

FOUNDATION OF DIGITAL BUSINESS

Surojit Mookherjee

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Week 01

Lecture 20

Lecture 20 : Linear Regression and Deep Learning Model

Good morning. So, in my second lecture of this module artificial intelligence for business now and beyond, I will briefly discuss what is some methods of linear regression and deep learning model. So, as I told you linear regression is one of the most popular the simplest regression analysis technique and we will just take a look at that. If you are interested you can always study the other regression models all of them are available freely on the internet in the text books etcetera. So, and any analytical course.

We will take you through all the regression different regression analysis techniques. So, it is a big subject by itself called mostly we use in management courses, we teach business analytics in introductory course and then we also do a advanced course because analytics has become a very big career option of today for most analytical person is very high in demand. So, because the simple reason is that business wants information, business wants insights and where else can you get this data from other than past data. So, nothing no better teacher or no better educator than your history.

that is the history of data. So, if you have data how can I utilize it, how can I process that data and that is all about regression analysis. So, you have data and these are the techniques for analyzing the data so that you get the desired insights or output which you are seeking. And the last topic will be deep learning, a brief introduction to deep learning again a very complex subject. It is a very large subject and now being heavily used in various fields including science, research, chemistry, physics

and you will be surprised I hope some of you might be knowing that 2024 Nobel Prize in physics and chemistry were given to one AI scientist from in each of these area like

Nobel Prize for physics was given to three people and one of them was a AI scientist. he is known as the godfather of AI, why because he developed how this deep learning model this artificial neural network the physics behind that where the how he developed that and found out the physics of this the neural network and made it so useful and popular. And in chemistry the same deep learning technique was used to predict structures of neural networks

So, we have millions of proteins and prior to that analyzing a structure of a protein used to take months or years, but today they can do it in minutes or hours. So, by today analyzing one protein used to take maybe sometimes one year, but now we have already analyzed most of the protein structures that are known to us and people are now predicting new protein structures using this deep learning models. So, what is new materials for example, new materials, new compounds. So, new with new protein structures we can find out new drugs which can take care of diseases which currently cannot be addressed.

Treating advanced complex things like Alzheimer's or cancers or various things of things which for which we do not have allergy for example, do not have medicines. So, the new proteins which we can discover utilizing these deep learning models can be utilized. The same thing which happened for the RNA the molecule which was discovered to for the COVID-19 to tackle COVID-19 through vaccines. So, this was the first time a artificial generated molecule was used.

to create a vaccine previously always all vaccines were always you know done from real viruses taken from wherever from human bodies or animals etcetera. And those were injected as in small quantities, but now the thing is changing now we can create vaccines using chemical molecules artificially created. So, moving ahead talking about senior simple linear regression. which is basically we are trying to find out the best fit line amongst a scatter of data. So, we have plotted the data we have x and y. So, the basic equation you want to fit the best fit straight line and the equation of a straight line is y is equal to m x plus c

or from in this picture it is probably y is equal to theta 1 plus theta x. So, as you can see here y is the dependent variable. So, that is the output. x is the independent variables that is what the parameters are this is what we are measuring and this is x and and then we want to find out what will be the output which is y, theta 1 is the intercept and theta is the slope So, intercept is when x is equal to 0 the initial point what is the value of the y and

slope represents the average change in the dependent variable for every unit increase in the independent variable every 1 unit of change what is the change in y. So, that is the tan theta or slope and the intercept represents the predicted value as I told you when the independent variable value x is 0.

We have assumed that our independent feature is the experience that is X. So, X is what we call as the features of the process, something which is influencing the process. So, it could be temperature, it could be density, it could be viscosity or whatever you are measuring something is happening and the output could be whatever some change conversion of a chemical reaction into something a product you are getting a product. So, with influence of temperature on the X you are getting different levels of volumes of whatever the product is the output is increasing or whatever decreasing etcetera. So, you want to express this as $\hat{y} = \theta_0 + \theta_1 x$ this is called \hat{y} this is called $\theta_0 + \theta_1 x$ or $\hat{y} = \theta_0 + \theta_1 x$

plus is equal to $\theta_0 + \theta_1 x$ these are all x_i means the various multiple readings you are taking of that same variable the independent variable. Let us say in this case temperature. So, you are measuring temperature T 1, T 2, T 3, T 4. So, these are being represented as x_i the levels. So, the y is the levels.

