Tokyo Workshop on Genomic Medi...eutics, and Health 20240413-01

Sat, Apr 13, 2024 9:02AM • 1:35:29

**SUMMARY KEYWORDS**

data, gaussian, correlation, beta, prior, estimate, lambda, called, term, solve, causal, bayesian, type, individual, talk, variance, snips, case, fact, assume

00:00

This is a UAV is really reveal somebody doing some interesting method development associated with data set so he's really happy when he accepted to come and do these. So way over to you

00:26

filming for the lighting here

00:29

the blast

00:32

apparently, really for me and I have learned a lot from this presentation contagious, like the workshops style because they're interactive, small scope. So just by raise of hands like just for me to know your background how many of you are raising

01:00

How many of you are in research on developing machine learning methods for computing? All right, thank you. So today, this morning, I will be covering three parts. The first one is Bayesian polygenic risk score requests and monitoring some of these large data data using car to life snips covering the Bayesian approach. The second part, we'll be talking about phenotyping using a well known cognitive particle model a family of properties to track their progress and talk about so I'll start with the first overview. This is a recent review by theater recharge. From this you have three different vertical axes. So from 2007 to 20.2. You can see that the number of our rivals in terms of the last study the number of low side also Griffin being freezing, up until around 2017 And then the sample size by classic is still continuing. So let's talk about something about this. So saturated on physiologic poverty. The knowledge is fine to me is a method that is scalable our scale is able to use the field

02:44

from yesterday we have learned a lot about the difference large smell banks. So here is a really good, a globally global overview of all these bags fabulous. So we're focusing on the UK Volvo also learn a lot about the MLB. So together it is around 2.2 million genotype samples. So it's a lot of training examples to work with. So as I mentioned, we will be working on it with datasets. The reason is that it's widely accessible so they make it quite approachable by researchers. So perhaps you only need to introduce a data so maybe you have normally the genotype information you also have an administrator so data like electronic health record data, different laboratory measurements, etc. So the challenge is how do we integrate the information and trying to really learn from the genome

03:58

so yeah, so zero. Wrong. Really. Overview overall research the writing lab is to develop methods to predict disease risk, more personalized medicine, elaborating is a heterogeneous genomic data. So it never would have been public as Madison This is the focus of this talk. There are records genomic functional genomic medicine is promising accessibility RNA sequencing. We also worked with a single cell genomics, as well as multi omics integration. Rather, how do we didn't have all of this information? These mechanism and therefore for disease risk more accurately? So I believe this right there. So yeah, here's another way of looking at it. So going from this information and all the way through the thermo phenotype. So there are several roadblocks so essentially, is how we understand problems. So in my research, we focus on three different dimensions. So the first one is to type type so we're using data to predict the field using our classical statistical methods. And then the other axis is on the single cell genomics which I will not talk today. So obviously, this is another really active research area, so that we also continue quite recently. The third one, the third one is probably contractor or where we use the DSR data from other data set and try to understand the field type comments. Among these observable diseases and Mexico.

06:01

Given that this is a lecture based workshop, probably summarize the learning objectives for part one, you know, we want to figure it out right off of the polygenic risk score. So we want to grasp the high level picture of a bench Score Fast priors. And Sir we want to understand the established Drs. Especially when the snips are correlated to how high dimensional features are correlated there will be some on this the other one and also other topics on how to improve release accuracy. For us, we all appreciate New Zealand various different framework, which is an efficient framework for dealing with high dimensional offers a problem that we'll get to know the chair as application data and using Wi Fi is shoddy to give you a far you know they're assigned some type of exercise using recently developed software by us, Verizon an interest apologetic risk warning. So lastly, we want to actually appreciate the difference between what makes the aperture is is different from what makes accurate assignments. So there is a concept of a multiple sample sets. That's the important concept but I leave it up and you know if you want to learn more about fine mapping, I've given you a couple references. So yeah, so by the way, this is kind of the lecture. So we don't we don't have to really spend some time because we're only speaker today. So if you do have passion, just feel free to raise your hand and it's very I also want to make this interactive. So the whole point is to communicate the concept clarity, and you guys don't have to ask in between

