it wants something to be consistent every time.

5:55  
And again, this is another motivation for using machine learning to be able to generalise.

5:59  
And we'll come to this concept of generalisation soon.

6:04  
The basic idea of all machine learning.

6:07  
So machine learning is essentially using an algorithm and the algorithm can learn and tune the parameters of a model from the data that's given to us.

6:17  
So we want to create a model which is say the bare bones neural network or support vector machine or all the various models that are out there that takes input.

6:26  
So we feed input into the sensory neurons, say on the neural network and gives output.

6:32  
So it can give an output on the output neurons.

6:34  
So we have this in our own neural networks, or we've got senses and we've got outputs, which are our motor neurons essentially, which give us the ability to speak, to move, to like move different parts of our body and so on.

6:51  
So this equation captures it where Y is the output, X are the inputs, and F is some sort of function that's performed on the inputs in order to give us the output.

7:02  
So this will make a bit more sense.

7:04  
This is like the very beginning of the maths in this course is if you can, we're going to try and most machine learning algorithms boil down to this.

7:12  
The tricky thing here is what is the function that transforms the input, say a picture of a face or something like that into an output where it says, yes, that is a face, or no, that isn't a face, there's no face present.

7:25  
So what is the function?

7:26  
What is?

7:27  
And that's where the neural network operates or whatever machine learning algorithm it is, operates on the input data in order to produce some output.

7:35  
At the end we figure out the right hand side of the equation.

7:39  
So we need to figure out what is the model, what is the F of what are the parameters of the model.

7:43  
By gathering lots of data and making them the parameters that we tune best fish our model so best predicts the output.

7:51  
A very basic example.

7:54  
So this is FX.

7:56  
This function here can be really, really complicated, but in the simplest case, it can just be a very simple function, which is this right here, which some of you might recognise as the general form of the OR this right here is the general form of the equation of the line Y is equal to MX plus C.

8:12  
So a lot of you would have encountered this or I remember encountered it in junior circ maths where Y is the output or the Y value, let's say.

8:21  
And M is the slope, X are the X values and C is some sort of consonant also known as the Y intercept.

8:28  
So the this equation of the line, the general form of the equation of the line describes all lines in 2D space or X&Y and a slope M&C.

8:39  
So this this is a model here, which is something like Y is equal to FX.

8:43  
But the FX, the function which we manipulate is this parameter W0 and W1.

8:49  
And we need to figure out what's the best W0 and W1 so that our model, which is this line fits the data that we have.

8:58  
So this is the data.

8:59  
If I just look at what it is here, it's a number of friends say on a social media platform, be it, I don't know, WhatsApp or Telegram or Facebook or whatever it is Instagram versus the daily minutes that people spend online on that platform.

9:13  
So it wouldn't intuitively make sense that the more friends that somebody has on Facebook, let's say, the more time that they spend scrolling to see what all the friends are at or something along those lines or on any of the app platforms like WhatsApp or Twitter or whatever.

9:30  
So the here, what we're trying to do is find what's the Y intercept here and what's the slope of this line to best fit this data so that we can make predictions about if someone said to us that someone has 40 friends, how much time do you anticipate that person spends online?

9:48  
Now we can see here that there's noise around this model, that it's not actually a straight line.

9:52  
It's a good fish, but it's not a perfect fish that some of the real examples here don't fit on the line.

9:58  
But it's an estimation of a kind of a generalisation of the data.

10:03  
Of course, our base model can be much more complicated, have lots more parameters than just these two parameters.

10:09  
W0 and W1 are in the kind of algebraic format of M and CM for slope and C for Y intercept.

10:18  
But the basic idea is the same.

10:19  
We might have lots and lots of parameters here, you know, for the big deep learning models like ChatGPT and stuff like that.

10:26  
There's billions of parameters that make up this model and we need to tune figure out what are these billions of parameters that will give us some model that models the data that we have.

10:38  
After we have created the model, we can then make predictions in the future.

10:43  
So above, how many minutes would we predict a person with 30 friends in the platform would spend spend time on the platform?

10:49  
So we look at 30, we look across, you know, we could say roughly between, you know, maybe 4850 minutes for someone with 30 frames.

10:56  
That's what we would expect.

10:59  
That would be our prediction based on the data that we have.

11:03  
Why does it take so long?

11:04  
So a model like this would train very, very quickly because we're only trying to seek out two parameters.

11:10  
So one in the above example, which is a basic line, is the model.