If you now take a thing from the supervised learning perspective, what is y? y is the label and x is the input the independent training data the variable. So, if I give this will be the output. So, these are the data is something which is known to be these red dots. So, for this in this for example, for this x for this input variable this was the output corresponding output we get on the y axis here.

So, for each of this point I have a x value and a corresponding y value. So, this is what I have plotted because this is my history and this is my sharing data. find the best θ_0 and θ_1 values we get the best fit line. So, we have that y is equal to $\theta_0 + \theta_1 x$ and $\theta_0 + \theta_1 x$. So, I have to find out what is the θ_0 and what is the θ_1 which is the coefficient. So, when we are finally, using our model for prediction it will predict the value of y for input value x, but now since θ_0 and θ_1 are known.

So, I have my line I can draw my line and now for this line whichever I x I identify it will tell me the predicted value y the new value of y the predicted let us say the predicted value that is more correct. Now, to achieve the best fit regression line. So, we could have drawn like the line anyway through those bunch of those dots. So, it not necessarily that you will start from here and go like this it could start from somewhere and go somewhere like this or maybe like that.

So, how do you know which one is the right line which one is the best fit line. So, here we use the term best fit line. So, to know that this is the method or technique. So, you predict the target value \hat{y} such that the error difference between predicted value \hat{y} and the true value y is minimum.

So, it is very important to update these θ_1 θ_2 values to reach the best value that minimizes the error. So, you are trying to minimize a function. So, you are doing \hat{y} minus y_i and you squaring it and divided by the number of readings and these values should be you try to minimize. So, for minimizing also there are certain techniques, but the aim is to reach the minimum error you get. So, this is the error function.

So, here if you can go back. So, this is the values are the levels where this y_i and then \hat{y} will be your the predicted value the new value. So, \hat{y} is your predicted value and y is the true value that is from your history. So, and the \hat{y} you are getting from the equation the new line.

So, the goal of the algorithm is to find the best fit line equation that can predict the values based on the independent variables. So, in linear equation as we call something called a mean squared error MAC cost function calculates the average of the squared errors between the predicted value \hat{y} and the actual value y_i or whatever. The purpose is to determine the optimal values of the intercept θ_1 and the coefficient of the input feature θ_2 providing the best fit line for the given data points. So, once I can minimize my error and for that corresponding minimum error that line I will have a θ_1 and a θ_2 .

So, θ_2 is the coefficient for the variable x and θ_1 is the intercept when x is 0 what is the value of y . So, that is θ_1 . So, when x is 0 what is the value of y and θ_2 is the coefficient or weightage of the feature or the variable independent variable x_1 . So, I give some weightage to each of these variables. MSE function can be calculate the cost function is again I repeated in the previous slide this \hat{y} minus y square over n number readings and divided by the number n . So, using this function the iterative process of gradient descent is applied to update the values of θ_1 and θ_2 . So, finally, we are trying to find out what is the θ_1 and θ_2 for which the error

will be minimum. So, that particular line for that θ_1 and θ_2 y and x is of course, there is the best fitting line. So, this is the ultimate objective of this cost function for linear regression. So, the final result is a linear regression line that minimizes the overall

square difference. So, this is the square difference $\hat{y}_i - y_i$ between the predicted and actual values providing an optimal

representation of the underlying relationship of the data. So, this was what you had a spread of data across y axis and x axis. Now, once you can draw this line then you can say this is the relationship between x and y and this is the best obviously, now we have done that cost minimization etcetera this is the best line we could draw through this. So, it will give a higher accuracy of prediction. So, now if I take any x find out anything.

So, what is the y for corresponding x? So, I can use it for my everyday say production which is happening and I want to predict. So, today x is this what will be the y tomorrow what if this x is x and so, tomorrow day after tomorrow what will be the y and this prediction will be very close to the accurate because we have done this and this accuracy again I am repeating again it is all depends on the number of data and of course, to some extent the quality of the data. The quality of data here we have not touched upon much is that something called outliers like some of the data points could be

some of these data points could be very far away from the line somewhere here maybe here or maybe here. So, such data they will bring in lot of bias. So, these data outliers what we call if a data point is here or a data point is here should not be considered for calculating θ_1 θ_2 . So, these we call outliers and we eliminate those data we do not consider it. So, then the training model becomes better. So, many such things have to be considered before you finally, ultimately come up with a very simple say straight line equation very simple looks good and you can use it for your prediction.