08:03

so, so let's jump right in. I would like to first talk about the very basics of prior for Bayesian regression. First I will talk about the problem. So the motivation to introduce is we want to accurately predict from you always has tremendous value in diagnosis. Northwest says we want to predict the outcome variable and so the challenge there is how we deal with also had a mention of going from 1 million to 10 million. And demonstrating example is I don't think I think he gave a very good introductory overview of the data science challenge in this province. I really appreciate your presentation and your service. And that, you know, the challenges are lawlessness have very small effect size doesn't mean that they're not causal. These are some small, so when you add them up, we have chameleons versus 100. Monkeys, you actually have a substantial power. Well, the question is how do we make use of these steps? So there is a very famous example of standing height, as you can see that there was the former all the chromosomes you have significant snips, so they are above you. So, there is a problem setup measured by a large number of individuals. There is your response letter. And then this is your design basic input in the results individuals. So you will feel tied to the measure 01 to zero is normally the case and want to do what else and although the what's my blue is trying to achieve it's trying to estimate this more or less a coefficient coefficients, dairy taking our products, or Esau these individual which is role or b multiply with same matter data. So the problem here we're trying to figure out what is requesting qualifications.

10:36

So we're going to use a Bayesian approach. So here's a simple diagram of the story. of the fluid moving from a frog to the crisis, and now evolved into almost a programmatic programmatic. So if you're going to have a knowledge in the approach, where you want to maximize the likelihood of getting practice, that's a maximizing for us. And all your xavion you model the joy, sometimes called is maximum a posteriori. And then in the Bayesian approach, you use finding your firm posterior using the process more challenging cases because of the mathematical.

11:41

So firstly, you want to talk about the simplest prior, which is a Gaussian prior. So this is there's a couple Gaussians on this diagram here. So the rules are assumed to fantasize follow the Gaussian prior. What we are assuming is very often that the Gaussian is zero. So that is the majority of the facts right size, or they're not and then the variance is defined by with a Gaussian curve. Now if you have a small variance, so in this case, this is a concentrated around two now I see a very, very large, Gaussian, or first that looks like this. So when you have a very large variance, you allow large classes, as long as you draw a vertical line here, that's the way you have a large American variance right. So as you you'll have a central tendency when you have a large data set, but now there's a very small, the largest online there's zero records because it's perfect. So that's sort of the Express. So this type of Gaussian prior over that size, and then through the barriers we can accept, repay, express our beliefs about facts. So to write down the Bayesian inference, so there are three major elements. One is the prior one is a data one is a lack of undecided. So now the prior in terms of the genetic problems, we assume, beta follow a zero mean and standard deviation and followed by the heritability divided by the number. So there the assumption is that each individual each snip explained playing equal amount of berries. So that's the heritability. It's the various blank. So that's the variance of x times beta. So now it's basically severity. So without knowing any specific snips, in their functional genome, so we're fools as

13:55

you can see that this very large, so in the visual display really small. So that's sort of our faces so the livelihoods here we're assuming also 1000. So that is doing the response variable, we assume it's a continuous response variable, and then the next time data so that's our basic and then the variance is something that we don't like cannot be explained by. So the query the challenge here is trying to figure out the posterior distribution of theta given y. So as we know Model T students or x axis always competition in distance. So whenever we write down the base, there's always an excellent divided by five. So the denominator of very often this problem, right, because the therapy of Y given X the involves integrating over the beta. So now this integral often doesn't have a form. That's women make the baby infant child. You cannot just assume any worries in fact to obtain the most for your use. However, you hear people like bouncing because it also has very nice properties. The way you have a Gaussian prior and it turns out the posterior is also so this is called the Council. Prior is conjugate priors, the posterior which deals in the form I'm going to use is the same form. So to solve this, as I mentioned, this is also an easier to do in Java. So it has to be independent. So the problem is how do we figure out

15:50

okay, for the

15:54

NGO, we're using that space how many of know how many how many of you know about this

16:07

domain so the way that we kind of saw this is kind of just leveraging the provided forum.

