### ML\_week1\_lecture1\_20240906

说话人1 00:02  
Before that. So that welcome all your guests to city university 101. That's the very first type of process to get a little is a brilliant choice. What is best? All right. So ii am kumar. I come to me in the system process. You make countries and the general, the sitting professor at 1621. So my research interesting includes perceptual in the world, for example, as computation, more than of our innovation system and computational work, and not to meet the forensics, and they have to make the stuff. There is six. We're learning so much. Learning is very much the topic of the moment. Please be discussed everywhere either under the name personally or under the name deep learning or under the more water farmer australia.

I believe you may have already take some sort of version of learning course on the website. So hopefully our person may be in different already for towards the last.

So first course, information production. And the investors were going to review some fundamental math answering some messages. Here is our for ancient teams. I'm going to doctors is your email and the office. I I don't have an office. I'm here at city and one, but ii haven't checked it before. I just went outside despite that was for better, the pipe is a I managed to be tripped in the big trains that here he delivered this perspective performer.

We have a total 6 years. So how we change our in ta and those six t as are all my ph students. Yes. The old girls have the 4 weeks. Then we have also a total of eight to force you coming. You have some to learn some apply immersion enough to do some employee gets that to. You can get an issue on the argument. The chemist is foresight, nice, your own responsibility, check, hands and reverse come back to work on. The topics. I will weekly announcement to summarize what happened towards in the previous great and what will be taught. So it sneaks more teaching activities, but you should sit in the classrooms to learn machine.

The only reason why i'm using room meeting is because it makes the artificial, either for me. I don't need to come back to this blackboard and to sit for maybe a time for a long time to walk. That's what I see. It is nice. And for the lecture, without present machine learning, language is emphasized on how those are. Some mathematics would be, right? Exactly the problem. We have a total eight followers, so who's scheduled up? Determine. And the purpose of this tools are done using machinery are not small. Point is that you get intuition and understanding paris machine learning areas. We have much money on which have we have a plan, for example, to but how the finders to get better performance on a particular person in the past six. You have a little four homework assignments for which require you guys to solve some difficult mathematical problems.

We have a possible which on a beautiful, higher machinery, she's sort of real world for during, I think the middle of this for chorus and our ta team will work on identify a particular measurement for all of you guys work on the central, compete in the past to say the performance mayor, they the number of parameters in the machine learning model approach is perform subject.

You can do the course project all by yourself.

And alternatively, you can formal group up to three. Do they for project together? And I think nature 500 % will come on the last more details regarding possible. There is somebody. So on the systems, we have course work. We have by the none. The course will only have 70 % to have. Remember, it was fast to one. We have to reduce for assignments. The couple of these course, number 1 + 14. Do we have 13 tendency? Group 13 is, by the way, right? The cost per ship be in possible. But it works for invention and optional presentation. And you have a right to choose trading or not for that. So what it is here and that never said a little about the course on project reports. I i'm not expecting you guys to give me a very density course. So when I was there students as the usdc university of hansa go to china. So we at freshman, freshman student and freshman students, they do physical experiment anyway. So all over and we do need it to be with all kinds of those experience.

So after each year, we do it report. Right? Initially, as this as we 125, the report limits should be less than 20 and 20 pages. As we go the medicine, it becomes 50 pages. Some certain union can be in 100. For a single side is small and experiment. I don't want you guys to compete in remains of possible. Make the course project concise, complete. And here that should be enough. Right? We have a better example, which becomes 30 %, and you must get at research on the product and research on the cost product. But of we recommend several forms freely available from waters and download environment.

The first is the introduction to machine is high, some financials and sarah. The book gives a change introduction, commercial learning commerce practice on board. I think it has some in the title. And the first was andrews has been a maintainer and one of the core contributor of second nerve. A machine learning to keep 5 years in the same opinion. In this course, we're gonna do a second term extensively in this course for us.