But prior to that all of these background work has to go on and including like I said the number of data set and then the quality of data whether outliers have been taken care or not there could be some negative numbers which is possible etcetera. So, all of those thing we do is called data cleaning. So, before we start training the model the first thing is to take a look at the data and remove all some of these unclean stuff like outliers or negative values or whatever. So, how are the coefficients interpreted?

Consider a data set where we are investigating a relationship between experience in years and the salary. Suppose you know employees when they work in a company the salary is you know related to the experience in years. So, when you hire or recruit new people you give them salary depending on say apart from other things one of the important factor is years of experience in that field particular field. for every 1 year increase in experience we expect the salary to increase by 600 dollars. So, if you have 1 year experience you get

the base salary plus 600, 2 years experience base salary plus 600 into 2, 1200 dollars like that.

And when you say the experience in 0 years that is the beginner this is the base salary, we predict the salary to be 2200 dollars. So, when you have 0 experience the fresher gets 2200 dollars, fresher plus 1 year experience will have 2200 plus 600. So, 2800 dollars and with 2 years it is another 600. So, if you plot that the plot will look like this. So, now, what you can use this plot is to predict salary for any years of experience.

Suppose the man has come with 4.5 years of experience, how much should I pay him? So, from 4.5 you can go up here and you tell you this is the salary which needs to be paid 6000 dollars etcetera. So, this is what is used this is a simple very simple example of a linear regression model to predict salary based on your sub experience of a candidate. The use cases, linear regression is useful in various practical scenarios where you want to predict a continuous, this is important continuous outcome like that straight line or any equation. So, for any point of X you can get an output, certain inputs here are some key practical use cases are sales forecasting, analyzing historical sales data and identifying

patterns to predict future sales figures, financial analysis, stock price prediction, credit risk assessment this will be sorry a typo will be real estate valuation, medical research and healthcare, predicting disease risk, drug dosage prediction and environmental science predicting pollution levels, crop yield prediction, marketing and advertising, analyzing advertising effectiveness, customer behavior analysis etcetera. So, various domains of business we can use linear regression model for doing the prediction. So, what some of the key consideration should be for a practical use?

One is of course, the first thing is a linearity assumption. See linearization works best when the relationship between the variables is approximately linear. So, first is the first assumption we make of the data set. Let us try first a linear model. If it does not work, does not fit, then the error is very high, then we think about all those other polynomial models.

Now, independence of errors. The errors the difference between predicted and actual values should be independent of each other. So, there should not be any dependence. This was the actual value and this is my predicted value. So, they should be independent.

Multicollinearity is when we use a multiple linear regression that is y is equal to like we said $c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3$. So, there are three variables three independent

variables may be temperature density and fluidity or viscosity or something of a liquid. So, these variables should not be related to each other. because there should not be could be 0, but there could be some relation, but it should not be very high relation. Because if it is a high correlation between the variables independent variables it can affect the model stability and interpretability.

So, these they should literally have independent effect on the dependent variable output, but usually it is not 100 percent independent there is some dependency. So, what we are saying is that it should be minimum. Sometimes transforming the existing variables, the independent variables or creating new ones can improve the model's performance. Suppose you are trying the result with those three variables I told you about this temperature, viscosity and maybe fluidity whatever. Now, you might not have chosen the right variables to get the desired output because you are after doing a prediction model.

So, maybe you can change. them or maybe take 2 or take 3, 4 or take replace 1 by some other variable which you are not considered. So, this is known as called as feature engineering, it aims to modify the variables in a way that can improve the models performance accuracy and interpretability. So, that is another thing which tool which you have in your hand is those independent variables, they are representing basically the features of that process or the experiment or whatever is happening which you want to use and you can use them the way you want to use. So, that your output is more accurate or better.

Now coming to deep learning, what is deep learning? So, deep learning is a subset of machine learning that leverages multi-layered neural networks. So, the multi-layered neural networks known as deep neural networks to simulate the complex decision making power of the human brain. So, ultimately our intelligence the way we think and way we find out or way we learn it is all happening in the one single organ the brain. So, what I had the AI scientists had looked for many years last 50 60 years is how to simulate this brain or replicate this brain within a computer structure that is your

hardware and software etcetera to represent the way the brain works. Let us explore this fascinating field in the deep learning revolution how machines understand, learn and interact with complex data pushing the boundaries of AI capabilities. So, these are not that simple y is equal to mx plus c type of regression model. So, these are handling much more complex set of data. Foundations of deep learning.

are neural networks, multiple layers and representation run. So, neural network is based on the artificial neural network it is called ANN which are inspired by the structure and function in the human brain as I am telling you. The multiple layers which in the term deep refers to the use of multiple layers in the network. So, if the data will move from layer to layer to layer. And you will get more you are doing continuous doing some calculation the computation and the results will become more refined as you go from layer to layer.

