Julio, can you hear me? Yes.

Yes. We can hear you. I guys have any questions regarding Umak or anything else? No? It's good? Okay. So now let's get started with our this lecture with Today, we will talk about the logistic regression model and the list of scripts.

Hopefully, we can cover all of it. Last week, we briefly mentioned about the differences between the regression and the classification.

I know we talked about the linear simple linear regression, the multilinear regression, and, the binomial regression. We also mentioned several metrics regarding the regression problem and the classification problem.

What are how do you use the metrics for the classification problem? I remember? What kind of metrics that, we quickly used was a classification problem?

Today is not from previous class. And here is outline that we will briefly mention about classification problem again, and, we will try to answer the question why not use linear regression for classification problem.

And then we are very good at details of a logistic regression.

For example, how the how this is representation And some cost function, but just the regression is a gradient descent and the regularization, and then we're gonna talk about The regression.

So first, let's review about the classification problem again. So, typically, if it, is a binary classification, then we have 2 classes. Right? We will use 0 to the negative class and Google values 1 to represent the positive class.

There are several applications. For example, in your email, detect whether it is a spam or not spam. In a medical part of also, we want to identify whether it is a tumor, whether a Alright.

So we want to identify whether it is a tumor, whether the tumor is, American or the benign. And for the transaction, we will detect whether it is broad or no broad. Right?

Yeah. Several examples about Classification bio classification problem. As a review about linear classifications that we discussed last week. So in this example, We have so many different, lines here, and those lines I can't what?

This this nice are what? Yes. It is a regression line, but we do it for this at the pipeline or we call it to call it You see your boundary is fine.

Yeah. I mean, the left hand, we actually give 1 class. You read it on our side. We actually do that as the class 2 here. Right? Why? We mentioned the problem of it is not, yes. That can be Yeah.

That's the one point about here. So, actually, we work out all the detail. What kind of new districts? Distribution? Okay. Yeah. That's gonna be another I'll set up for it. Right here, we not one is a linear regression.

The history is that linear regression We are being affected by some outliers points plot. So let's say more details of it. So now give an example of the tumor size and the malignant to see whether the tumor is malicious Also benign.

So now let's say if we want to use a linear regression, you feel these data points Where is the log? The line should be this way or should be this way?

Should be this 1, or should it be this 1? Which way? You see it this way? 1. Right? So, actually, so you guys need to notice that. But in your regression, we actually want to figure that once.

We want the data points that are close not the line are close to the data points. Right? So, typically, we will use this as the linear line That I cannot point to right because it defeats lots of points. Right?

Because this is the regression line you guys made, can calculate to fit those given datasets. So let's say in this case, We had defined as railroad, right, was, linear regression, and, I just did the s here, miss the to predict as 1.

If not, we will just predict as 0 here. So let's say sphere is a spherical 0. 5. So which part is net 2 and which part is? It means this party is is positive? Do you think this is the ad hoc team All that you part of shadow.

It's kind of negative. Right? So now let's see more details. So, actually, we'll be able to use this line as the handoff of our decision boundary. Right? So if we get below the 0. 5, we recognize this part as negative.

If it is bigger than they're quantifying those parts or those core points where we are saying it is quality. So in this example, we may think that this behavior, where should we be to norm for it to buy?

And Go to the right. Right. Because either can successfully separate it. It's to give you the classes. But However, let's see another case. So this is similar as before. Let's say we have another outlier. So now for linear line, Yeah.

It's for this guy this time. It's nice. But be close to this point. Right? You guys learn about linear regression. Right? Right? So now in this case, we're kind of well, this 3rd line should be close to this point.

Right? Because, you know, one fits all the data points. But now did you notice something different Here. So let's say we still want to use the previous 0. 5 as the. So no. We're chatting.

The answer is here. So if we feed the best from the regression line, The scale wouldn't be not to decide any point by which we can differentiate the classes either way of putting some positive class examples into negative classes.

So in this example, what 10 points will be pushed to the left team In this case, I'm advised. So now 7. 7 points? This is 4, and you have another history.

Right? So, actually, you will put these 3 points where Previously, it should have been in the positive classifier. So in this case, here is the DCM boundary. So in this case, it will run move to the negative points.

So in this case, the linear regression did not do a good job. Right? Because this is the key reason why We don't want to use linear regression, but classification problem, because either we are affected by outlays a lot.

But now Yes. You can point? Okay. Good. So let's let's talk about logistic regression. So logistic regression is one of the basic and the popular algorithms to solve a classification problem.

So let's say in the future, you can review and ask a question about what is the u usage for the Not just a regression. It is about regression problem, or is it look like classification model?

If you say the but just a regression is very important question problem, it's known as a logistic regression because its underlying technique is quite the same as linear regression, And the term logistic is taken from the logic e a logic function.