11:15  
Learning all the parameters is very, very quick, almost instantly on a computer, but the real world is not that quick.

11:21  
So machine learning models can take hours, days, or even weeks or months to be trained.

11:26  
But this wouldn't be unusual at all.

11:29  
Chap GPT is actually, as I said, eight different models, each with 220 billion parameters.

11:37  
So you can multiply this by 8 to get like the trillions of parameters that make up GBT and to solve each one of those parameters and get good values for each.

11:46  
So that the we get the output that we do takes a very, very long time to figure out.

11:51  
Now the human brain is amazing and that we can you know, when you see a, a baby growing into an infant and toddler and so on.

11:58  
And you know that you can see the tuning happening almost as they're as they're developing.

12:03  
But like to see, you know, it's incredible how fast that is.

12:06  
But the computation resource it takes to tune chat PGPT over months and months and months to be as smart as it actually is.

12:14  
It's a lot of processing power and it takes a long time.

12:18  
This training is repeatedly doing the same procedure over and over again.

12:21  
So just feeding in the same data over and over again and just tuning the parameters.

12:26  
And at the at the very beginning, you're tuning the parameters, making quite big changes until it's starting to look it'll be just totally garbage that's coming out at the very beginning.

12:33  
But as you get better and better and better, maybe the tuning and the tweaks of the parameters you're going to make is smaller and smaller.

12:38  
Or sometimes you might need to make big changes in certain areas of the model or of the network until we get the parameters we think are best.

12:45  
SO no one what what are the best parameters on when to stop is a tricky thing.

12:50  
We've got this kind of concept of like how much error errors are in the model.

12:56  
So, you know, we can see here that this still is error in this model, that this prediction isn't really corresponding to the real world value.

13:03  
But you know, it's hard to model exactly this with a straight line, with just one straight line.

13:07  
You know, you might need something more complex to do a more complex modelling of whatever is happening here and maybe there's more factors to be accounted for rather than just a number of friends that predict how much minutes someone will spend on a social media platform.

13:25  
There's probably lots of other factors that I can just think of off the top of my head as in like that person's, you know, work schedule or how many real world connections they have.

13:36  
All of these things will actually impact that.

13:38  
So maybe one single input isn't enough to really predict this.

13:44  
We'd need something a bit more complex and that will introduce more parameters in the model.

13:50  
So this is your machine learning system.

13:52  
Yeah, pour the data in this big pile of linear algebra.

13:55  
Just talk about this word linear algebra for a moment.

13:58  
Linear algebra is the maths not even in secondary school, but that we learn in primary school.

14:03  
It's when we learn our plus tables and we learn our multiply tables, and linear algebra is essentially just doing adds and multiplies.

14:11  
And the reason that all these machine learning algorithms, modern ones, only use adds and multiplies and essentially only use linear algebra, which is a whole field, very big field of mathematics, but that only relies on those two operations.

14:24  
The reason they do that is because computers are really, really good at doing adds and multiplies.

14:31  
People, sorry for interrupt.

14:33  
Then collect the answers on the other side.

14:34  
What if the answers are wrong?

14:35  
You just stir the pile until they start looking right?

14:37  
So you just chew the parameters, keep tuning them, keep stirring it, And then once they start looking right, you need to know what right is, of course, and you need to have some sort of a measure of a way of measuring what's right and what's not right.

14:50  
Most of the data we want, if we want to train a neural network or train a model, it needs to be labelled.

15:01  
And it's the same thing if you're trying to train, say, a radiographer or like an oncologist, someone who reads, you know, MRI scans or, you know, maybe is looking at mammograms or something like that.

15:14  
And they want to discriminate whether a tumour is malignant or benign.

15:18  
If you sent them into the job without ever seeing a mammogram before, they wouldn't do a very good job.

15:23  
So what you need to do is you don't really need to show them lots and lots of mammograms.

15:27  
You need to also have an expert there who's telling them, OK, this is a benign tumour and this is a malignant tumour.

15:35  
And we're not only going to show you one of each, we're going to go through hundreds and hundreds of examples of what a malignant tumour looks like, what a benign tumour looks like.

15:44  
And that with that training with an expert labels data.

15:48  
Essentially you're looking at the output of the scan and you're looking at whether it's malignant or benign.

15:54  
That data is what you need to do to train a system, train a person at any particular task.

16:02  
So any data we put into the system to train it is a pair of information X, which is the features of what you're trying to train it on.

