Imagine being able to predict stock prices. That would be wonderful. And we would be then super-rich inside. We have an AI-powered algorithmic trading program that trades in the stock market. More specifically, any stock price prediction has to adapt to machine learning techniques for time-dependent data, which has been called a time series.

To understand this more in detail. Data was always set for us. It is a set of examples, an ordered set x1 x2 x3 x and our examples we say are x, x y y is our example pair input-output pair, and we have examples run from 1 to n. But there is no dependency on the pipe, like what, 2 is after 1 type of notion we never had. But several data that we work with also have some amount of dependency. There is a sequence of nature to it. Like the price of a commodity that I buy has something to do with the price? In the last time, little instant time, the instant of last day or year or month or whatever baby the unit you take. This space has that hour; it is not a set, but a sequence ordered set in a way so then x1 and y2 and y2 x1, x1x2 and x3 x4, etc. So, most of the time, the x is a type but may not be the time. The example that I took about a sentence access door time. While I did talk about the price of a commodity, or some stock price, the commodity, the axis is usually time-series data has time axis, and how we predict the problem is almost similar. So you have several packets sold population of the certain country across the years. From the learning problem point of view, what we are interested in predicting what will be the next, for example, the population in 1991 will be the given this data, can we say something about this. So this is the type of problem that we address. Let us see how do we formulate this. So this problem we also say there's a time series. So the signal processing community looked at this problem extensively to formulate it. The Statistics community looked at this problem extensively as time series is a big deal of this time axis. People also call this a time series and many popular applications in meteorology finance, marketing, and many other areas. So now that two problems or situations, which we'll go one by one. The first is quite straightforward to see; there is a price that is x1 x2 x3 x and n supplies available to us. This is one model of more parameters, so this is just an observation. It could be one parameter, it could be a scalar, or it could be a vector; think of it, for the time being, it is a price of commando available every day for many years. And we are interested in building a model that will allow us to predict new or the past better. The second is more complicated because we believe that rain influences agriculture or the economy. And that influences the stock price. So it's not obvious how the rain is changing the stock price. We believe that we don't have a parameter, but they plan on having influence. So, but this is also a mechanism that can also be modeled as x influences x rain can, I can measure. So they influence us at Etsy some hidden parameter, some big like an economy of a state of the economy, what are some things like in it, which is usually a hidden state, and the edge influences Why. Why else I can measure up. Why is the stock price? So we, we can, so this is a second model. So, in this case, we observe x one y one, x two y two x three y three x and y, etc. So we know the line, how much happened we know the stock price. Still, we don't know whether there is a relationship between the till now what we have been doing is, we would like to predict y, y1 on x1. We want to predict y3 given x3, and we'd like to predict y given x, which is how we have been formulated, along with we assume that there is a time dependency and the time dependency could be exploited in this entire process. That's why I told you, now it is a sequence, not as a set. We were randomly sampling. An example like when you talk about Stochastic gradient descent. We are always; you have a pool of examples. We take an example from it. So this model doesn't work well for us. At this stage, because there is a time dependence. How do you break that? So we want to break this type of dependency and put it back into our neural network how to do. So let us assume, let's start with a simple problem. We have a sequence x0 x1 x2. So the price of a commodity x p, where t is a time index of the sequence most of the times in this sort of notation, people use a time t as an index variable because of the condition called time. So we made an assumption. When we are assuming at this stage, it's a strong assumption. But believing all the previous patterns also use it. That, that the measurement of the period depends on the three previous ones. So now we can convert this into a nice representation. Our role data was a time pone x one p two x two t three x three t four x 45 x 46. So this is how our data was. We relied upon this data, such that, given x one x two x three, we would like to predict y x for given x two x three x four we would like to predict x y given x three y x four x five, we would like to predict x six given x four x five x six, we would like to predict sir. Nothing new. We just rewrote it, so now this becomes very nice for us because we can go back and say that our look, this is your x, is your y in the traditional setting for a neural network or machine learning algorithm. This is a feature representation. The three features. This we would like to predict. Our problem is now reduced to the same; we'll want to learn a function, such that, given these three measurements predictors.