16:19

I know the slides maybe looks like a very dense view. So when we have that about the guy you have about him prior to

16:29

going there, we're assuming just by convenience, learn to express the merits of numerous positions as fast as possible. These are jobs we're talking about. So magazines like this. Now we multiply them together. We want to join the model. The idea here is we basically have the virus the firm because the longer the better. That's a constant because they don't evolve. And now we have because they have an exponential rate. So what do you multiply them rather? So the product product equals sum. And I just sum and now you add that inside a term that don't involve beta which is the first term without defining the concept, now, we deal with this form here. So we formed this into this quadratic form beta transpose. And then this form is the data products. So by completing the square we recognize the term Tom street with that said Florida, called the universal awareness office in Dallas. This, this term here we denote as a, which essentially lambda, where lambda is the inverse of the covariance known as here, although in the prior we're assuming that the facts are independent. Because there is a fire orchard. We do this across here. We find out that a fullback is no longer and I see that I see that for anxiety. It depends on the excellent cause X was experts was x, which is also n by n matrix that is proportional to the correlation. That is your your covariance is being used by the clock correlation. What's the correlation between says, Listen, sometimes we call it the Olympics is deliberate. So that is the cause of accidents. So the you see the mean there? Which essentially for various time, all this is persisted for air worthy Americans. Excellent. Excellent. Suppose Why is proportionate Martin so centrally, the fluoridation East individuals, and more importantly, that's a fire. But this is adjusted by their correlation between others. So that's how we interpret events and see that this is also intuitive in a way that either expressed in facts or even evidence. So here we have samples of correlation because there's one line of evidence second most important line on evidence and excellence was live for Asian

19:51

so once again, so we have our posterior also has a Gaussian covariance matrix and the popular immune so popular immune is that our s beta, so we're going to say after retrain this model, we're going to make predictions on the test. So what we do is we take the posterior mean, so this is our request, in essence, and now we apply to the new genotype information to obtain the predicted So in here, you might wonder, okay, explain this. You know, what if you don't have individual to have, so this formula actually allow you to learn the fuzzier man and covariance with access only the summaries of physics. So we're talking about the privacy yesterday about the challenge of sharing and facing data. But by sharing the Sunday citizen, you don't actually reveal anything about it easily. So what is some summary? Summary Statistics is certainly far from what we know is that we have experts all blacks, which we denote it as four legs and divided by the sample. So assuming that you standardize the genotype, you don't have to worry about the mean for each office near zero. It's already basic, it's one. So the chairs information, essentially reduced to one over x equals x, but that's one part of the song you can see that this is operated on m by n. It doesn't involve any specific individual to hide information. And the other sovereign citizen, his marginal tax is excellent. Here, beta tilde is one that's the marginal effect size. So in the large gy study, the authors are required to release the summaries ready submitted in the middle of the field at bat for people who are interested in access that apply for but the summary stats are publicly available. So here anytime see that callback is on the Express there you have access. So this is where you put on the other. On the left hand side you have the spark. So in this part is we have some fields directly using a reference. For example, you don't have any individual genotype from the accuracy of our score. You can estimate this, this correlation using solid zero, which we don't have access to. And then the beta tilde is this one. So this there's a class here and in here you have access to as why we express this as an but in other words, the passage drive the warm up area used to build for the backside stata using the solid staff are building and then the rise of the data confirms that all astronaut Alveda are very often estimating the maximum Eigen or are you research using the validation set? Or the question to fulfill following