And the second book is pattern recognition, machine learning. This is a very clearly and very reasonable. We'll give you a delightful basin tower of machine learning and doctor christopher bishop, what it is. Company is a magical technical failure and the recommend swords research ai works and he is a expert in the artificial intelligence. And he recently just finished the deep learning, foundations and concepts. And this is a very beautiful. I have just give me this for a reason to do that. If you have come, I had you come and you guys to take it back. What's interesting is that so christopher has aphd degree in theoretical theory, cop the circle is, it's come back to the population by stephen boyd and naming manager. This book is very red. One is starting to talk to this more than two context of nation. And stephen boy is a chance in the field of position. And he is a professor at stanford and he developed a very, very successful online course, read, reading, available you guys on complex of nation.

So when I was on understanding when I was an undergraduate student. So I learned this online course by myself when I was graduating, why is everything is on my performance? And for those 2 times up, there are a lot of fun from this regarding me. If you ask that ownership and all of the nation, all right? They remember was steven's phd student and comments be the professor and music area. The 4th is the matrix. This book is just a collection of facts and identities, approximation in equations and in qualities regarding matrices.

For example, if you are not sure about what's the gradient of some matrix quantity respect to some vector or matrix, you will probably find the solution in this. I would also upload this up to the cameras. For example, the last is the new article on its application by paper string. This is super little by paper strings professor at mit and it also developed a very, very successful online for something on from which is which is still currently, which is still constantly evolving until recently, because you put a job requirement. If you want to recap some fundamentals in your article, or you want to learn, in fact, for this case, excellent resources for us to check it out. So whose abstract so this, for the goal of this course is to introduce students more than the machine learning. And machine learning basically allows computers to automatically learn the class context patterns from computer data, such as taxis. That document sounds into this course is intended to give all of you from most theoretical and attractive standpoints with emphasis on how to apply machine or not to build for the problems.

As at the end of all students will have close as I want to find you and have experiences which may not. Therefore, so in some occasions, so you may feel like you are sitting in some method. In case you find different, not because you should drop it during the annual period. I'm not sure I think this course, I think the woman is a mandatory course, right? About. Yeah, it's a manager, right? Therefore you have to bear with it pretty much. And we have four and sales. The first is to identify and explain summary analysis. The second is a time machine learning partners to solve real world problems. The stories divided into different machinery outcomes is suffer technologies and technologies. The forces on the mathematical aspect of we do have some formal prerequisites for this course. So also then I expect to be with this method or have ab to make a test in the background.

For example, in the article that is not application, you need to know and understand. The most fundamental concept is in the article that's in the cast as you need to low, or we have the definition of union function. The range will come as a function how to take. The derivative will function respect from congress, because in machine learning, but not using a particular opposition to it. I in the grading base for commission to learn factors in mercenary. I don't think we will touch the interval of some functions, but we will make extensive use of gradient, global function, and probability and statistics. What's run? You need to ask this question. What's that means less dangerous? If you will give one number of definitions. What's one variable? The answer should not be. A random variable is just a variable. That's a dummy. That's a down house. All right. So mark, you need to know duty variant documents. You should multivariant documents. Not nervous, as I mentioned, was an observation. So what's they come to accept? What's the complex function, what's the complex of nation? What's the duality that is another very important concept of the solution?

And if you don't know those concepts, there's nobody will get to cover to define and introduce me some good example for the universe. For these concepts, important concept, important basis. This was required a little python for primary student experiment, python, as we go, self evaluation and not ready to go. What's the derivative of x this stunning notation? He wants this funding efficiency. So you answer, transpose, it's transpose, so x transpose, it's read as x transpose w with respect to w where x transpose is a one by n rho vector, and w is in. They end by one column lecture in our course.