That means they use this method for classification. Okay. Good function. So after x is between if you want, 0 and y and the y should be between Use it?

So let's say, what is the range of the margin? The logic function. What is x range? X range is That's the entry. 0 to 1. Right? So what about the x range e of sigma function?

Okay. So now let's go review about the previous classification. So even in final classification, we have one as positive and 0 as negative. So is that the case? Let's think about the previous linear regression model.

So Is that right? Okay. Yeah. So, Zelle, what about the logistic regression? So, actually, we want the output is between 0 and 1,

Any guess? So that's what we do. So that we can gather a interval between 0 and 1. Any guess? No? So previous, this w x as p can be any number. We can do what so that it can be between 0 and 1.

It's kind of so, actually, once I point out, you guys can't remember. So this is very about this And this function is the so called sigma function, or we call it the HABDA, not just the regression.

So I think, up to now, you guys will No. What do we needed to do to apply for is, live at addition, what's a single function for linear regression.

Right? So that I need to change this in between adjust the regression. And then eventually, we have to make it as the h x here as the one divided by 1 plus e minus the c negative times with x.

But this part this part is really about the z y, but the z is dominated in this sphere about the z g x. Right? Let's continue.

So let's try to interoperate, hypothesis output. Let's say the actual x that is equal to the estimated probability. That is y is equal to 1 on an input point x. So give an example. Your x is equal to the transpose of the x0 and x1.

What is x0? How do you guess about x0? What is f0 here? It is can the file issued by the way, you replied as a box. So it's it's the x there is to find a box. So previous but in your version, we may have, x b.

Right? So this b, the bias so, eventually, the b, we are can be incorporated with x. We combine them together. So that we will find that, You know, waiting. Your first dimension should be, like, 1.

So that is the box. Give me example. Let's say h of f h is theta x Yes. So actually so there's a piece that you will can you can tell the parents is that 70% of chance that tumor will be manicured.

Right? And that has more details that we try to integrate as a logistic regression. So suppose the predicted is y is good to 1, once as you see the x Is it bigger than 0.

5? Why is this is z is bigger than 0. 5? Is this correct? Y. Why is he should z should be bigger than 0 here? If y is equal to 1, So y is less than 1.

We try to predict the y is equal to 1. You have to make sure you like, c is bigger than 0 because as you can see from the But directly, right, if it is close to 1, then definitely your the z should be bigger than 0.

Right? Okay. So let's, to view this function again. So k. Now we have x x1, xi to x, capital I. So we have a mass of amount of weight that equals w 1, w I to w capital I here.

And for the linear regression, you guys also notice that we we usually have, as a sum of all the weight then plus this plus b b as bias here. Right? So that we can get our panel of g z, and z is what?

They used to be just as a single mode function. Right? So the opposite, you can't get this. What's this? Properties. Yeah. This is the probability. What is w and b here is what? Yes. And we can also call them as what?

Together, we call this as what? Parameter. So this is the great advice together. Just to repeat. Repeat. Just about permeability of x in the category c. So let's try to see more about the decision boundaries.

So let's say, for example, the h x is a to gc0zeroplusc0onetimesxoneplusc0twox2. So here is more example. Let's say if the c zero zero is equal to negative 3, and the c of y is equal to 1, c of 2 is equal to 1 here.

And then we try to predict y is equal to 1 even this cannot what? This is a button. Yes. A function is more like a what what kind of function? It is either line or circle or what. This is what? This is what?

There's a line. Right? Actually so this is actually this line that we brought out here. Right? So you put this bigger than the arrow. It should be the positive. Right? Even in that time, you should k. It's about another example.

So let's say if the h x is equal to g c the 0 plus the c the one x one plus the c the two x two, Then press the theta 3 x 1 square and, plus the, theta 4 x 2 with the square here. He has more details of those different parameters.

That's about this. So then this should be If we practice this different numbers to to it, well, we are getting well, we get here. Yeah. The what? Is what? Okay. Yeah. So you might as well get a lot of zips, and this is so called Circle.

Right? The circle here. Right? Good. And is that our point, it is the. You side is the. By measuring. And the further we can extend the and you see that there can be more, like, from 0, 0, 0 to 1 The c l 6 blah blah. Eventually, right.

So maybe we have product kind of function like those, and then we can predict some kind of point like this. Right? Sima, in this case, you may have question about why is this logistic regression has so many different patterns.

Right? Where does this form come from? So here is the overall representations of the logistic regression hypothesis, And we have this, like, at c l zero plus c l one blah blah blah is to the c l n x n here.

Right? So you're wondering this problem. How we can get this kind of functions here?