23:06

this feel free to see what I mean

23:09

by this lecture slides are mostly from the forest that talks from euroscience 565. So the all the lecture materials are on the website. I can also release this slides afterwards. So if you're interested in anything that we talked about today, so the materials are also available

23:35

I explained it in the context of the last spring well but obviously these general principles apply in

23:49

just another example, I guess a district sort of joint distribution could use

23:58

Yeah, so in general problem when you deal with McGlasson and these are sort of the classic folks, that's not just limited humans. So when you deal with the emotion, or the variables, and you try to get the hell out of here excuse you know this is a very sort of General ranepa. And then, you know, in some situations you don't have access, maybe you you're protecting the customer satisfaction is quite different lab is the summaries standard

24:44

clinical trials, where a lot of clinical trials a lot of clinical trials utilize Bayesian approaches, just because sometimes there are issues with getting helps with a you have a smaller sample size, use a Bayesian approach where you can incorporate prior information. That's kind of another approach that uses a lot of these activation functions.

25:12

Yeah, thanks for bringing that up. I think that reminded me also the ways to express uncertainty. So normally, when you're just offering your estimates there is no uncertainty massive liquid the pot area difference, your uncertainty that called as the various of your assets. So you can say click on file. Let's say you have a set of panels of markers, and you want to pass obviously you have a limited budget you cannot have 1 million markets that you want to choose the one that has a large effect size at the same time also have a small small variances tired, sir that type of legal residence as you know is

26:10

three important buy avoid.

26:13

Bases are underway. Consider is important. Is says correlation between the coordinates and billings sensing is of interest to me

26:39

it measures the relationship of our billing other snips and

26:52

so are you take away both ours and the correlation between all you take for that gave you the forest and for last year with the Irish pudding this level itself? That's the bag so that's itself that it started out I saw four aces

27:14

on the bench here that will be the two lilies and marginal and pick.

27:21

So I must say that the reason I'm reading the law is the effect size and the estimates the effect of addressing it by considering the effects of others and vision.

27:42

So there you go over time for other facets on top of correlation excellent close, right. So right so you can take a snip, so I suppose it's wrong. And now you just calculate a dot product which is one. So then when you copy that, you know in real time format SPIRATION this

28:12

dot products measures the

28:19

distance, which is the product measures the distance that are

28:24

measured based on

28:31

that right so let's move on now speaking of validation I mentioned so you already have this simple these are the two hyper parameters are necessarily universal. But there is something almost cannot explain how beta is the virus person or the universe. So, so those are some things you can measure. You can estimate them. Imagine that if you plug in your estimate back to your library, now you can maximize your data, like either data or you can select using the validation set. So now you'll save time so you're familiar with the machine learning set up behind the training set validation sets live somewhere else and you have any type of learning method when unless you have a hyper friends. In its neighbor, decision tree you'll have in your neighborhood a number of neighbors cybercrimes between a regularized regression and then your policy parameters. So now you have a validation set. So the idea is that you can choose one type of parameter and then you train your model based on that type of parameter. And now you use it or Fe model to apply it to the validation set. But it does add some something that you don't use and now you want to use it. I think it's I think you got it a cybercrime and all you can do that for at least the values for the same kinds of crimes here and now you have a site of validation performances and now you choose the one that timeframes the value that gives you the highest validation so that's a sign of writing in standard was learning. But also, in that case, you need to have access to individual data for the validation because you have to test these individual pieces to evaluate and you can divided by area under ROC curve, are you predicting IOC f1 score? I refuse to use the R square to kind of nicely also do the summary stuff from the validation set. The reason is, you don't need to access the individual data even you're using validations as an idea. So what is our for our slides, they're seeing how to do so ours was just wanting to subscribe to some words from the procedure. So that's your prediction some of your patient has divided by total variance, some of Squares Total. The sum of squares on the receivable, very high response error, you have the beta hat, but as we said the beta hat is being set to zero and then you have your use of visa does not matter access data has to have your credit. You know how do you compute the difference when you break this up for radix and you have four terms? This mark that is about these valleys and then at this point now you can say we can see that here you have violated x. So here this part here you can also recognize this is the marginal effects is computed from the foundations. So that means that this part, you have a summary sense that we don't need individual and then this far, this is coming from extra data they have access to a better hotspot as well. So here you can see that this is also correspond to the LDA LD matrix which is why we use what's known as a Saturday. So we can substitute this with the LD from the reference time. So this is basically a massive covariance of your estimate beta weighted by the correlation between now the CRS as I mentioned, this is the marginal cost and you can see that this equation here doesn't require individual now this is you assume that you're still tied to center this Saturday, basic one here, follow some sample data, so that you can see that if I want to use a valid sample truth, my chiropractor, I can say this aside, let's say to do a training, training validation slips right you have only one single thing. In practice, let's say you can split it into training, validation and testing. So now if you use your validations and you compute your thermostat now you can see at the top of our message for macro prudential performance higher status. So you can use this place passively. So which is also very efficient, because it doesn't require you to predict municipalization patient individual.

