### ML\_week3\_lecture3\_1-20240920

说话人1 00:00  
Is being extra videos is level and one questionnaire and where they come from to

说话人2 00:11  
before that.

说话人1 00:12  
And so on. Or concept or comments, position. The same nature are going to

说话人2 00:19  
have

说话人1 00:21  
it first nature for the machine

说话人2 00:25  
and the 4th final floor

说话人1 00:28  
and the

说话人2 00:29  
local in china.

说话人1 00:30  
The most function is so that machine that is not just as function is in addition to pressure is higher, what are these things on the europe generated person screaming last time? The best the beijing, baltimore, africa and the 94. So when I mean lga leader is agreement is as it is, generally ii because it's at least for the germany, pos student posi

particularly the general class stars, learning more of the joint as commission skills, given this is a high until one is cheap, right? It's a one time to building for this customer or the enemy.

Always, this will still probably be because I want to

说话人2 01:38  
make prediction

说话人1 01:39  
based on the by maximum is

说话人2 01:43  
just your

说话人1 01:48  
based on direct estimates to yp and x you. Because you don't care about the past conditions. You don't care about

说话人2 01:58  
the paper

说话人1 01:58  
generating courses, given the class paper, how to generate this picture and how to generate

说话人2 02:07  
object

说话人1 02:15  
just refresh. Is it basically discrimination?

Direct models is doing about using a

说话人2 02:25  
linear function is equal to a couple of times?

说话人1 02:29  
Suppose that I think after this, they read that term and be

说话人2 02:37  
the best of

说话人1 02:38  
this. She's for transports very transport. So fundamentary, the

说话人2 02:42  
world is in the product.

说话人1 02:48  
Poy is equal to what is the one given x is similarly sigmoid of that. We know this special thing. We know that, which is, and therefore, the norm of the racial is equal to the inner function. And the normal, this issue is cheap for you as a motive. Not just

说话人2 03:21  
here

说话人1 03:22  
is a cleaner function. And the proof is that is straightforward. We have poy poy is equal to one. Us is equal to this. Poy is equal to n to one uns equal to

说话人2 03:43  
e

说话人1 03:45  
to the power of each is function.

说话人2 03:49  
Go to china,

说话人1 03:52  
divided by one

说话人2 03:54  
us.

说话人1 04:01  
And this developed this. Maybe you don't know the castles.

Therefore, they are here for the time it is that one is make money is equal to one divided by this is the group remains as long as the negative for this. Not it. This is equal to that function of

说话人2 04:33  
it's

说话人1 04:35  
not b because yesterday design in the innovations, if you take the laws, you take the laws, the law is, and you are done in a function. Any questions on this

说话人2 04:57  
simple evaluation?

说话人1 05:08  
The in answer patient function? If for this relevant? Again, this is on month c is in the respect for c is you ask that poy is it the next maximize

说话人2 05:24  
the posterior

说话人1 05:27  
course? So the ability is more than the signal function.

说话人2 05:31  
It seems the

说话人1 05:36  
question of this regression. So based on this module, right? So you have the equipments between several thing

is in the world, right? You

说话人2 06:06  
have

说话人1 06:06  
this is bigger than zero

说话人2 06:08  
of

说话人1 06:08  
quality, this response to p for why is it? For one payments,

说话人2 06:14  
right appeal,

说话人1 06:15  
right? With one payments

说话人2 06:17  
they do not

说话人1 06:18  
want, which is not for us as a community on this module expression.

Now, comments of musician, so much of the machine learning can be reading as an alternation called before. We are dealing with, in general, reform that they be as a

说话人2 06:48  
minimization

说话人1 06:49  
for by the meeting, my with respect to some primacy. Cf is a trying to venture law, and the most function is defined over a set, the chinese set b with respect to the time that the scene. This is equal to the minimum with the zeta, one over n sum from s one over n this is the average over the most function.

说话人2 07:23  
For

说话人1 07:24  
individual examples in this 27, each individual most function, as we see, here is t is a financial incident consist of the

说话人2 07:39  
examples.

说话人1 07:40  
Here we have another r which

说话人2 07:44  
is a function

说话人1 07:45  
theta, so better rather to prevent over patient to and we have already come to 1 type of realize. One is for five when we call our mortgage regression, see the passports. It is also we can sit on. It was too long school to see. And this is simply for the sun of k on one, for the cardinality of so the lens of the sea not see the trees for.