So let's try to consider value from a function f. That is mapping from x to y, and the way x should be a vector of real a very the features from x, man, to x end, And the y case, what it function is is just 0 or is just 1.

So assume our x I are conditional independent in a given y, and the model xi depends on the given y, is equal to y k as a function.

Is this one. Right? And we also try to model p y, the probability of y, that's only function as pi. When I use p y is equal to pi. Right. So let's say it was going on about the probability of y giving x, y, and q to x n.

So first, let's try to, pass this, visual rooms. So here is our probability of y is equal to value given x. So I I I'm trying the Fisher rule that we can guess here.

Or should it be the probability of y is equal to 1, then Hamzow is the probability of x giving y is equal to 1. Right? Can I add my right to the patient law of this one, boss? But what is the best result of this month?

It doesn't matter. Yeah. You had just the right. What are you using about the visual human? Of this one. We got not because you have all this. Just about Right. Is this correct? You agree with this again? No. I have property.

It should be what? I have property. K. Why the bus? Yeah. Okay. Why the bus? Yes. But yeah. Finish. So which Huawei did Which one are you aware of that? As your vendor, though, your mind? Which one? Either. Which one are you aware of?

So, actually, let's try to compare with those 2 again. So this is the p x for beta of x given y, it could be 1 should be just this. Y no. 1. Right. So the arrays can be separated in this case. Right? 1 and 0. Right? It's correct.

That's got it. So then, next month, we need to try to divide by probability of y is equal to 1, the probability of x given y is equal to 1, so we can't get this kind of error. Right? Why is it trying to divide divided by this one?

We get This case. Right? So next, we will try to apply the expression on that map and, also upon the log group. What does this mean? So you were getting bots? So let's say if the line with x here should be equal to what?

The sem y, Right. So then we try to just apply this to this case. Right? So we have this experimental map and the logarithm part here. So next, we try to separate here, right, with this. That is this part.

What? As previously, as I mentioned, that is the this p y is equal to pi. Right? Let's say it's a. So that that's the reason why this part the first part, we can say this is pi, and this is the one minus pi. Right? Probate.

Right? And then let's try to plot all the cases above all I in this case. Right? Because now we have this column here. So, eventually, we'll try to plug in the probability of x I given y, so that is really about our partial function.

Right? So I will not expand too much here. But, anyway, so, eventually, this party just means here. And then this party is so called the CI that you want to mention. And this is the eventually why we can't get lots of. You get it? Okay.

Okay. That's good. Let's try to talk about next part is about the cost function of a logistic regression. So let's say if we have a telesat with m examples, x y x 2 to x m, and then here's x Should be should they have x0 to xn again?

X0 is 1. It's a bias again. Right? And there's a y should be in the either 0 and the 1. And, actually, the x is just above this So now the case is how can we choose the parameters, see it right here?

As you already you What function? That is this is what? I just say that Xi is what? This is what? So I mean, So everything from admin I do see that as x I. What is this? From linear regression. So this is actually my output.

Yeah. It's the output of a linear regression. And this is the y I, is the branches of the. And this is the sum so that we can calculate the cost and the we can donate it to cost as the this is the gradient.

So eventually, we can say the cost I just see the x. So y should be equal to 1 over 2, then the bar of this this is a linear regression. Right? What about the the just the logistic regression here?

So actually, this is the function that we defined the what about just the regression? So if y is equal to 1, we will use negative log, and you see the x. If y is equal to 0, that we will use negative log.

1 minus, And you said x here. Let's say in this case, if the cost is e equal to 0, For example, even let's say h x is good is glued to 0, if it's close to here, now the cost should be Evening in your mind.

So how can you draw this kind of plot here? It should be this way or it should be this way. And, guys, how this line how this line changes you in this part. Is close to 0, is close to 1, and this is Yeah.

2 guys for for you guys. This month or this month? I don't have blood. You know what? So which one you you cast now? That's the one? Are you guys I believe it's mute. Yes. Confused. Confused.

Okay. Because why confused here. Let's say if y is equal to 1, the cost is equal to 0. Could be here. Right? This is why this is why this prediction. Right? This is a prediction. Prediction is equal to 1, So that cost is 0 to 0.

So then you have one point a year, but, obviously, if Hash x is good to 0, so that cost this cost should be given it. It should be you have these 2 points. Right? So that's naturally, it should be the first one.

But is that correct? Why you have these 2 key points? One point here, one point here. Should you done this right? And that's good. Good. So this is the part of the path function here. Professor on mute.

Can you hear me now? Oh, yes. So let's say if we try to combine them together, so eventually, we will form the overall loss function our cost function for logistic regression as I can see here. So why why should we do engagement?

So this is, Why we want to use use 5 minutes? So let's say the y is equal to 1. This equation is equal to what? Y is equal to 1. This power will be done by so you should be just in that keypad, and it's still the x.