So when you encounter meat, opal letters, it's always gonna be the column vector by default. It's a column vector to convert a column vector into a row vector. You are going to transpose it. That's why x transpose above it left. What's the derivative of this quantity with this topic? Can you the answer from the students? It's just x what's maximum at the estimation? Or I make for sure what's maximal posterior estimation? What's freedom is? What's so ask? Freedom is, what's principle component analysis for singular value information? If you score 3 to 4 points up, so we go 0 to 2 convertible future power. That's why we should be sitting.

So then honestly, so if you mind as own rule, I think we found it. Reason is there is a lot of picture rise up. I finished assigned by yourself at your own program as it's fine to discuss with a fellow classmates to start with the tears, starts with me. And on the other side, this is particular solutions. Don't give me as a reference for plagiarism is between those similar and progress guilty. And you hurt your own respect. Not report cheap. As instruct we have the responsibility to report and dishonesty cases. So as for classical. So we take suspect plagiarism and then all is in the presence of general. Ai and the basic control of the department program in the united. For example, child gpt to be or not also compliant, maybe during these during this course, we may expect gpt five, right? The students are not allowed to use cni for so funny part. As always, ii always believe that programming is something another shelter need to programming.

You need to compare different version post, so why the particular version is more efficient than the other one? Why the version version is more numerically safer than all? Then you can learn how to that. And students are allowed to use for your writing assignment reports. So the argument must make super citations because generally have reference to some source on the internet. And you need to refer to the original source. I gave the whole person creative salvation, the original source, so that of the cologne or quality is considered pleasure list. And then here are at the emphasis don't perform self. Maybe you have so maybe you have some assignment course project or stuff that has already completed in other courses. And we should submit this course or some other courses.

Now is considered a simple page if you have a 10 page report, right? If you want to write a similar report, don't simply call kind of place paragraph from this document. This document now is also consensus pages. Even this first document is completely relevant by you about 7 pages, right? So useless caution as genius, good at fabulous 18 for others, they can force prepare for option. We will in this course, we will introduce on the line memories of some popular children was the end of the city as part of the early century.

A question on the course information. So let's do a broad introduction, right? Before there are machine learning. So we are always gonna ask this question, what's normal? Right? If you enter new field, we have some key words. We have some key concepts. You need to always ask this question how to define this concept. So what is, which? So from the behaviorism aspect, learning is a long term change and behavior build the experience. So singer was an american psychologist. So human action as dependence on the consequences of previous actions, which makes intuitive things, right? This is a principle known as reinforcement. And if the consequences of action can also be back there, you see, 49, we thought this action is gonna be repeating. So it comes for you. The outcomes of the action turns out to be good. So it's highly active on this action is gonna be repeated. Here, it's important to look at the goodness of a action is individually independent. From the cultivation. Learning is an internal mental process of interface, new information. You start already established new frameworks and topics is frameworks from the connection you do. Learning is a physical process in your brain or other animal screen in which europe going by developing the synapse, us, three definitions of learning from different aspects.

So what's machine learning? From the science view? Machinery is of the study not give computers we need to learn without being explicitly good. Man. So samuel, in 1959 actually coins the term machinery. I'm coming to in 1997 if machine learning very working definition, a computer program is set to learn from experience e with respect to some task t and performance measure p if the performance at task t I sneered by p improves with the experience in approximation.

If you don't understand this definition is nowhere that we have a company examples images. So tom mitchell is a professor samuel and actually founded the words first machine learning department. I see. So jordan at 2050 is returning another abstract relations found is rapidly growing technical.

Now, you have the perception of cultural science and statistics. And I call a lot which we can decide why they give such a definite course. Jordan works in the fields. The drawing is a japanese statistics or regions, at least that using drug. This from the samples, will this have a conference with visual example? So that is recognized 100 pages in image. Here, the input is a 28 × 28 image. It's just aa 28 × 28 °. We have 28 rows and 28 columns. Each grade for the pixel can only take one or two values at zero. It's black or one is white. If it's a better image. That's a 25 image, and they can fly from this image into a 784 dimension vector. That's important. Here shows examples of digit zero, digit one, image of digit two, and you can see for dg four. So there's a lot of vibration of communications.