16:10  
Say it's a picture of like a medical scan and Y is the output of that medical scan, whether it's a healthy or unhealthy sample or Y is the correct answer for each.

16:21  
\*\* then is the collection of all the features, but they could be pixels.

16:26  
I'm trying to think in other example areas that have worked in over the years.

16:29  
There could be other features, you know, have worked quite a bit in medical diagnostics.

16:33  
So in cancer imaging, you know, they could be like the, you know, there could be blood results going into it.

16:42  
There could be various different other features, the person's age, the person's BMI, the person's previous medical history, etcetera.

16:53  
So let's say we have a collection of images that are cats or dogs and we want to build a model that can separate them.

16:58  
For a particular image, there's a correct Y, which is, is the image a cash or a dog?

17:04  
An X is all the features of the image.

17:06  
So all the features of an image are generally just the pixel values.

17:09  
All of the pixel values in that image feed it into your model.

17:13  
Train up your model with the correct answers that you already have.

17:16  
And then in the future, hopefully you'll be able to generalise, to be able to recognise cats and dogs that's never seen before and be able to discriminate them if you feed in some new data that it's never seen or never been trained on.

17:28  
The idea is same with someone like a radiographer, that if they see a new specimen that they've never seen before, they'll be able to make a prediction based on the kind of general knowledge about the topic, their generalised knowledge.

17:42  
So just as a look at a kind of an example, you know, not cats and dogs, something that's still quite hard to do, but not at the same level.

17:50  
We, the human intelligence must do a lot of work here.

17:52  
We just don't know we're doing the work.

17:54  
But when we look at all these characters, and I sometimes have to do this and I'm correcting the exam scripts, it can be a total nightmare to figure out what someone has actually written.

18:02  
So in order for us to be able to read handwriting, and again, you see it with children that it takes lots and lots and lots of training data and labels associated with them.

18:12  
And you see, I think it's pharmacists I think are the best at deciphering scribbly scrolling, writing in the world because I've seen so much of it coming in from prescriptions and all of that, that they've become really, really good.

18:24  
So the example is the endless data set.

18:26  
This is the data set right here.

18:28  
To learn to recognise handwritten letters, the algorithm must be trained on 60,000 samples, each of which must be labelled correctly by hand.

18:35  
So someone has to go through this, look at this and say, OK, that's a zero, that's a one, that's a 2.

18:40  
This thing here is an H and that's six and not AB and so on.

18:45  
And this is just a sample of the data set from Wikipedia.

18:50  
But there's a whole load of samples, 60,000 samples that you train on.

18:54  
And then there's a whole other set of samples which you hold back to test how good the algorithm is on data that it's never seen before.

19:03  
Again, this 4 here, you know, you could be looking at an A or something like that, so there's plenty of room for confusion.

19:09  
Sometimes you need to be aware the more advanced algorithms aren't only looking at individual characters, but they're looking at the overall context of a word to see if that four crops up within other numbers.

19:20  
And a kind of a telephone number probably is a four and not an A.

19:24  
Well, if something like that comes up in the word, maybe like if there's AC on one side, an or E for care, then dash is probably an A and not the number 4.

19:34  
So there's all this kind of context being built into the more advanced algorithms.

19:39  
Here again, we are looking at handwritten digit recognition.

19:42  
So MNIST I think is mostly looking at it's not actually.

19:45  
There is a data set which is alphanumeric, not just numeric.

19:49  
This data set is a numeric data set.

19:51  
So it's a bit easier.

19:52  
You're only trying to discriminate 10 different symbols or with alphanumeric there's a whole load of more complexity.

20:01  
So there's lots if we were to try and do this again with if statements, there's it's very difficult to handcraft rules about digits, you know, if you're trying to say, OK, there's a straight line and a little curve and.

20:12  
If you're trying to do this with if statements would be impossible, you have to hand it over essentially to the machine to learn.

20:19  
So as an example here of our inputs and our outputs, this is typically how it would look like in a machine learning algorithm is our XI is our input.

20:26  
So these are all the pixels here that make up this number four.

20:30  
We can see it's a four because we're pretty good at this.

20:32  
And then the Yi is the output, and you can see here 0123 and four.

20:38  
So this output here is that whichever 1 it thinks it is, it'll put a one on that symbol.

20:44  
So if it's a zero, it'll put A1 here and zeros everywhere else.