34:06

Yeah, I hope so I have a question regarding previous slide. So I think some of you expand this so you should find the Thung data using much more likely research. And I think that you know, each square is heritability, right? So I was wondering whether if you know, now here we are be prior knowledge. Do you think you should use information as acquired information?

34:36

Yeah, that's also a good point. So like, the last way we did we also, right, so if you have a heritability estimate, so then you have to be careful you're having to carry that estimate is only estimate using your tweets and so then there's no information your estimated favorability Google's entire other use is 90 Use your test cases. So But then, in the end, it's really kind of mad because you want to you want your message to perform well, right. So sometimes the theoretical part of it doesn't give you the best prediction performance. So that's another What if you're going on up and you want to maybe validate on dinickel Different, different cohorts, right. So in that case, you'll have the ability to maybe then differ a little bit every four or five years is similar, but it's still not identical.

35:36

I just have a quick comment about that. That's a very good point because actually, that's how the LD print model works. The estimated regression and then they plugged it in into this estimator. And that's how the LDA model works it so that's

36:01

it so now, for some of you know about kind of questions you might have heard about progressive techniques, Khan raises the question. So raise your question, actually has a very direct connection with Bayesian regression where prior to reworking the formula, this is the last. So their first term is predicting loss, and then plus the time of the data. It's a square of more than the sum of the square of the individual effects. If you have a large effect size, you'd have a large loss here. But this will contribute to the audit logs. So that's very relaxed and you have a Title parameter lambda alerts, you set a lambda large, now you go allow your batch size to grow that much more than your beta. So in a great request, and very often people want to account for overfitting. So when you have a high dimensional features, you want to have a professional lambda to control the magnitude of the effect size so that the model doesn't over. What does overfitting explain very well on the training side, you have very good performance on the same set, but I generalize really for the test set. So in machine learning, there's an order maybe in January computational science, there's various bias trade offs to overfitting your model is too flexible. But then it has a large variance explained it pretty well and isn't a small bias. Now when you pass on the test set which has individual data sets to model performance.