说话人2 08:28  
This is an answer is

说话人1 08:30  
that those are the equivalent terms. The transportation is equal to the l two conspire of the same and the sum or over n squared. Any question on this slide? The initial regression, right? The country is a bit better as a faster, and the competitive rating as a unified kind of the procedure.

Therefore, the north for the individuals, animals, isn't it just right? Not so just. No logistic loss is more one aspect function of the z here, we define z as y as a label

说话人2 09:33  
times.

说话人1 09:34  
Is it as a speaking? Right here? Predict function is a linear function. That's why it must be. As example, for the superstood, man come from and is the from squares while saying that say you have done. And we both thought they had plans see into this term markers and remember all defensive. And his cases. Is it a bit better for any questions about several types of populations? And you can have class cooperation as a legal operation, some organization or have a reasonable as comments or commission, which is in the case. And it includes not just regression in a regression for support for information, is then as a special reasons. And the non condensed, which is half in general. And it includes because so the very first concept in complex operation is 9 seconds, 9 seconds between 2 points, x one, x two, consist of 4 points,

说话人2 11:19  
x

说话人1 11:20  
which is equal to r five plus one minus rx super alpha less than 01.

So ask is straightforward. And how do you define in mind statements in companies or where alpha is equal to one? Is equal to 61. Alpha is equal to zero. Is this two. Alpha is strictly in the chain. There is a quantity chain based on the concept of nine savings we can. And why convex set, the convex set, company's mind statement between early 2 points in the set. So beijing is not, we randomly sample 2 points, x one, x two in this set. Pretty good test. You randomly specify in alpha, which is the all of them, not the number equal to zero, and this time equal to one. Then if this is a complex set, this must of x one + 1 and ask for part two, also 90 %, not. So definition,

说话人2 12:56  
of course,

说话人1 13:01  
for comment set, we run on some of the 2 points. This nine segments, the also man is set something that

说话人2 13:12  
is

说话人1 13:13  
he probably, and this is the mathematical working

说话人2 13:17  
with.

说话人1 13:17  
So and here we give three examples. One is a complex set, and the other two is not. Here, this is a complex set it. For any 2 points, they said they're not saying that is always set, but this is not the case what they said, right? Given this particular points, and we create a 97 for some points in this 1997, our company need to set, right?

说话人2 13:57  
Some

说话人1 13:59  
are company to set, but the other are not, which means that this issue not a compensated, by its definition, compensated here, right?

One more dealing with a close set area is an open set for this particular once. This 19 plus open recession, it is slowly, it's not connecting. That's why it's more complexity. But if we connected with this war, have this four line segment into the set, this becomes the

说话人2 14:41  
economic set.

说话人1 14:58  
The definition of hyper plane and half space, hyper plane is a set of this form consist of cortex where a transpose x is equal to b where a is not equal to zero. And a is a normal vector. When we talk about union and logistic regression, we have booked on this, is this with a difference from the future of peace? And we have some space. Instead of ones, x not satisfied, a transpose x is this sample or equal to b where it is of zero. We come, which consists forms the ct the so higher plans are half space upon. My question is, I hope, to, let's say, the half space to the company set, right? To prove it, by definition, given 82 points, this defines listed as some profit x given two points, x one, x two. Is it also in this set and we specified in our farm?

Obviously, our farm is good on. If the 01 is on equal to one, we verify whether alpha x one plus one minus alpha

说话人2 16:55  
x two

说话人1 16:56  
is also in the same by examining whether it is equal to those any transfers points in this 9 second. This is equal to

说话人2 17:12  
alpha

说话人1 17:14  
a plus x one

说话人2 17:16  
plus one

说话人1 17:17  
minus alpha. Since they explains to behind this set, right? This is nissan. We go

说话人2 17:27  
to alpha.

说话人1 17:29  
This is nissan. You need 1.5, but

说话人2 17:36  
is

说话人1 17:37  
this?

说话人2 17:41  
This

说话人1 17:45  
is which means that this said this . also that is exactly also

说话人2 17:56  
company is

说话人1 17:57  
that and we are done. We have available is a common sense since x minus two arbitrary training. Hong kong are very true. Right? Any questions on the group?

说话人2 18:24  
Through?

说话人1 18:37  
We tried to ask for is that t complexity was set. We'll apply that nation. We have read that a couple of

说话人2 18:48  
examples

说话人1 18:48  
in this slide randomly specified points, they said, is up to where you are from that they want to prove this point is also in this point, 1909 segments specified by this x minus two, is also in itself.