20:49  
And similarly if it's a nine, it'll put A1 here and zeros everywhere else.

20:55  
So the image can be represented by the values of each pixel.

20:57  
This is the XI, all the inputs, each pixel colour is called a feature and they're independent variables, meaning that each pixel value, if it's just black and white, it's a zero or one.

21:09  
In this case, I'm not sure if it's grayscale, say if it's just zero and one, black and white are independent from each other.

21:14  
As in any of these pixels, if they go to, they come from white to black.

21:18  
It doesn't influence any of the other pixels, they're all independent.

21:21  
We represent this input image as a vector.

21:24  
So this input image has actually got 784 pixels in it.

21:27  
So this XI is 784 real values.

21:30  
So I'm just going to try and describe this mathematical notation as we go along.

21:34  
So we've got 784 real values in our input vector X.

21:39  
This kind of example here that we have is supervised learning where we have an input and we have the correct output already annotated.

21:49  
It's a discrete finite label set that's zero to 9.

21:52  
So we've got a classification problem.

21:54  
And if you notice, these aren't continuous.

21:56  
It doesn't go from zero to 9 and all the numbers in between, it's either A0 or A1 or A2.

22:01  
And that means discrete means you know no, nothing in between zero and one, it's either 01 and nothing between 1:00 and 2:00.

22:09  
It's just numbers 1 to 9 or 0 to 9.

22:12  
Giving a training session, which we have here, which is the set of all of our inputs from X1 all the way up to XN where we've got N examples.

22:22  
The learning problem is to construct a good model or a good function F of X, whatever form that looks like loads of parameters essentially inside a neural network would be one way of doing it.

22:32  
We could do is using lots of different types of models as well from these data that we're putting in.

22:37  
So essentially our function is trying to map a 784 pixel vector of real values.

22:44  
That's this right here into a a vector which has got 10 elements in it of real values.

22:51  
So these are our 10 elements here.

22:53  
Now at the moment these are actually holding integers, zero and one, but most algorithms won't actually return just 001, you know, where the one is, the 00123 or so on, but will instead return probabilities.

23:06  
So it'll say the probability of that number here being a zero is 0.01 as in kind of 1%, likely 3%, you know, 0.040.010.88.

23:16  
So it's .88 confidence that this number here is 4.

23:22  
Sometimes it'll be a bit undecided.

23:23  
It could have .6, you know, for two values, and then you have to make a bit of a decision or go with the biggest one or whatever it is.

23:31  
So the image is most likely A4 rather than definitely a four.

23:34  
It's 88% confident that it's a four, essentially.

23:40  
Looking at another example of face detection, it's another classification problem where again, you've got discrete values of 0, not a face, one face and two.

23:51  
Oh, yeah, sorry, I forgot to say in the last slide, this target vector, this target vector is essentially the output here.

23:57  
What's our target?

23:59  
So we've got it's either A0 or A1 face or not face.

24:03  
It's not, I mean it is that particular number or it's not here, our target vector for every single image or region of an image, even for every one of these bounding boxes.

24:16  
In fact the target vector would either be A0 or A1 or A2 which is a non face face from the front which is here denoted by a square or rectangular bounding box and a profile face which has got a kind of a triangle or an angle at the edge.

24:34  
But The thing is this target vector could be expanded into the face of a particular person.

24:38  
So the possible TI set could be very large.

24:41  
Here it's just saying the, the target vector is 01 or two for any face.

24:50  
But imagine here if you're looking for particular faces, if you're looking for the face of say someone of interest to the police or a criminal, you know, you don't want, you want actually to just flag that particular person rather than just any face.

25:03  
This could be, you know, all the potential faces that are out there in terms of everyone on the planet that you've got like 0 for me and one for somebody else and two for somebody else.

25:13  
And you know, 10 is the person they're interested in and the only one.

25:16  
So TI is a set of all possible faces that are out there.

25:20  
We map, track.

25:21  
We, what we're doing is we're mapping or transforming this image that we have here to a particular set of target vectors.

25:28  
So in this case that which was done in 2002, it was pretty revolutionary at the time is it was able to detect faces.

25:35  
Now, since then, some of the newest algorithms again can detect specific faces and they're being rolled out across the world and surveillance and you know, certain governments are very interested in knowing where their citizens are at all times.

25:53  
Stock price prediction is another example.

25:55  
So problems in which the TI is continuous are called regression.

25:58  
Again here, it's either not a face or a face or a side profile of a face.