Therefore, it is very difficult to identify rule based programs to recognize the rule, basic provide these emails or stuff like that. Right? And it's 13 people too hard about those kind of things. But we have a machine learning solution. We gather some example data, and this role, it in professional apples, machine learning apps use the opposition to be 21 men. The trainers are not being able to regulate class of those hundred digits at the recently actress.

Again, let's review me, trust you. We chose definition of mercenary sense is very different. Very important. Let's repeat the definition. So a computer program, he said, to learn from the extreme scheme with respect to some past t and performance majority, if its performance at past t as mirrored by t includes e task class of task t means on learning is task specific. You have to space about very different machine learning atoms.

Such as in the previous sites, we have featured with the vision. You may have some other machinery as subjects last week to gather items in the one cluster if they have similar patterns. The performance layer p we need to define a quantitative performance measure. They use it as the most function to optimize the prime interest, immersion on. And at the same time, we need to prevent us into lots, learn to adapt to the noise patterns, use this noise pattern in those training examples, nice to prevent over feeling. And we will use some forms of you make we will get over beijing realization of detail. Definition. There is no more. So external c last example later, man, so machine learning is the agreement. More data is better. So extremely in the previous exam is just example images of the speeches and machine learning, deep learning and artificial intelligence or that control. Emerging learning grew out of early work in ui and other for such as statistics, physics, singular processing, neuroscience, and they filled by more powerful computers.

We have different fields in here, such as general artificial intelligence for artificial general intelligence and machine learning and language processing, conservation and robotics, and so on.

You from this paper, you can see that you can view deep learning and the subject of machine learning, which is a subject of artificial intelligence.

We have fought. Typically, we have several folks in machine learning, ship rest early on should press learning, reinforce learning and learning theory. For sure, press learning training that has input and outputs. Right? In the case of each replied reputation, the input is 28 by 20 battery, and also considered as seven and four dimensional better. The output is one of the ten digits, 01 up to now. And the goal is to learn the function mapping from inputs to offers. And we have two passage, super learning classification and regression. And the classification is when the output only consists of limited number of categories. And in this particular example, the output only consists of two cases, three, digits three, and digits eight. It's a binary classification problem. This course is represent the image of digits three. This stars represents an image of digits eight. This straight line is the decision boundary given by some already trained person learning articles.

Here is my question when we have a new data here. So what's this prediction? Is it three or is it eight? 33 ranking? Because it below this is and similar for regression, the output is continuous. So it can take any value within the interval. In this particular example, the input is doctor per million population, the outputs as a continuous extreme romantic expectancy. So I think expected. So those two companies are positive correlation. This curve is the basic model relies on trend regression model, and based on this regression model, valuable informations.

So, for example, when we know that the total per leading population is around 2,500, so what's it predicted to be? There's no lack of expectancy. It may be around 74, 75, right? Last prediction. And for answer plus early. So training only has a team, so there's no outputs. We have this destination, which is the mother of four country president boss. It is to construct or believe in model over me, but ask them, discover groups of single examples of damage and reduction. In high dimension data two or three dimensions. Ok for example, for visualization, here is a example of classification of one data set into three tasks. So that one has the same color, which means that there are lots, right? Not the same. And for this particular task and ask for examples, we group data ., but are closer to each other within the same. This is a dimension and reduction example where we project a dimension of each image. It's a completely better date down to 43 dimensions. So that is embedded those high dimension images into the torture that is dimensions.

So we as in purpose, in this particular example, different colors for different pages, as you can see, each image of percentages are project into the same classes and in the reduced dimension of space, which makes perfect sense.

And we also have reinforcement learning, which is to make a sequences or 1/7 of action is given the common state, but there's a lot of practice environment. The goal is to maximize the world.