Now, this sense of from the seven we got four of them. Now, if you have 100 of them, you will have too many more than three of them, you will get here. So now they're here, you can do random sampling. So from here, also you got four examples here. Similarly, you may get 100 examples here, and from 100 examples, you can take 10 of them. Give it to a neural network to learn, etc. So a simple model, a linear classifier if you're using it will be more like w one into X t minus one plus w two index t minus two plus W, two X t minus three, plus some noise, and our objective will be learning w one w two w three. These are our linear classifiers, and we have seen them for enough time. Our problem is given x zero x one find w one w two w three. And this is nothing. But, your linear classifier confidence of the linear classifier. But we don't want to do that. We want to use MLB or MLB what is it what is the difference, given X t minus one t minus two t minus three, we would like to predict XP, with the help of some weights in the neural network. And this is the model that we want to fit. And this is, we will give us a mean square error, the predicted minus. The neural network predicted, minus the actual minus neural net to predicted square around you minimize for all the data, and you will find the best w we know what the blue. Such a W is what you really notice: the MLP can actually make this prediction, either with no big change in the solution. Look at this, and what we're doing so we have a time sequence. Being t plus one t plus two etc., given what we would like to do is this. Build a neural network, which takes these three predictors as functionally saying, We made an assumption, the only the three batters if you believe five matters. You put five. If you say that again one-week previous firewall matters, we need to use it a little bit differently, so you're making an assumption. So, this is how the raw data, we rearranged, created a new data, which is more like a given v one point x four u and v to predict x four u and v three predict six years before predicting x seven sorts of a problem. He created a new problem which the suits are a meta way of looking at it. And we asked MLP to learn that. So if you want to look at this as a neural network, we have pictures and visualizing, you will do x one x two x three predict x four. The same thing being taught is, how do you measure the performance need absolute deviation means, so all this, you can use similar things for me as a loss function so that you will. What are matters to you, you can use it as a loss function for your MLP?

So the summary is that predicting future samples is a new problem for us; because we never looked at that sort of problem in the machine learning setting, our solution is very similar to what we know. It is costed as an equation model can be linear, linear regression or nonlinear with an MLP on how many past samples or future samples will depend that the order of model of the model could be guessed. So that's all till now is that okay look we saw a very, very different class of problem but solve with existing code, with no big change. No big effort, be solid, so firstly, you should appreciate the power of MLP as a nonlinear function approximation, the second backpropagation as a learning algorithm.

I hope you have understood the time series of ml techniques. It was very interesting to see how a time series problem. Recast itself so that it can apply powerful techniques such as MLP and backpropagation. See you in the classroom.

--------------------------------------

Imagine being able to predict the stock prices. That would be wonderful isn't it. And we would be then super rich inside. We have AI powered algorithmic trading program that trades in the stock market. More specifically, any stock price prediction has to adapt to machine learning techniques for time dependent data. This is called as time series. Let's hear from Professor Java. To understand this more in detail.

For Baca was always set for us. It is a set of examples, an ordered set x one x two x three x and our examples we say are x, x y y is our example pair input output pair, and we have, ie runs from one to n. But there is no dependency on the pipe, like, what, two is after one type of notion we never had. But a number of data that we work with, also has some amount of dependency. There is a sequence nature to it. Like the price of a commodity that I buy has something to do with what what the price, we had in the last time instant slight time instant of last day or year or month or whatever baby the unit that you take in this space has that hour it is a not a set, it is a sequence is an ordered set in a way so then x one and y two and y two x one, x one x two x three x four and setelah. So, most of the time, the x is a type. It may not be the time. The example that I took about a sentence access door time. While I did talk about price of a commodity, or some stock price, the commodity, the that the axis is usually time series data has time axis, and how we predict the problem is almost similar. So you have a number of packets sold population of certain country across the years from the, the learning problem point of view what we are at this moment interested is in predicting what will be the next, for example, what will be the population in 1991 will be the given this data, can we say something about this. So this is the type of problems that we will see. Let us see how do we formulate this. So this problem we also say there's a time series so there's a classically this entire thing was formulated signal processing community looked at this problem extensively statistics community looked at this problem extensively time series is a big deal of this time axis, people also call this as a time series and with a lot of popular applications in meteorology finance, marketing and many other areas. It's quite obvious to discover that B word applications are. So now that two problems situations, which we'll read go one by one. The first is quite straightforward to see, there is a price that is x one x two x three x and n supplies available to us. This is one model of more proc so this is just an observation. It could be one parameter, it could be a scalar, or it could be a vector, think of it for the time being, it is a price of commando available every day for many many many years. And what we are interested is in building a model that will allow us to predict new or the past better. The second is more complicated, because we believe that rain influences, agriculture, or economy. And that influences the stock price. So it's not obvious how the rain is changing the stock price. We believe that we don't have a parameter but they plan on having influence. So, but this also there is a mechanism that is, it can also be modeled as x influences x rain can, I can measure. So they influence us at Etsy some hidden parameter, some big like an economy of a state of the economy, what are some things like in it, which is usually a hidden state, and the edge influences Why. Why else I can measure up. Why is the stock price. So we, we can, so this is a second model. So in this case,