And we can shoot out, right? This set is open from simple complexity. Have the planes have exposed by all bridges, in fact, is very complexity.

说话人2 19:24  
Okay?

说话人1 19:24  
Especially provide function and so on. So if you want no more operations, such as the capacity, actually is hundreds of issues.

说话人2 19:36  
For

说话人1 19:44  
now, we have the definition of how to accept it. Let's move to the definition of the common function. Convex function is a function from rn the units is ten dimensional space. From r the output is the real number, is complex. If you decide to have two conditions, first, the input domain on the set, second is satisfied. This be included. If rx one plus one message two is designed to our market x one plus one minus one eight, and steps to for all points, the same domain.

And for all alpha, that's between zero and one, that's the emission.

说话人2 20:41  
Response

说话人1 20:43  
of the function has two foundations. First, don't forget the first foundation. The input domain is a complex set. You will look at this complex set. The second sets that is equal. This is a typical earthly complex function, which is profile. This is a typical, long complex function where we can confirm the local, optimum, or verbal meaning of, which is, notice forms

说话人2 21:26  
over the whole function. We

说话人1 21:27  
have no home anymore, which is the lowest points around its neighborhood. We

说话人2 21:36  
have this sample ones,

说话人1 21:38  
which has the same first order conditions. Ingredients of this point is zero. The center of a minimum for the convenience of reading is also there, but it's never a lawful, minimum law or global, meaning. In some direction, for example, in some directions to stop, but the gradient is zero. We get when they have to might be not to be universal.

For most of points, it's really easy to learn the sum of points. Instead of because you have so dimension in some dimensions begins a lot something other than to be profound.

说话人2 22:25  
It's kind of wants that.

说话人1 22:27  
So f this the concrete function is matters of its public function, which is not. We certainly believe this in

说话人2 22:44  
order

说话人1 22:45  
if this is written out of this as a complete function, if this is a lesson based on the complex function, but they both should satisfy the first foundation, the input set and the input domain become accept. Now, it's, yes, this visual better on. Here we have. This is x one. See,

the corresponding points in the map axis. Right? X one. This

说话人2 23:43  
is

说话人1 23:45  
from the and it's that one more in between this are from the x

说话人2 24:03  
one plus one

说话人1 24:13  
responding out to this

说话人2 24:19  
formula are

说话人1 24:21  
from x one by the one as hong

说话人2 24:24  
kong.

说话人1 24:26  
We just as 3 points and three for one of hong kong

说话人2 24:35  
problems.

说话人1 24:35  
And here is interesting happening to draw a non statement between these points On this .. What's the corresponding of for this .? So here, these points should be alpha or a plus one plus one minus r if is the power of something, we have this this this is greater than this like this much. And if you can compute this distance mathematically, to verify, it's really this offer. Right? You've been sitting in the travel stuff about that. We can draw this zero and nine. Here.

We know the distance of baseline, everyone, the distance of baseline.

As soon we know the distance,

说话人2 26:14  
this

说话人1 26:14  
123, and a little computing business here for and use this. We will find we do this, and we will find we have four classes,

说话人2 26:29  
which is equal to this line.

说话人1 26:32  
According to

说话人2 26:37  
similar triangles,

说话人1 26:38  
i'm not simple geometry assignment you covered up by yourself. I worried about this one is really this one. Right? That's a visualization of this inequality. A in a kind of time. Any questions on this? Here are the examples of complex function. On our one. We have the one function, ax plus b with adrr we have that financial function, u of asaa is the number of power function xa that's r plus plus means, as extremely positive number, is strictly equality number. R is a real number. R us is a non negative number, which contains zero. R plus is strictly for the design. For r minus bigger than one. Is that absolutely horrible, absolute value for peace? We can come on. We have had two entropy. It's not x 15 of us.

This is called invention, right? And what central entropy is? X log negative x or not x this is entropy. That entropy is a convex function. Then entropy function is a common d function.

We often need, as always, be medium as a complex function, and the map is negative computer.

Therefore, in many circumstances, you may heard this description. I go to the maximum entropy, I go to the maximum,

说话人2 28:51  
come to.

说话人1 28:54  
That's why it's easy to do

说话人2 28:56  
not.