At some point, not everyone and some responsible, you're gonna receive a report of functions and actions may professor beautiful. This gives you an excellent example where the most tries to solve business. So only when the most successfully solve business, it will receive this reward, make the key. At the very end, there is only reward another way, there is no reward. Another way, maybe occasionally they may receive some punishment because it may hit the wall. Right? And they turn one each return ones and find another possible classes. Regarding learning theory, why does machine learning more? Do we have performance guarantees? Do we have found unexpected area? What types of functions can be represented by artists? And how much data do you need to train? Some particular measurement artists? Right? We train machine learning album on some training data. We get training air, and we test our machinery road on some new data. We get test data. Do we have some performance bond on the gap of training? And they say, that's 1 kind of nursery. And it's still extensively investigated in the modern days of paper and the worst learning in the real world.

And we have a lot of which learning applications when you from the campaigns, so many system levels of these cameras, visual recognition, your smartphone, so driving cars, advertising, personal assistant chat gpg by chat report, filtering in your email and recommendation system. Produce some items on actual of unit networks.

And then ai generate confidence. Here is an example of chapter two. When you from chapter two is this. With this is we can tell you how china gpt works. The response is complex is china city is a large average models are uses the burning to generate. Human ethics is natural language. And it's based on the gpt the gpt is short for generative treatment. Transformer. Transformer is a particular type of your level and pretend this particular machine learning or position to give you learning.

And then general generally is a particular type of machine learning arguments. I as they can revolution, describe generation. The model is turned on massive texts like those articles so that you can understand the patterns in the structure of that. The way giving a promise give a promise means some input text to this mercury per permission and model. The model is a technology general text that continues giving in that matter. The key technology, it mathematically that chain technologies is cheap. It's just learning to predict the next word or learning to predict the next token, given previous words. Right? Essentially, we are going to build a conditional probability model, p of the next word, given previous.

After we have this conditional probability model, we are able to sample from this product model to produce a possible export. We puts the produced x work behind the conditional bar and pro produce an expert. And we do this repeatedly to auto tools, the coherent human life and the coherent paragraph. So that's the ta but in terms of engineering, how to scale up, how to with the efficient and empowering, how to write efficient book on us. And we also have takes two image models where such a model takes. Let's see, a short description of natural language has been put and produce the image that matches this text.

Right here, for example, a close up your hand on with this group from it. So text to image models actually emerges 10 years of 10 years ago, but with image success, up in twenty, 2002, and this particular field has experience the numerous things. And the underlying two is a circle, I think, institutional. In our course, we will touch on some mathematical background of your institution. Now, it's give a general introduction on the machine learning training pipeline. So first, we gather example pairs, right? In this particular beach classification examples we have by a beach classification. And we have examples of three, each three, examples of each eight. There is a typical procedure for feature transform. The key question is how to represent the data sample as a vector numbers. In this particular example, for an image, we may compute the mean of the image as the first picture. We may compute the variance of the image as a second picture. We may simply count the ratio of but I think so as a circulation, we may do it more fine greatly.

We may partition the image into four long overlapping blocks or not overlapping image patches. We can't the ratio of black pixel in each patches, give you another extra four papers. And we get all creatures in for this picture. So that's they include examples of how you perform pictures in your mind way and in the face of deep learning. So the feature extraction module can be automatically and under 20 organized. You don't need to have this priority on picture or our particular personal task to be extracting this.

So after future destruction, my favorite switch records to a machine learning algorithms. A machine learning album has 2 types of parameters. One is for model prime first or higher prime. Those prime are gonna be set managed by you have. In this case, we have to be 1/2 prime proceed, which is randomly set to ten. The second type of parameters are training parameters that have automatically learned using some populations. Take this. And the key question is, how to separate the data in classes, which mercenary model do you choose? Which of the relation techniques do you choose, which most function to optimize after the much more is trend, but we may use it to make predictions. What's so called education sector? Validation set is another set that is completely independent of the training data. In this case, we make it an error rate of 21 %. To measure how well does the model work on new data. The validation data is not seen by the train, by the machine learning more during training is completely newly. That's more virtual machine learning training pipeline. We have also emerged model, select pipeline, so being sending data gathering, picture instruction, we said 90 % of those type of primary to turn in the model, the prediction on the validation set and get the area.