37:50

So now, we'll take a partial derivative of this losses to activate the arrow and if you have a false one, kind of similar to the very thing that's a normal equation lambda plus lambda times. So now we can see that this forum, this is Jordan Warner these were requested with so this is a kind of very similar to the popularity or not but just by looking at this it doesn't seem to be obviously separate. He had this excellent worldwide theory. But then if you started doing this land members before and only had it for an hour to bring out or rather you bring in the towel after them, so you'd have good work here so they can kind of go out and on this farm become the ratio. So that's how I feel over the last year said this to be the same as the lambda. Right, you can see that the solutions are the same. But in Beijing forum, you have a specific type of injuries. Right for the data, types of data and constant mass whereas here it is interpreted as the Bayesian is more natural property. So that's the climax so in terms of was not allowed to close out the workshop? No question. But the reality is that we have this tricky problem of LD labor now without LD. So the rat, the rat dog is Hogwarts. There's no correlation, very high lower triangular matrix, okay. So the way to determine the correlation between this guy and disguise and then the correlation. So, now inserted all partitions are zero. So now when you compute and backsies the extra X or Y or N, you will see the facts as large only from causes. So the rest are you know, they're not working. So they will have well now with that said suppose this is highly correlated with the quality and also the other. In that case, this will also have that's why when you reach them, it might have been blocked. You don't see just one single sentence go up. Down here. Do you see a sort of a building? There's a trading setups and there's that leading to that at least? That's because the correlation the LD correlation. So this make the request requests and problem mortality. So here we are. So this is our sort of beating assumption is that in most of the fields, there are only a small fraction of the reason I see a large amount of snip that being very significant is because of it. So now that sort of calling for the sparse, you want to give further toss airy facts that were only a fraction of the Senate as a non zero in fact, since the rise of zero so that's falling into our thumbs up if the snaps are long follow zero. So here is the real data.

41:36

So because the previous slide you explaining the correlation with these this was taking into account to what you're seeing here is that it's not sufficient.

41:50

Yeah, so previously, you were performing this type of pavement with a Bayesian priority and Gaussian prior then if x is zero for all unknowns. So that is sorry, it doesn't follow the information, majority snips are non causal, right. So this will have zero and also a kind of make the interpretation difficult. Right. So now you still end up with a meal as it was not yours. So that's a misnomer. It also makes it hard. So that's the introduction here. So in theory, people can have real data for body mass index. So here it is. My husband Bob. Now this is one of the most well known I feel notice is I can see that a lot of very significant super sensitive parts and minor seven. But it was all very fancy that a lot of them here are very significant and are small because they are highly correlated to the Olympics. So in theory, there's a lower panel you can see the more red is the higher there's a correlation with the correlation here, and that's why the last floor for data and therefore they allow there's sort of two bodies one is maybe all of them are causal, but now I'm not going to do it or the other way is only one of them or very few of them. But the rest you know they're significant because of this.

43:32

So now, it really like we believe in that I think the most of the cases is our towels or the restaurant, and I'll call them their passengers.

43:50

So yeah, that's what we just discussed. So now we will review the Gaussian prior as Gaussian as you're always united. So here, you have this, we call it called no configuration. It's a binary vector. It's one damage is causal, zero. And then there you have the tool called effects. So there's two factors are not known that we already know that some people have but if you actually use the Gaussian prior, now your posterior estimate is going to be nonzero phosphor sort of motivated as far as the where are we trying to get for First of all, even this back to the right, which is causal, so they didn't know. The second we tried to also start the out there facts, if there are causal what they're positive. So that is the problem of requesting for polygenic score. So later, I'm going to show you, you kind of have one single model that your third vote so that is the sparse request. Bye for now. So a very new reciprocals were drafted. Rightfully so we're the American version is complete. So all you do is that you take your your your have yours yes. Now you had your least diamond shape, and then the other sets are the circles. So here what you do is you first cube the correlation between the shops with Lisa. If there a correlation for our clients are indicated by this red, red, they have our five there's a correlation with the least. And now you've got to remove the basket or the land you're going to do this recursion. Right. So now once the reward is, maybe we'll go somewhere else in the genome, another snip that is a second morphine like that, but at the same time it doesn't correlate with the currency. So there will be the pay around 500 base pair and so now you can see Secondly, you your your distance there within this window around, say 550 or one negatives, and then outside of that you can do the same exercise. And now after that you remove all the things that are correlated with the most significant step and those who are thresholding. that's holding you to the arbitrary key by this rational rational, which is five times negative eight, and now you are paying your neighbors that are multi dependents. So that's the sort of type of crooks who are paying a very small fraction. So now you you basically set them to be sort of causal. So when you perform request and it was only about making use of the facts as of those that survived the threshold so here it is. It's just a mathematical expression. It was a set theory, you talk to the terrorists. First, you take the food side of address after deploying on plumping and now you sum up their effects that didn't that typically is below the threshold. So the theory has a good indicator to one but also as always one otherwise zero, and now you waited by the z score, Z scores basically, the reason we knew the marginal effects is because we're left with snip that don't have correlation, right? And test occluding exercise so that we just take them multiply with the genotype. So that maybe GRS but all that starts with two arbitrary choices. One is Let's wrestle for LD. So if you have a very high LV, we removed more of a field