we observe x one y one, x two y two x three y three x and y, etc. So we know the line, how much happened we know the stock price, but we don't know whether there is a relationship between the till now what we have been doing is, we would like to predict y, y one u one x one, we would like to predict y three given x three, we'd like to predict y given x and this is how we have been formulated, along with our assume that there is a time dependency and the time dependency could be exploited in this entire process. That's why I told you, now it is a sequence, not as a set. We were randomly sampling. An example like when you talk about Stochastic gradient descent. We are always, you have a pool of examples. We take an example from it. So this model doesn't work well for us. At this stage, because there is a time dependence. How do you break that. So we want to break this type of dependency and put it back into our neural network how to do. So let us let us assume, let's start with a simple problem. We have a sequence x zero x one x two two x. So price of a commodity. x p, where t is a time index of the sequence most of the times in this sort of notation people use a time t as an index variable because of the condition called time. So we made an assumption. When we are making an assumption at this stage it's a strong assumption. But believing all the previous patterns also use it. That, that the measurement of time period depends on on the three previous ones. So now we can convert this into a nice representation. So what is our role data was a time p one x one p two x two t three x three t four x 45 x 46. So this is how our data was. We relied to this data, such that, given x one x two x three, we would like to predict y x for given x two x three x four we would like to predict x y given x three y x four x five, we would like to predict x six given x four x five x six, we would like to predict sir. Nothing new. We just rewrote it so now this becomes very nice for us because we can go back and say that our Look, this is your x, this is your y in the traditional setting for a neural network, or for any machine learning algorithm. This is a feature representation. The three features. This we would like to predict. Our problem is now reduced to the same, we'll want to learn a function, such that, given these three measurements predictors.

Now, this sense of from the seven we got four of them. Now if you have 100 of them you will have too many more than three of them, you will get here. So now they're here you can do random sampling. So from here also you got four examples here. Similarly, you may get 100 examples here and from 100 example you can take 10 of them. Give it to a neural network to learn, etc. So a simple model, a linear classifier if you're using it will be more like w one into X t minus one plus w two index t minus two plus W, two X t minus three, plus some noise, and our objective will be learning w one w two w three. This is our linear classifiers, we have been seeing for enough time. Our problem is given x zero x one find w one w two w three. And this is nothing but, your linear classifier confidence of the linear classifier. But we don't want to do that. We want to use MLB or MLB what is it what is the difference, given X t minus one t minus two t minus three, we would like to predict XP, with the help of some weights in the neural network. And this is the model that we want to fit. And this is, we will give us a mean square error, the predicted minus. The, the neural network predicted, minus the actual minus neural net to predicted square around you minimize for all the data, you will find the best w we know what the blue. The such a W is what do you really notice that say the MLP can actually do this prediction, either with no big change in the solution. Look at this, what we're actually doing so we have a time sequence. Being t plus one t plus two etc given what we would like to do is this. Build a neural network, which takes these three predictors as functionally saying, We made an assumption, the only the three batters if you believe five matters. You put five. If you say that again one week previous firewall matters, we need to use a little bit differently so you're making an assumption. So, so this is how the raw data, we rearranged, we created a new data, which is more like a given v one point x four u and v to predict x four u and v three predict six years before predict x seven sort of a problem. He created a new problem which the suits are mlta way of looking at it. And we asked MLP to learn that. So if you want to look at this as a neural network we have picture and visualizing, you will do x one x two x three predict x four. The same thing being taught is, how do you measure the performance need absolute deviation means, so all this, you can use similar things for me as a loss function, so that you will. What are matters to you, you can use as a loss function for your MLP.

So the summary is that predicting future samples is a new problem for us, because we never looked at as a that that sort of problem in the machine learning setting our solution is very similar to what we know. It is costed as an equation model can be linear, linear regression or nonlinear with an MLP on how many past samples or future samples will depend that the order of model of the model could be guessed. So that's all till now is that okay look we saw a very, very different class of problem but solve with existing code, with no big change. No big effort, be solid so firstly you should appreciate the power of MLP as a nonlinear function approximation, the second backpropagation as a learning algorithm.

I hope you have understood the time series ml techniques. It was very interesting to see how a time series problem. recasted itself so that it can apply the powerful techniques such as MLP and backpropagation. See you in the classroom.