Active log. 5 minus x. So it should be the same as previous one. Right. And, again, it's, overall That's, just a regression for this, thought function. So eventually, we tried to plug in previous just one.

Right? I tried to plug in with m examples. So it should be I just see the x I, y I. So you mentioned that the first the previous that we mentioned is The same from IHF eventually about Harris was a response.

And this is the case that you guys need to learn. Good. Let's say, join the free join the learning, which I consider parent the sitter, and then it's trying to minimize, here, j here.

Then general prediction, you give a n u can x. We tried to output the x I use this budget to that and we try to directly grow, but predict the appropriate evidence.

What about our our high students? I guys have I mean, past years is so far. So let's talk more about where does this cluster function comes from in another way that's Try to use, not to lack of let you see it.

How can we get this? Similar as before, we have examples from x y y, y y to xf yn. The maximum likelihood estimation for the parameter setup. So that means it should be set m l e here so that we try to x miles.

I've kept the same thing here from the all the probability from x one to x and y our end here. And they that can be donated with this part. Right? That can you do the product from I to 1 to the. That you go. Right?

And as for the next goal, conditional, not the likelihood estimation for the panel to see that here. So let's say our goal is what? Our goal is try to choose the parameter theta to maximize the conditional likelihood of training data.

And this is, like, the probability of y given the x, and this is probability of y is equal to 0 given the x here. And this is just typically, and this should be one minor. Right? As you guys okay. What about our data log?

Likelihood, you should be also, like, some continued project from I to 1 to m about all the parameters of p c tag here, all the points, and this is the conditional likelihood that is of And here is the overall stuff that we tried to optimize.

That is, maximum conditional look at the likelihood. This is about here.

So as for how we can get a close phone solution, we will try to discuss more about expressing the conditional, but the likelihood of our list. Pricing will take us a lot of all the products So we got 12 for the county guests.

I still try to tend to log all the for that. This is what? Then we have to come to the sub. Is that correct? Right? We cannot change it to the sub. Right? So that's the reason why this part changed to this one.

I got it. And, further, we can donate to this The final equation about heresy here, they can see about the first this part is referred to what? I say why is this why is it equal to here?

Why is the probability of why human x is equal to here? You should notice that what by the why? The why we have how many it's we just have 0 and a y. Right? So that's the reason why it can't expand file.

This This should be this should be 0. What? This part is for. This part is what? It's by what? And this part should be working. And this is similar as before. Otherwise, we try to get, this kernel cross function, but y is equal to 1.

This should be, like, negative. Next, point of the one time, try to discuss about the just the regression using gradient descent method. How we can get What is the GSC right here?

GSC what? Hello. Jessica here. What about good news? The good news is that it is a convex function. A convex function is what? There are no proof of constitution. We have to use the pre designed trigeminal, approximation.

These Numbers. Right? So that is remember about this is the key key point about gradient descent. Right? We have I see the g I here, and that is equal to c l the g I minus alpha.

That is with directive of the, Jay, I see that. Yeah. We we respect you to the h see the j here. Right? What about it? I think that's a problem. What is At the next step, we try to simultaneously update all the parameters theta.

So here is this about how we calculate these directives of of this bare button. So let's continue about it. So that that's when we try to plug in, it becomes this one.

Right? This is correct. Okay. The slope? It's kinda max slope. It's really about the levy rate, how fast it is. Let's try to remember what is the greatest and linear regression. This is the previous for the linear regression we got.

And he has now we have a logistic impression here. What? This part? The last part? So this extra last post here. Right? What I think is that what is I just see the x. And you see the s? I'll do the step.

But I need a revision on them, but just a revision. I'm just here. I remember, we just assumed this. But exactly exactly the sigma of the functions. So that's the reason that key difference in between I can see the x function.

So previously, it's just about, any accommodation, but in this case, it's really about the Yes. So CLT is correct. So it can combine both the weight and the bias here. Right? So that we call the safety.

But always remember, in the first foundation, we have parts here. And, this is function for the logistic regression. So this is the key difference between linear regression and the logistic regression during the gradient design.

Next part, we will talk about the regularization. I'll start. Yes. That's the point. It shouldn't be eventually, it shouldn't be just about this part.

Right? You see it? That's the reason we've already checked that graph here. Reason in fact, reason with that team to the data, JV, that becomes here. Next, we want to discuss about legalization here.

What is MAP? It's not absolute. It is what? It is a maximum, vital the MAP. Right? And this is the case in that is the how we'll be delayed as maximum condition of posterior estimate, MCIT, And this is how we're doing it in this case.