48:11

if your keys are LDS or so we're more

48:17

looking to remove

48:20

right for the fact that you have a five lb for your move there. But you will need let's say this guy or not, for example for French fries, but now you only move it let's say find a party on the other side if your LTV is low, then you will remove it make more sense. So that's obviously you can see that this is obviously what was the price of what's the optimal price. And then the other one is the feedback. It's actually one more so that's kind of your instinct, of course these days, probably one of the most popular techniques because it's simple, it's efficient, that will give you the optimal performance. Okay, so, before I talk about the slacker slacker I'd like to first talk about another kind of popular pirates class. So let's look first prior in contrast to Gaussian, you have a couple of students so first, I was disappointed. Disappointed, zero when the minimum. So now we have a classifier here it is instead of taking the square of the difference between the effect size they're excited equal to zero is absolutely basically the absolute difference. So now you do take the love of the fire, because very often it was a lot of love fire. And now you see this is content to us. And now this product. Love some of the law classes also financial family of distributions along with us nice and casual always have this. So now you here is D there is a color scale residue similar to the various of the Gaussian. So now you put the ring on this scale one over B. So the sum of the optimal value which was 1014014. So now in theory, you should assume that that one over d is in fact size, lambda. So solving this There's a very famous request in our philosophy. Not always where you have your love is still were of no tomorrow, but I do our paddle a few days. Now one more, which is the sum of the absolute so here we've had maximize a puppy around you using this lasso. So that is when when you're facing prior isn't the last part. You can use the lasso to optimize your radar use lots of repetitions. So not a charity Pepsi that there's a comparison between reg rewatching. So that's the equivalent of golf and prior and lawful, which is a class you read requested and here you have the one over Lambda. So that is lambda is a penalty if there's one over Lambda, going from left to right, relaxing your handle the Lord way accounted for our large