说话人1 28:58  
And we have a complete function to find ours in this french. We have no open function, not just missing our restaurants. And here I will call the attention that, all right, the final function is both complex and complete. Do we have other types of functions that are both complex? Or do we have other functions? So the answer is absolutely below. The only type of function is not a host on websites. The higher idea is that you can see is from history. For comments, something if you have occurred to, if you do, you have this could come up with you for combat function. If you pass the, put down on it, those comments complete straight up, have a point. We won't provide it the last the only function that is supposed to comment on the company to

说话人2 30:27  
find out.

说话人1 30:33  
We also have complex function in where the information is the international space or in pattern dimension

说话人2 30:44  
space.

说话人1 30:46  
We have this defined function. We have this small function to tell complex function more. We have examples where a function takes a matrix estimate. You don't be terminated by this matrix as you can consider a matrix, as a bachelor asked, the best classes in 500. In by end, by one way, they flattened into a condom. Just a great question. You can define the function temple x which is equal as an adjacent to the idea, except it must be. And this summation can be rating competitive as this. And this is known as the trace of it. If you don't know the trace of matrix, just google it for us at 50. And we here we have, let's say, the trees, if a square matrix is simply sum of its stand for me and start the trips, if you have an by n measures that this is the sum of the animal and try to figure it out why the trace of a transpose x is equal to this summation. I remember how to define matrix multiplication. Could I just entry is the in your product and as role of the first matrix, just 1 %.

Then the definition in the product, you combine these things to get it is something. Here we also have this spectrum law, which is not a singular value. Spectrum, is denoted as this one here. We won't write this, as they are too long, is for better. Now, we have matrix as in this is for the scripture normal. That is matrix. Yes, you don't like this signal, the maximum signal, whether that's which is equal to the maximum icon value of x transpose x and we take the square root of the root of it.

Here, I show this example that would be force. You may think this has a very complex time, but it is a complex one. If it is a complex one, it's easier to deal with it, to minimize this function with respect to the input.

说话人2 33:57  
This matters.

说话人1 34:07  
Here is an example, the right function, how to prove this simple, correct function. This context right here, we only remember the definition of proof by definition.

Now here is a group with, when I say 2 points, when I say, how about this is between? And we put this our expert on that last words. So you guys should be able to regard for this detail.

说话人2 34:43  
Simple art.

说话人1 34:50  
I'm so similar for this absolute function with them. Use the final imported to.

说话人2 34:56  
All right.

说话人1 34:58  
This expansion. So they to prove that the that's it. Your exponential function, by definition, it would require the taylor expansion of this ex which is equal to the sum of n from zero for the unit. For this is known as and

说话人2 35:25  
that for

说话人1 35:27  
the fire holy 71 us to has to

说话人2 35:33  
turn back for

说话人1 35:33  
1 × 2

说话人2 35:35  
× 10. So I

说话人1 35:40  
have ask you to figure out these details. Right? So pay attention to this change of this study that's from 01. Here being back to zero. And we use this how to expand national reversal, the properties of commerce function 89 seconds before between 2 points, mass of our growth, and every more convenience of global meaning.

That's a very powerful argument. Every normal meaning is a global meaning. Why are going to have a formal mathematical are not too hard to understand code in the next road, in history. And hopefully the less nature we're going to derive or prove this together. If you have time during the lecture

说话人2 36:52  
with

说话人1 36:54  
this argument is why to make the comments on the population easily. It's fine to find a local meaning. And we know about this global meaning that's done, right? That's why it makes complex or local meaning is available.

Properties of commerce function, none native compilation or complex function are complex. Here are my beyond total numbers. We have a fine scanning of comments, functional comments. The . was maximal. Complex functions are complex. Here is an optimistic respect. And to almost done this with opinion, draw a simple figure.

说话人2 37:58  
This is fx this

说话人1 38:07  
is the mark. The . was not simple. Is it here? The point is x gx is greater than fx we retain this. In this part. This should be a problem. Other composition of complex function are not necessarily complex. Your levels is like this, give your levels. We have a simple example. Let's ask this question. Can you give me an example that simple examples of the composition over commerce on the role of general commerce?

Here? Let's see, I will give you one very special as an example. Fx is explaining function. You have put up each comments on, right? How pox is negative x not linear function, linear functions, both comments and computations. Right? Let's compose it. This becomes negative ed of x this is a complete function, because you put that sign from instead of combat function. Right? Compositional comments from the general comments. And here, we not, we can make it complete. In general, it's neither convex nor complete. It's not convex. We have first order condition for the common function. If f is differentiable and its weakness, respect to x is defined as the partial of x respect to x one. Part of that, you are at the next one, and it's an income better. This is read as a gradient of fx with the facts, or this symbol is is may not be seen as not like this.