Representation learning is models learn to represent features from raw data allowing them to automatically discover patterns and hierarchies. it is actually considered almost like a black box because of this complex nature of this layers and nodes and synapses. It is very difficult to perceive as of now the knowledge is how exactly each of those computations is being you know the results are coming out, how the results are coming out So, people are not still as I said not very confident about using these deep learning models with a lot of degree huge degree of confidence. Because again all those subjectivity factors many unknown things are there unpredictable things can happen.

Like in chat GPT you have heard about the term called hallucination. So, it is something which is some unexpected behavior unexpected response that sets you back. So, you did not expect this type of answer, but you got that answer and some of them can be very very silly. So, they are representing the features from the raw data allowing them to automatically discover the patterns and hierarchies. Some of the applications of deep learning are computer vision, deep learning powers image recognition, object detection and facial recognition system.

So, one of the examples here is of course, if you have you must be travelling in the airports we have using called BG Yatra. So, once your face is recognized and then rest of the whatever processing you are doing onboarding, security check etcetera everything you just camera sees your face and you go through all of these. steps without any problem. So, this is completely working on this deep learning technique called computer vision. Natural language processing, it is used for natural language translation, sentiment analysis, chat bots and text generation.

If you have been using Google translation, So, what we have been experience say 10 years back the quality of translation and if you doing whatever now you will see vast improvement in the quality of transaction. And the number of languages that it can handle today we can handle multiple languages, several languages for doing translation

because it is being done through these AI models of NLP is being used extensively by Google and any other translators for doing. So, translation has literally become very easy and you can do it very quickly and accuracy level is very good.

Obviously, if the accuracy was not good you would not have accepted the translation. Speech recognition enables accurate speech to text conversion. So, when as I am speaking now for example, the transcript is getting recorded it is getting typed out nobody is doing it manually. So, my voice is being heard by and captured by an audio system which has got a speech recognition software working behind it tool which has been trained by again using deep learning methods. So, it can very accurately

convert into text like when you talk to your phone for example, you want to send audio message or all the audio books which are now we are using. So, all of these are applications of speech recognition system. So, either it can convert text to speech and speech to text either way. Recommendation system personalized recommendations in e-commerce and content platforms. So, these are what those ads which come up when you see when you are you know surfing the web for some you know something product you go to Amazon etcetera.

And then you will see lot of ads being popped out at you which are contextual that means, they are relevant to what you are searching for. So, that is what is the back of the mind is back of the application was working is deep learning because your previous searches your previous purchases your likes and dislikes are known to the advertisers or let us say to Amazon. because they are capturing all your clicks and what you are buying, what you are looking at, what you are surfing etcetera. So, from that they know things that you like.

So, next time when you go for surfing or searching for something maybe you are not probably even want to buy that time or do you say, but the system knows he is now on the web. So, let him let us push the things which he or she likes. So, those ads starts popping up. So, this is the recommendation system. So, you can they can recommend if you are buying something say Amazon you buy a particular book and they know that you probably like this type of books.

So, they will recommend if you are buying this why do not you buy club you know together other things and give you the recommendation. Autonomous vehicles of course, we all of us know the deep learning models they are analyzing sensor data for safe navigation these are the driverless cars which are there and probably will become more

common very soon may be in a year or two by 2026. And finally, generative AI which is now one of the also the most talked about topic or subject, it provides the engine for the sophisticated generation capability we see in tools that create realistic images, write human like text and compose music and videos and what not.

So, whatever you want to do answer this question answer that summarize this text write an essay write an story. So, whatever we want it is being able to do it. So, we will talk about generative AI in more details in subsequent classes. So, this is a pictorial view of this deep artificial neural network. So, here first we have this input layer and then we have all these hidden layers which you do not see and then we have an output layer.