What are the difference? Because, the only difference is about in the file. What is the file? The p ceta is the file here. By normal distribution with 0 being the identity covariance, and the pushy's practice was zeros.

And here are different kind of regularizations, and, it'll help it'll can help avoid very large weights we're feeding. That's the my reason why we want to use MAP because We really try to avoid bandwidth and the whole video.

So it's, again, the case about all different kind of regulation that you can see I didn't see here. And then eventually, we we try to see a better difference about the ready to send.

Right? So first part, you guys already noticed that this is about, maximum condition on what the life of the estimated MCLE results. Let's try if you use the ready design.

And this is difference between MLE and MIT here. We can take a 5 piece break. We can let you guys in professional. We have lots of masks here. Right? This is, confusing because you guys very good to know about what is a mass mechanics.

Let's try to have one I mean, we can't get started. So, actually, up to now, I would like to Say, majority, I bought purchase a profession that shouldn't for the best comment about it's about everything up to now.

So the next one I want to discuss is about the multiclass classification. Just maybe I just talked about the binary classification problem.

So what about the multiclass classification? If we try to use So just a regression method. For example, in an email folder or tagging, you may have multiple tagging plugin.

For example, emails from work, emails from friends, from family, from hobbies, etcetera. Right? It's about medical diagrams. So some sometimes, we may like, you have not idea of of and catch a cold or maybe you catch a flu.

Right? The weather, it is, like, stomach, it's cloudy, it's raining, it's snowed, all different classes for different, micro class classifications.

Right? What about the difference between the binary classification and the margin cost classification? So from the prob box, it's just very intuitive.

Right? So because in a multi class classification, we have more than 2, let's say, applied. So how about the logistic regression in the multiclass Classification. So we do call the web versus the all auto web versus all the rest.

So this is about the So we cast that as the right so in this case, for the logistic regression, we're usually Performing as assuming as a one classification problem, what we do here is For for example, for the h c l one class, we will try to pass this class, this is the class y, y, and this is blue, the real class.

All the rest x 2 here, this is class 2, and this is all the rest. Right? What about the additive 3 here? You should be just care about these are real numbers, and others would talk all or the rest.

Right? And this is so called web versus all here, and then this is the Equating that we try to pay attention to this multiclass problem that instead of xi xi1.

Yes. And this is best about the probability of the y equal to I that given x and the parent, you can see that here. Right? So this is the x. We'll try to pair attention and see that I x here.

To check, logistic regression class y, x c I x for each class I to predict the probability that the y is equal to I. So let's say given a new input x, we will try to pick up the class I, then try to maximum I in the I can see the I x.

This is investment maximum. It's probability by this part. It's really about the probability of value. That's right. Suzanne, you guys need to remember what about the cost of the money class that's very tragic maximum.

What about some methods of comparison? The directive method and the discriminative approach that might be best. So that I tried to estimate the probability of y and the probability of x given y, and this is about prediction.

Pretty late function. So that we will really study about SPM. That but you guys need to remember for the logistic regression, we really try to predict this probability.

Y over x. Right? And, yeah, some more than release. Guys, I'm just trying to do the details of it. And then next time, we want to Try to let you guys remember what we learned from this call so far.

So we have learned About the top, hypothesis representation, it's really about, signal model function, and we learn about a cost function regression.

So this is really about this. Right? But y if y is equal to y, they should be negative log.

As you see the x and y is equal to 0, they should be negative. But one minus, you see the x here. And over about, pretty decent, we Guys, we need to remember that we use m as in after alpha should be y over m.

This is the sum. As as you see the x, a minus y then times x I, xgi here, that's about representation. Remember, we put the, additional addition was.

For the addition was, we can't get it. Really about the massive penalty here. So lastly, we learned about the multi class classification, for using the just a risk That is tracking maximum is probability in the corresponding task.

And this is the key point from this lecture about Logistic regression. Next, let's try to extract about the details, how you guys did implement In this case, you guys may need to implement by yourself. And the yeah.

And the details that you guys can I'm looking for so first of all, it's about, civilization. This means about what? This is the first step. About what? The sentence is about the in your in your normalization of your pandas.

So you will have the feature vectors that is m size m by m plus 1. M plus 1 x 1. Why is m plus 1? This is m means number of points, And I think he has number of automation or number of features in each.

But right? So why is and pass 1? Yes. So that is it's a box. So, like, you guys need to remember once you try to implement it, You should measure the dimension edge.

I'll send you have n plus 1 every month. And then next, you needed to, implement the signal functions. Very easy. Right? Other factors, this is d.

That's need to remember. There is about both part of this feature vector and this weight here. Next step, you guys need to define a function of the cost function. So this is the cost function that you guys need to define.