In one. I'm not sure whether c is equal to a lot more or not, right? When we travel in value, c is equal to one. If we change, c is equal to ten to c is equal to one. We're going to retrain the model of the century data and gets a second machinery model and testing on the same vibration data can get, in this case, that there are ten when we try to get another one, says ten return the whole model, tested on the validation set, and get the area. And then which state do you choose in this case? So impressive. The one with the lowest validation area, which is c is equal to one, the one with the lowest validation set. Now, the situation is gonna be a little different. We use training data to determine the training parameters. In the machine learning model, we use violation set to determine the hyper families in a machine learning model, which means that those students that are already been seen by the machinery model already, we may even serve data set, which is called the test data set that is independent of the training data, independent of the valuation set to test the generalization performance too much anymore.

Last training for training primaries, validation set for harbor finals, and the test data set for test, for the generalization performance will perform of a particular learners anymore.

Any questions on? What was that about? So that's the model with no expectation. Now let's do a a little bit with you. First, in article, the real vector space is a set of elements, apps, which consist of ten numbers. It's one objects. And why we would be transpose here to think about it. Because when we write something like this, one up to the extent it represents a row row vector, by convention, when you encounter the need for both letters, it's common.

So that's why we put the transport scheme. It's a common by default, where each xj is a real number and elements, x support vectors, they satisfy the formal problem. Addition. So we can add two records of the sentence which amounts to adding the corresponding agents. We can do scalar product, multiply a scalar with the vector, which amounts to multiply each elements in the vector bias, the scalar and return respect. And you also can define inner product of vectors of the same length. If x is a dimensional column vector rise and then column vector, then x transpose side is just the summation of the corresponding products of the summation of the curve, of the products of corresponding elements, right here.

So let's work out the dimensions. X is n by one, then x transpose is one by n why isn't by one? For many for matrix product? Matrix multiplication, the inner dimensions must match in order to perform value. Matching my application. Therefore, this will give you one by one vector, one by one vector is tensor. The standard does not. That's what you have. This is the number, the sum open numbers, which give you a number. And here's my question. What's the dimension of xy transpose? It's n by n because this is n by one. This is one by n which will give you an by n matrix. This is known as the outer product. You guys should be able to work out each element in this out environments. So if you don't know it, search online, ask you, right? The matrix and n by n matrix is just a rectangle of real elements, where I is the war index, there were total animals. The j is economy dense. We have to end UPS, and the metric supports the following operations. Addition, we can add matrix is at the same size, which we must add in the corresponding engines.

Again, similar, we come to scale product. We simply multiply each element in this method by the same.

Now, the same scale, what's more important is matrix multiplication. If xx is a n by n matrix, why is the empire c matrix? You see the inter dimensions match, right? Then we are able to define matrix multiplication, x times y which will give you z where z is an by c matrix. The bank hazards, because it has been sucked over where the, I guess, entry in the matrix is defined as follows. The sum of k from one after nxik plus yk change. What's this? Any others? We have defined this in the previous slide, in the figure slide, it's in a product. It is in a product of water weapons. This is this is critical for you to understand the matrix multiplication. The zig is the inner product of the I slope of the first matrix and the ps column of the second. How to compute zig the first matrix, you have the ice rope of x and the second matrix. That's why we have the gist on its inner product of those two vessels, which will give you this engine. You can understand matrix multiplication through inner product of two vectors. Now, let's take a 10 minute breakthrough or the resume I answered.