26:04  
There's nothing in between.

26:04  
You can't have something between A0 and A1 of like a bit of a face or a kind of a face or maybe the face of an animal or whatever.

26:11  
We're just looking at discrete outputs here that we want the neural network essentially to give us.

26:16  
But you can have examples where the target vector is actually continuous.

26:20  
And what we're trying to estimate, these are predict rather based upon some inputs.

26:26  
These problems are called regression problems.

26:28  
So we're looking here at say stock price.

26:32  
So XI, the inputs which are trying to make the prediction of the output based on is the company profit is something that might determine stock price, the level of debt, cash flow, gross sales, number of spam emails sent, etcetera, etcetera, etcetera.

26:51  
So there's lots and lots of things and like the global market, all of these features, the XI, the inputs will somehow predict the stock price.

27:00  
No one has got this absolutely right otherwise to be very, very rich or as some of the companies may have good insights into how to predict, you know, whether stock price is going up or going down.

27:10  
But there's a whole load of features.

27:11  
Again, you'd have to build a model to the parameters of those models to be able to predict what the stock price of the company is going to look like given these inputs.

27:19  
And this is essentially a regression where you're trying to predict, in this case over time, what the stock price would look like.

27:27  
There's so many examples of machine learning being used in the world.

27:30  
All the big companies use it.

27:32  
Like I said, I spent a lot of my career working on IT and medical device imaging.

27:37  
You know, there's all sorts of segmentation algorithms now and diagnostics that can actually look at particular regions of interest and depending on the shape of them and depending on the characteristics, you know, make predictions about diagnosis.

27:50  
Similarly, you spent a lot of time working in ECGE e.g.

27:53  
signals.

27:54  
You were looking at an ECG.

27:55  
You can also look at brain signals, make predictions, you know, automatic prediction algorithms.

27:59  
Now they're working to tell whether a user has got certain arrhythmias or certain other heart conditions.

28:06  
Surveillance is a huge one in the region of computer vision.

28:09  
Our shops, you know, if you have a loyalty card with a big company or Amazon, I see here are really, really interested in making selling you more stuff.

28:18  
So the more predictions they can make based upon what they know about you, the better they can do as a business.

28:24  
Netflix, we look at an example where it uses a lot of data that it has on, you know, what bits, you know, what do people like to watch?

28:32  
What are the patterns of life for particular users?

28:35  
What kind of stuff do they watch and what do they like?

28:37  
What time do they stop watching a particular episode when they want to eventually go to bed?

28:41  
Or what are the kind of moments that keep people, you know, engaged and hooked into the platform so they can make all sorts of decisions.

28:49  
And we're looking at the example Microsoft, Yahoo, Facebook for selling your stuff, selling ads to you.

28:53  
Google similarly, it's been used in this really nice application of alpha folds, autonomous vehicles, computer games.

29:01  
It's a huge exploding area.

29:05  
And if you look at some of the high profile examples I've cropped up over the years, see some of these aren't the most recent, but I do remember them coming up when people took note and were like, wow, this is something big.

29:16  
So there was an article back in 2012 for Target, which is a big shop in the US, figured out a teen girl was pregnant before her father did based upon just certain things that she, that the, that she started, she was maybe using his loyalty card and based on the food that she started buying.

29:35  
They were able to predict really early on that she, because they've seen so many examples of people going in and various, you know, shopping habits change, let's say during pregnancy or something like that.

29:46  
There's so much data on so many people that they can make, they can see these signals.

29:52  
And they started sending coupons to to the girl's father for, you know, pregnancy and baby related stuff.

30:03  
And the father got on to target saying we like, why are you sending me this?

30:06  
You know, this is not at all relevant to me.

30:08  
But then it turned out it actually maybe was relevant.

30:11  
Netflix then is turning viewers into puppets.

30:15  
This was with House of Cards.

30:16  
Remember seeing this story again, it's a few years back.

30:19  
But what they had to say, it's so much data on what people liked, what kind of storylines, what thematic stuff that people liked, that they essentially put all that, all that data into an algorithm and kind of out popped at the other end.

30:32  
The series House of Cards, which is not might not be to everyone's cup of tea, but it did predict that this for, you know, a sizeable or majority of viewers will be giving people exactly what they want in a, in a, in a series.

30:47  
Another example from 2012, which is kind of the beginning of deep learning for object recognition.