And if we know that f is differentiable, the first order condition for differentiable f with complex domain, right? For function to become as the first condition of it has complex domain with comic coming. If i'm only this fy is twitter now in fx plus, we don't know whether it's possible to find an answer. X usually, here, this is fy we expand use taylor first order taylor expansion of x the first order taylor expansions, or is why at this ., as, therefore, the first order of the occupation of xy is equal to ii don't

说话人2 41:57  
thanks last for

说话人1 41:58  
not long

说话人2 42:01  
as

说话人1 42:02  
if it transpose one minus x not first order pay the expansion of the one xx images for I previously mean only see the period expansion for one become. That's what we made. Here is the universe. This means that the first order in a script approximation for f is a global underestimate

说话人2 42:31  
return

说话人1 42:32  
for which is reasonable, right? Because for comments, function is profile, we use the straight line or have a plan to approximate this line below. Make this comment on.

If we also have a second order foundation, if there is question,

which means on the patient matrix, which is unbalanced, symmetric matrix, symmetric matrix, the ids entries, the second order part of the world with the xxi and xj if we have passed this phase of that, with complex domain, if this f is convex, this is real as if i'm only the haitian matrix is. This is not greater now able to. This is + 711.

So if half past differential f with complex into remain is a compass from if i'm only it's haitian matrix is paused by the same year and then we have a new concept, right? If we don't know if you haven't learned the definition of quality semi definite, here is the definition. If the matrix is a square matrix is positive and semi definite, if i'm only for any here is for work for all x we have x transports ax is with these important numbers. The court will send definite or graphic is such a reservation of the positive positivity of the number. If we reduce the dimension to one by one. Therefore, x transpose, it is simply a a times a number, right? Which means this is a experts should be greater than or equal zero, which means ah should be greater than or equal to zero. Positive senate, that means simple generalization of the productivity were under in one b to any party.

Therefore, to prove the convexity complex convexity of the correct function, they take the second order to relative, which is a positive number. You got that exponential. Take the second order derivative, and we are done for logistic loss. Right? We in the early slides, we see that logistic regression response to the comments, population power, the logistic loss, you will take signal derivative is strictly greater than zero, because we're working on one, right? We can see this number is pretty important. We can also see this one by one matrix is safe for the fixed impairment. And you want to be reduced to this number is given to. But in inclination space about 4 % 7. So great, decent are more. The humor learner goes to optimize the loss function, finding english learning problem for time for explanation. We come up.

So this emerging learning is medium as a complex small city. And here,

说话人2 47:09  
right?

说话人1 47:10  
Already 2,000 for that functions of congress function. And if we minimize the progress companies respect this partner on the basic problem. And here we are really recent or gpu this is the following update formula. You have sit on is equal to or is updated by see how manners are from the region they always been seen. Alpha is only made. The we know about the novel signal is agreements of your signal evaluate. And here is the example we have. Lc types for graphical agreement is to set up. We learned is this. Here you see for one.

And then, according if you apply this, are basically you have two,

说话人2 48:17  
So on.

说话人1 48:18  
And here we draw this figure,

说话人2 48:22  
forms occurred,

说话人1 48:23  
but they make the update on the space. Gd first choose an initial point. If you go to find she stayed up and repeat the theta at the t plus one inflation is equal to theta at the t minus Alpha is under the root na, na ol the most damaged defense over the 2017 hundred and five theta and statistic equation. This is equal to theta, t minus Alpha, and Irish of the grievance evaluate.

In particular, as example, as the teeth parameter as so how to stop? When in general, we have to several ways of stopping the nation. The first, if they are true norm of the difference of the conservative managers, is this are

说话人2 49:48  
or is

说话人1 49:49  
this are equal pretty much issue? The bill too long of the grievance of the most function evaluate at the counter prime estimates. The game is based on our group, is last year. And we do that. It looks normal. Theta is square root of the sum of all enemies squared, l too long square. When we get rid of his spirits, right? It becomes some more or Indians. Good. If seven is a small constant, it was said to be ten to the left or six or 10 to the eight. Let's take any break. If you resume at 10:00, 30:02.