Outside, you can calculate lots and lots. I'll tell you a mouth file. Y hat I and the outputs here. Right? So next, you're ready. You need to use this how you should move the to us. A maximum loss. This is the how it updates.

Just let me wait. Another time you guys need to find previously. Right? And the event is the part of, great designing. That's what you guys need to use. Yeah. Mark, and, I was at 6 studies about the prediction.

How can you make a prediction? So for previously, the way you decide you guys already So what you get is the way that I'm biased from the grid design because when they commit prediction, you know, adding new points.

Right? So that I get the y hat. So that after after that, the some steps should be captured as accurate. This is just about what does this mean?

1st part means what? Yes. This is the total number of So then I will try to review it again about the logistic regression. So now after now, I think we finish about logistic regression part. To be honest with you, it's not always this.

It's not not always this. You can it's better to use this, Sometimes, the diverse trend does that. But they're set. I Not necessary for for this part. This is something that I want you guys to keep in mind.

Maybe in the future, you guys need to do this kind of So then we will talk about, I'll quickly talk about the cycle and then data about the multiclass. From 2 reports about just the regression.

Then we try to use dataset as dynamic prediction. So this is just a few lines, guys. Starting in the future, it could be that's so we do have the future would be a lot asking to implement your look for just the regression from scratch.

But this can be one to build match with you guys. So which means that Jerry, you guys' homework. You guys need to pay attention to this. Not try to Google it. But useless for you. You Right. It's not visible for you guys.

So, anyway, if you guys have any questions, let me know about your of it. Let's try to put our another section about this lecture. Let’s quickly talk about Least Squares Regression. So giving, dataset that x1 y1, xn yn and to x and y and where x It should be x y, x and the y, not random.

The yi is equal to alpha pluse beat xi where I variable as random where random variables U1, U2, Un have zero expectation there were in the variance of sigma scale square here. So the method of these squares, let's try to choose value if alpha and beta tries to satisfy the equation.

This is what? There's a very bad error. Right? But this part is called what? Yi minus alpha This is called what? I don't know how. So later, we may have last slides to discuss with the details of it.

So next, the algorithm, observe the value yi correspond to xi and the value alpha plus beta Beta multiply x side on the regression line that is y is equal to alpha plus beta x here.

So, typically, this is just a fitting line. Right? And this is just one point and this is what another point.

And then this is the point from the regression line here. Right? And then here is the object function that wants to achieve, but, this is the least square estimation. Is that clear about this function?

So next, we really want to discuss how we can estimate these parameters. That will be our set as a direction. It's equal to 0. Is that right? Okay. So then, if let's say, let's try to track a derivative with alpha.

Please respect your alpha. Yes. Just the please respect your alpha.

No. It doesn't matter. It has the 2 here. Right? They are the square here. Right? Once you have this kind of problem, you would you probably will move this to to be here, you know, front, then you will copy this.

Right? Then do a directive with this to alpha again. Right? Yeah. Move it to the front, then copy this one. Yeah.

Yes. Yes. Continue. Yes. Yes. This is, 2. Yeah. Yeah. Copy it again. Copy the support equation, and then you will Times with another directory, that is with respect to alpha. Yeah. So now alpha should be negative one. Right? So Yeah.

So it should be no. That's x I, but you still actually be to the the rest that should be here. Right? Yeah. The combination of Xi with those. Right? With regards to this. So first thing, you this is just very easy about this this one.

Once you try to tell a graph here with respect to alpha, you are of course, this 2 is a wrong to write. As you can see, you are putting a 2 in the It should be negative. Right? So you will have negative 2 here.

So now because we need to set it to 0. So Yeah. The neck to his gut. Right? So this is really a body for the first of my life. Does that make sense? No. Then try to expand this first. What is the answer for this one?

Once you try to make handles, what can you get? Yeah. I mean, the better. Yeah. Yes. Yeah. Yes. Is that is this correct? Good. Right? You get it? What about the second one? The second one, what is difference for the second one?

Second one is what? What is the answer for a second? What is it with respect to the beta? What is the answer for the beta? Yeah. It's it's just gonna be the same n times with f side.

Did you get it? Where is our come from? This is 2. I'm sorry. I'm sorry. I'm sorry. It's 2. Okay. So now next, you need to take it with respect to the Beta. Yes. What is the answer for this part?

What is the answer for the whole? So the beta. And in terms of that side, is the in 0? This is 0. Yes. What about this part? This should be just the x I. Right? Just the x I. Just the x I. So we we will keep the same measure here.

So, eventually, it should be just x I in the angle. Right? Right. You guys tell it? Okay. Good. Go back. So now this is the final issue provided as this one. The only difference is that there's Additional x I here that is equal to 0.

So once you try to expand those 1, you will get a disaggregate. Right? Next is the most important part for you guys. How do you solve this alpha and the beta here? This is the casing again.