30:50  
And this has huge impacts on the field of computer vision is they just, they develop an algorithm which if you feed an image into it, it can make a prediction, tell you what's in.

31:01  
So there's a what's in the image.

31:03  
So there's a huge amount of training that goes into these things, annotated images where you know everything in this image.

31:08  
And when you're off and filling out Captchas online and you're like select all the squares where there's a traffic light or select all the squares where there's a bicycle in this image.

31:16  
Essentially, you're building these data sets for the companies that they can use to train their machine learning algorithms because it's very, very expensive and boring and tedious to actually sit down somebody at a computer And people do this as a job all day long and be like, that's a mice and that's container ship and that's maybe some a city and there's a scooter and there's, you know, people in cars and a leopard and whatever Madagascar.

31:40  
So some of people do this, but most often now they're trying to get people to do it kind of covertly or whatever to train these things.

31:48  
And then it's amazing, you know, how the performance of these algorithms know.

31:53  
Now the good example is Go.

31:55  
So alpha Go, which is part of the Google DeepMind, and they also have a chess implementation, learn to play Go.

32:01  
And starting from completely random play, it became the best player in the world, which wasn't expected to happen.

32:06  
You know, this happened maybe close to, I don't know, 7-8, ten years ago or something like that.

32:11  
It wasn't expected to happen that a computer could beat a human at Go for another 30 or 40 years at the time.

32:17  
And now this system can learn chess from never having played before, nobody teaching it how to play, just sitting down, playing against itself.

32:23  
It can become from scratch the best chess player in the world within 48 hours.

32:29  
So the kind of competitiveness in terms of how good these machines are and how competitive they are with humans is, is exploding.

32:36  
Art is another thing.

32:37  
This is a story from a couple of years back where it was the very first artwork created by artificial intelligence or Christie's was the first auction house rather to sell art made by artificial intelligence.

32:49  
And people thought it was a big deal at the time.

32:51  
But now people can create with the likes of Dali and Mid Journey and tools like this art whenever they want and create any type of art.

33:00  
So yeah, looking at these high profile examples of the latest stuff, ChatGPT, the kind of Google equivalent, Mid Journey, and Dali being amazing image generators.

33:09  
And as I said, this field is exploding.

33:12  
So what I'm going to do, and I'm going to go back for a minute and look a little bit more at some of the kind of basics of machine learning.

33:18  
So looking at the different types of machine learning, it's often divided into supervised and unsupervised machine learning.

33:24  
And most of this course we're going to be focused on supervised learning because it's probably the most common that's out there at the moment.

33:30  
And it's techniques where we have training examples where we know the correct results already.

33:34  
And that's the way we essentially train, you know, we train ourselves or we learn is, you know, we're taught and we learn how to read.

33:43  
This is the letter B and that's the letter D.

33:45  
And we learn the kind of difference between them.

33:47  
One's facing this way and one's facing that way, you know, lower case D and lower case B.

33:51  
And it takes us a while, like everyone gets that wrong at the beginning.

33:53  
And then we try it lots of times and eventually we learn.

33:56  
So then we take it for granted.

33:57  
So everything, no matter what we learn, usually we're given examples by a teacher or someone, you know, a book or something like that, that knows the correct result.

34:05  
And then we learn.

34:07  
There's different types of supervised learning, as I said, classification where the target effector is discrete, you know, it's either a face or not a face, and regression where the target effector is some sort of continuous variable.

34:19  
Different algorithms that are out there, there's regression, different types of regression, Bayes K, nearest neighbours, SVMS, neural networks, decision trees around the forest, and we'll look at some of these during the course.

34:29  
The other type then is unsupervised learning where there's no right answer known a priori meaning in advance, but the algorithm given just a heap of data, tries to find some structure patterns in the data.

34:42  
So it's really, really costly to annotate data.

34:44  
Like all the big companies even in Galway, like Valeo and all that, have a whole team of people that they pay full time just to sit down in front of a screen and be just like, that's a traffic light and that's a person and they're drawing boxes around it.

34:55  
But if you don't have that resource or you don't have data that's annotated, you could just throw a big bundle or bucket of data into these algorithms.

35:03  
Principal control pond analysis came its clustering page rank, where again, there's no one who's working for Google who's like discriminating, you know, or saying what kind of page each is.

35:12  
It's kind of just all going in and then it's clustering.

35:15  
It's doing its own work to try and figure out what are the patterns here.

35:19  
All supervised learning techniques will group things together that thinks are similar, but we don't necessarily know what the groups are.