So, this is basically the brain. So, we input things through right away here eyes or you know human beings or inputs are through eyes and ear. 2 input there and the output is of course, could be anything or action whatever we act or whatever we decide or etcetera etcetera. And rest all of this is what is happening in the brain which is not of course, we also not very clear how the brain works and this artificial neural network also we are not very clear as to how these things and layer by layer the computation goes on,

but something definitely comes out and this the quality of that output will of course, depend on again the training data and the number of data the huge. So, the number of volume of data the size of the data set and of course, the quality of the data will decide the accuracy or quality of the output layer. So, what is deep learning continuing on that approach deep learning with caution. So, all that is achieved great success in many areas in dealing with text video images, it has raised serious questions on interpretability in practical applications. Specially how and why the DL model arrives at a specific answer, this is known as the black box nature of the DL model.

and can raise concerns in say for an example medical diagnosis. So, if a DL model after examining several scans medical scans of a patient predicts that the patient has a lung cancer let us say. It is critical for the doctor to know why it arrived at this conclusion because it could be a false positive or a false negative. So, in case if it says it has a lung cancer then it could be a false positive means the patient may not have a lung cancer. But once that you know if you depend on the system too much and say fine this is a very well developed model trust it etcetera the doctor and takes

the decision from that model based on whatever scans is provided and tells the person. So, then he gives his prescribes the treatment for whatever disease say in this case lung cancer, but actually it does not help. So, that is the ethical question people are asking how

confident can you be on the treatment predicted output of a deep learning model in such cases because medical cases etcetera with different ball game because of playing with the life of a human being and these treatments can be very expensive. So, the doctor has to know that the whatever prediction is coming is reliable.

It also and that is the reason it needs much more data than traditional ML. And it also needs much more computational power because of this deep neural network they run on very high level GPUs or chips. They are very power consuming because they are doing huge amount of computation at a very short time, they have large capacity and that is how it works and hence very expensive to develop. These GPUs have helped in rapid growth of machine learning and deep learning. However, they are quite expensive and for deep learning is required in very large quantities.

So, no wonder the company called NVIDIA is today one of the most valuable company in the world. And it this training also consumes very large amount of energy raw power electricity for this computational task. So, that is one of the drawback of these deep learning developing these deep learning models because expensive these chips buying these chips getting these chips very expensive and while you are running it consumes lot of power electricity. Coming to computer vision, self driving cars see through the images captured these are used to determine things like stop sign, advance in there.

So, all sensor based and BG Yatra I have already talked about how it works if you once you go to an airport and once you first see the your picture is clicked and then rest of the thing you can move through your security and checks and other things without any physical display of your documents like passport or boarding pass etcetera. It also can be used in very industrial applications of quality control through computer vision it can see detect and find out defects in a product. And another application of course, is a traffic lights where you have traffic signals where you have a cameras to find out whether somebody has crossed

the line red line or speeding limit crossing the speeding limit this is nowadays very common in most of the big cities metro cities they are using these cameras. And of course, we have seen cameras being used in various other applications. So, that is all about computer vision. NLP text classification is automatically assigning levels. So, you are classifying certain things as whatever.

So, like topic for example, you can automatically assign tags such as this topic is related to budget, this is related to finance, this is related to product development. So, you can

classify a text through NLP tools or say incoming problem tickets. So, customers are raising different tickets in the system they are having some problems and you can use NLP processing to classify these tickets into different categories for example, the high priority, medium priority, low priority etcetera depending on how you have trained that.

And also you can route those tickets to this type of ticket it goes to this person, this type of ticket this goes to this person. So, all of these processes can be automated and the basic technology behind this is NLP. Sentiment analysis is another very commonly used the customer feedbacks are analyzed to find out whether the customers like the product or service or not like the product or service. This is called sentiment analysis and this is very much used in all by most marketing people in most marketing domains is very commonly used. Read from unstructured these are the feedbacks which is being given in say in Amazon about

when you go to a travel site trip advisor etcetera. So, all those feedbacks are extracted put into this engine system and the customer's sentiment is found out that what whether they like what they like etcetera. Conversational agent chatbots we type some query and say help desk thing and they can give a response. So, behind that chatbot is a NLP processing tool and could be also in connected with a GNI system and it is coming up with whatever answers to the questions we are raising. So, more about it in the future classes.

So, they are doing various tasks they can work as assistants, setting up the calendar appointment, dialing a number, finding restaurants, playing songs, book travel. So, these are all using NLP like tools like Alexa, gadgets like Alexa, the series which many people have in the house you can ask Alexa name me a restaurant good restaurant for Chinese food near my house. So, Alexa will find out and tell you which is the best Chinese restaurant near your house. So, with that I end this session on regression analysis and deep learning.

Of course, this has been done at a very high level because these are all each of these topics are very big topics and we can spend hours on these topics, but they are of course, meant for more of we go into more technical details which is not our scope here. So, we are sticking to what can be used in business for business and we should know what to use and where to use and things like that. So, thank you very much.