You guys have around, 2 minutes to finish to solve this kind of Problem. Our problem was through demonstrate. Can you share a can you show us for this So then after because the past you get a beta, but how much should be very easy?

Yeah. Go ahead. Come. How would you solve this for the app and beta? Right now, I am I will check each of you as they can get some service requests.

For example, if you want to calculate the beta, you need to cancel this alpha. Right? To cancel the alpha house Well, for the first one, what do you need? Do you first need to divide and divide by n, so that cancel is this part.

Right? So then you can cancel this Alright. The first equation divided by n, then times with this part. Just a waffle n here. Waffle n. Waffle n. Right. Okay. So now go ahead. Hamza times with this part.

So this equation, we need to still need a time like heading button. When it comes to extract the beta Okay. I tried to that about the yeah. It's fine. Are we gonna need to also. Is it easy? It's not that easy as you saw to write it to?

Is this crap? I'll have this one. Is this what? I explained to you, Asaba. What's this this part? You want to remove? This is what? This is this part? What's this? Yeah. Just, that's first step.

We get the RFI, and we take the RFI into the stack in the. And then, And then divide the 10 account. Go to the website. Okay. So let's try and verify whether this is correct or not. Yeah. So this is about really about alpha and compare.

Is this correct? And I try to pull this out. Should be times with n d and times with a here. There should be echo. Right? In here. Is this right? Yes. What about if you will be have this big idea? Well, how can you get an offer?

Is it easy to get a alpha? For example, in this case, there should be alpha just equal to what was this through this part of z divided by 3 y. That's correct. So you should get. Mhmm. This part, the miss the spare.

Right? There's no spare. You missed a spare. It should be enough spare to be here. I'm. Yeah. So, yeah, more the details of of it. So this is, again, about alpha and beta here. Alternatively, in the middle, you probably have this part.

And this part just about the covariance x y and this part of our variance of x I. So, eventually, we're gonna try to get above this is covariance xi divided by variance xi, and, also, this infinity is equal to this one.

X y bar, and x bar comes with y bar, And divided by the x, square x, which mean and is that me have x with the square.

Right? And you mentioned this is the case. If you guys do not believe, if you have time, you can try to verify whether this equation is equal to this one. Just try to test your knowledge whether this one is equal to this one. Okay?

It's kinda of a telephone market. Let's continue with it. Once you get alpha and beta, it should be very easy for you guys to, calculate another one. As I said before, what's this? This is ready. We need to talk about the residual.

So this is a way to try to explore whether the linear Regression model is appropriate to a model given a wide range of this set is to is to inspect a scalar point of so called residual Ri against the the xi here.

So the i's residual, ii, is defined as a vertical distance It turns the ice point and the estimated equation line.

So for example, this is could be a y I that minus Happy here. And this is the sort of residuals that we can get from the model.

But I always you needed to remember the sum of all residuals should be equal to 0. Right. Is that true? Okay. It's good. This can be a quick Test of your knowledge about this, this estimation can get the up and beta beta big.

Let's try let me try to write down this. What? Should be 1, 2, 3, 1. 8. That we had. 51. I have screen 1 scan. Right. What is Avan data? Can you read Avan data?

So according to this bill, Those are three key points. Let's give you several seconds of our online screens. So what this is is, 3 points. You can write down the 3 points here, and then try to estimate the alpha and beta.

This was a respect to this 3 points. Right? 1231. 851. If you can't read it, just, let me know what's your answer for your big idea. Oh, what's those 3 points again?

An issue again. So the 3 points are these 3 points. 123 1. 8 and the 5 one. Do you see these 3 points? Did you say anything else? Yes. Got it. Okay. Well, on my students, do you get some results about beta and alpha?

Not yet. So I should be just the ecotrist rewrite. I guess, can I stand with us? Oh, no. We're happy to answer. Did you get some result now? Yeah. I I got Data I have. I got data I have for, 9 to 0. 2 249.

245? 2499999. Yeah. So 245. What about your alpha? Alpha, not yet. One second. I see you guys are probably good. Kind of different answer. You're right. Offset. But you're right. Okay. It doesn't matter. What about the rest of you guys?

2. 349999. Yeah. Right? When they guys have a good number of ways, let's address the answer here. So the answer here is about the alpha should be neck That could be a negative 0. 5. Yeah. It should be good. And the position 2.

35. Right? It's not just the number. You guys can remember, but always try to be patient. Even though this is so simple calculation, a high school student can do it. Well, you see, you still can't get something wrong.

Right? So always keep in mind, be careful with everything. I think for the rest, it should be very easy for you guys to calculate about our residual. How I adjust rewrite? How I just skip this part.