35:25  
It often takes a human to go back and be like, ah, that's a group of, you know, it's seen a lot of things that they're all cats, you know, if we haven't told it what a cat is, but it's put them all together in the one category.

35:36  
We'll do less on supervised learning in this module.

35:39  
And we do some, we do a lot of maybe when we're learning, we do a lot of passive on supervised learning.

35:44  
We often ask for validation.

35:46  
You know, we're like, what was that thing I just saw?

35:49  
Was that a boss?

35:50  
And they're like, no, that's a Ben Lorry or whatever it is.

35:52  
You know that this is the kind of, you know, so that is supervised.

35:56  
But sometimes we're learning where it's unsupervised, where we're just taking in and we're making kind of associations of, you know, what groups of things are.

36:05  
There's other types outside of the scope of supervising unsupervised algorithms, namely just reinforcement learning.

36:13  
But we're not going to really cover this cover that in this module.

36:16  
The two main ones are supervised and unsupervised, and of those two, supervised is the most popular.

36:22  
Where you've got a expert teacher, these different types is with supervised learning.

36:27  
We've got classification, which is discriminating between two discrete sets, and you're trying to draw a line between them, saying is it malignant or benign, or is that a horse or a donkey or whatever it is?

36:38  
Regression then is continuous and we're trying to make a model that fits a whole set of data.

36:43  
Clustering then is where you don't know what any of these things are in advance and you're just trying to put all the things you think are similar together.

36:50  
And then association, which is sometimes kind of what's the dimensionality reduction is it's looking for connections and linkages between variables.

36:59  
Are variables kind of correlated with each other?

37:01  
Are they associated with each other?

37:03  
And could we, you know, do with less XI, less inputs rather than feeding in millions and billions of inputs into the algorithm, could we somehow like reduce that down a bit so that we've less parameters to tune?

37:15  
Essentially, could could we make our job easier?

37:17  
That's kind of dimensionality reduction, which we might do a little bit on.

37:20  
I'm not sure if we do so much of that.

37:23  
Machine learning overall is often referred to as narrow AI.

37:28  
The difference being as opposed to broad AI, which doesn't exist yet, that we're only concerned with a specific task.

37:34  
So the goal of machine learning is not to develop general intelligence, kind of like what we are as humans that can do really broad range of tasks and it's not far away.

37:44  
Maybe in machine learning.

37:47  
It's what it's rather doing is it's seeking an algorithm that learns a very specific or particular task well and will result in a system that probably or hopefully carries out the task correctly on most occasions.

38:01  
So if we want to do machine learning, what do we need?

38:04  
So we need to define or limit the task somehow.

38:06  
We just can't say be figure out the world for me, you know, we need to say, I want you to discriminate between X&Y or make predictions based on like how much time people will spend online or how much time we'll see some other examples you know of through the course of what it is the tasks can be, make a prediction about like an image or something, the experience.

38:27  
This is some correct examples.

38:29  
And the more data that you have, the better examples of wash is the right answer to this.

38:35  
The performance measure.

38:36  
How do you know, how do you measure how good the machine is at doing it's job?

38:41  
And it's difficult for the machine to learn if it's only right or wrong.

38:44  
It's like, am IA little bit right, even am I?

38:47  
You know, that's for us.

38:48  
We get that kind of feedback when we're learning.

38:50  
It's like, ah, was that like wrong or right?

38:53  
Like, you know, where do I go?

38:54  
What's wrong?

38:55  
Where do I go from there?

38:56  
But if you're given more constructive kind of feedback as in, ah, you are a little bit closer to it and maybe that was better than your last attempt, at least you kind of know the direction you need to go in order to get better.

39:06  
The learning algorithm then is the recipe by which we improve the performance and chew the parameters.

39:12  
How do we chew these parameters?

39:13  
Do we just start wiggling them randomly and hoping for the best?

39:16  
Or is there something like directed way of guiding these parameters so that we get closer and closer to our solution all the time?

39:23  
And we'll look at some of those methods.

39:24  
And then the intelligence, the actual model itself, that's kind of that that is to be tuned the brain in the case of the neural network and lots of other kind of linear algebra arrangements, it all just points boils down to linear algebra.

39:37  
I don't really differentiate in my head between SVMS and neural networks and all these things.

39:41  
They're a bit different, but essentially they all kind of use.

39:45  
They're using linear algebra because that's what computers are good at.