You guys, if you are interested, you can try You calculated, once you have time, let's try to continue. So that should be everything from today. Okay. Is really about, to implement, well, just a regression model.

And then once you implement your model, you need to compare results from the function from the socket event, and then try let me try to quickly go through, the homework with you guys.

Hi, guys. See my screen now about the homework week 4. So, basically, this time, I Try to load some kind of package. So this is the load of package for you guys, and I already created a test set for you guys.

So this is x train, y train, x test, the y test. As a what key things you guys need to do regarding the implementation? Let's try to go the 1 by 1. Remember that your first step in trying to get a normalization.

Right? So here, I already write down the period. There should be x minus with mute times with, sigma here. So in your first, assignment not assignment. First function, You need to implement this part.

In this part, it's really about calculating the normalization of each Each row. Okay? So that's the reason it has a follow-up here. This is where it mean for each row, you need to Nominalize it. Okay.

So that the second part, Bob, number in your utilizing your pandemicers, As I said, we have a x and, weight here so that you need to remember, for example, for your weights, The weight should be has a size of n plus 1 times with 1.

So, specifically, because your dimensions would be 20 you said they use 21.

Imagine 21 by 1. It's a dimension of your weight. As for your x, so, eventually, it should be dimension of 800 times with 21. Y you have 20 y, so you'll have a bias. So remember, you need to add add additional column for your x.

Right. Remember that. For the first one, for a weight, you you you you should use num numpy dotzeros should be general generate all zeros by our weights. Next, you need to generate a new x.

A new x shouldn't be The first column always, then the rest should be equal to the previous x. Okay. Get it? I really hope you guys can write down because if you forgot it, you were not able to complete it.

Cost the function. So the cost function is a little bit of complex. As you can see from here, so this is the overall value of your loss function.

Plus the function, you need to re implement. That's because this part and this part this part is so called the cost of 0. Cost of 0 is referred to y I dot. I just see the s. So you basically, you needed to implement this part.

Yeah. The cost of 0. Okay. This is the cost of 0 here. And then this is the cost of y. It should be referred to this part. The second part is called the cost of 0. It's cost of 1. So, eventually, you can combine these 2.

You guys need to remember that you do not need to use a for loop in this case, Which means that you will play a little bit trick between the for example, you may need to use dot product, and, also, you may need to use Test post visit.

Okay. I already put that in the here. My skies increment. This is why time with times with blah blah blah. You should use dot yt blah blah there, and they're similar here.

Okay. So, the 5th step is about the gradient descent. The gradient descent is ready. So once you calculate this cost, As a greedy designer, we are trying to indicate the cost together.

Right? I already write the first two for you guys. You already do that. So next, you need to implement the weights the weights, which is so it should be the setup. And there's a cost list.

There should be a pen for the all the cost for each. Probably, we are using this cost function here. But the case is about something after alpha. So after alpha, you really need to calculate this blah blah blah, this all those part.

Right? This one over m, this is sum, which is just above the mean of it. Right? So when, once you calculate as you see the x as you see the x, it should be We already got it. Should it be the signal mode of the z.

Right? The signal mode of weight and x There, and then minus y, then times with xj here. So you you so in that case, you can finish this part. Eventually, I use a dot product that is x t here to refer to x this x here.

So you put a x t here. So then the next part should be Really, h zeta x minus y in this part. I guess Guys, no problem. So eventually, once you finish, how the grid design, you will try to go to the prediction step.

The prediction is ready, so it should be This part should be easy. If we need to be because I'm 0. 5, it should be equal to what? On this, it should be equal to To what?

If your prediction is bigger than 0. 5. Yes. So your list should be a one. Otherwise, this should be just a 0 here. So you have your so your prediction have enough of the prediction l I s. This should be a prediction.

Eventually, you have 10 calculated the accuracy, y and y hat by using some summation and divide by lengths so that you gather it. And, after that, after you define all the functions you can try to run and complete in the model.

For example, first step, you normalize your data, then you will try to define your learning rate And the iterations for your gradient descent methods, that says that we try to get the weight and the cost from the logistic regression gradient descent.

Because from this function, the way it give you a ways, it's about your setup, your 1000000 cost. So you can't get this.

So, eventually, you can't get good accuracy of your trying to test that and test it. Test that. How does that you can plot out across the list so that you you can see whether your grade descent measure is convergent or not.

If it is not convergent, which means something runs here, And, you eventually try to compare your results with the logistic regression from.

Okay. Does that make sense? Okay. Good. Well, just, right time, we're finished. I hope you guys kinda finished this earlier.

Right? Because Yeah. It's a little bit difficult in there. The face finish there. Side as well and parcel. K. Well, all my students, do you have any questions? No. Okay. That's good. I will see you guys. Bye. Same for you too.