52:15

Okay, so we have had a very large majority of you know, using the sort of gos data base or another separate data set for predicting tensors on your backside art. So then started when the penalty is the one over Lambda. Now when you become a small one over Lambda Mark 30 when you're relaxed the penalty is up and down. Very positive in fact without so you can see that it wasn't when the getting sick. When they go off the off the wall they all become downfield, rather than here they are not exactly they're just American close to you based on now your last okay? Because when the when the penalties are large, they're all here. So now when you read when you learn from one for America so the farmer so as you relaxing because we're lowering the Thank You that are the variables that are showing up synchronously. So that's sort of the last sort of penalty tax solution. So that you have this kind of nice orbit visa. So to fall off, we have also have very efficient approach which is called 40 cents. So what is important for the family is that when you are trying to figure out the backslash for you working conditions, all the rest of them that sort of sort of bagging question I kind of assumed in fact I've sort of rest of the minus ones this has already been solved. But in other words I take this are here to Slack dating that I picked these types for all the snips inside first of all my prediction using all of them all I had to do my prediction for Cara's you can also get them separately. So that's a lot of explaining the logical way so if you've taken a few steps here referred to as your Medicare for all. So now you you will that level. So it is not covered or below the maximum duration it will be more and more each individual. Now you solve the lasso regression claim base based on these three scenarios. So for the timer moving inside of time, I have not shown you that there is sort of like you can derive some reasoning from this. So I have seen what I've said that facts are sort of day to day. So here I'm comparing the excellent post j with the refusal letter. That's why it's a practice for nation without us if the CFA is poorly 554 days. I mean it's that's that's very helpful foundation on the presidency because I knew it either. Well, it's helpful to be either negative party with a renewal or positive. So that is your comparative with your kind of, if it's a more negative than to make the lambda have to be positive, or negative or negative. You're going to show the full form solution for large number divided by the variance is a more positive philosophy in the sense that there's a strong case it's a more negative, you make it less negative, because here's a looming value added on that you think if it's a more positive, you make the best policy for your strength. Now, the negative is referred to as the slim fix or being divisive. So that's my last thought also known as bias. So because now if you do the right the fact is is small, right candidates and the rest of you guys are excited to be back with you. So that's where the sparsity comes in. So again, this iteration if it survived the first round, meaning that because of a lot of negative I'm not worried. We didn't make it on. I'm glad you said that. But that explains it a little bit. But maybe in the next iteration not going to survive anymore. So that means you're going to set it to zero. So based on the lambda, you do have a few different kinds of sparsity densities when I showed you how to compute is the last of which port bounce flops so, oh yeah, so here actually, what I will say is that you can actually come to your laptop or you can also somebody searches for formula so this is really the data. So now without just looking at this is the requirement available. Now, if you do a little bit of mathematical, you multiply x with a wide here, so basically found the marginal success. Now you might think I've tried both with the X scholastic here, below, you kind of see that this fact even though you could come and find the marginal success so that's a summary. No extra dollars a day with an X plastic, which essentially is the sum. In theory, you would transfer X and fortes one roll. You multiply the x plastic which is n minus one. Sorry N by N by my mind. Now you have to do without product. So you basically computing the correlation between these two individuals nearby with a very fast correlation so that's the LD. So it's asking me that from the Reference panel, just what I mentioned. And then the beta hat is something that you have estimated to be the theory cycling through each one. By the time we reach beta j, you're assuming the rest of it has already been activated. So you can see that in here it is only required summary stats for success. And then the wrexham's ld different, meaning that you can't claim loss with only the summary statistics. So that is the method actually. It's called the local farmer. So this method actually, kind of modification is one of the most popular methods because it very efficient because this when we take a univariate disciples for one snippet, and now because you're only using some instances, you can run it on your laptop

59:49

so here it is. These are published in 2017. So the author has a try either time, I think the data available wasn't available yet. It has to be valid there. Welcome trucks, right here. This is production of binary field theory you have bipolar disorder, or wonder Are these communities so here to compare a lot of songs with certain key values wrestling, and now also with a free party, or with uploading without and now with the LD friend. Another method that was published earlier that uses more sophisticated prior public slack and slack and it's something that we'll talk about next. For a while for making binary response, you might think okay, so far we learn is only offered by you here we kind of think of this as a liability. So that is you have a although you have this a binary case and control, a single Imagine having to liability type of score, right what is a process wrestler is bronze wrestling with competence in his clothes is always a console. That is we can still offer continuous requesting. But if you take your sandwich that's maybe computers too long, but there's a request for binary. And now you can train your model to obtain the beta beta hat. So sort of sort of pretending that this summer science and I said compute it from the continuous trade. And then once you have that you can now do your prediction. For your production has got to be real bad, regardless of whether you're dealing with violence or binary classification. And now you can use your production. Now, you could take a while tools binary, but you don't really think about it the area under ROC curve. So based on different threshold, you're going to have a true positive true negative TPR fer and now you can talk to the area under Roc so that's pretty bad. So you can see that in this case, the local song lessons performed pretty well. Definitely all performed. So for LD practice, often regarded cases, maybe it's APA or in this case a few years before or so here you have a sorry, I forgot the method behind this lambda stuff. So last year, we had a lambda. So the lambda fantasy theory fact This always has to do with your one by the performance of the law when our students schools, lambda using the validation set, which counts will be the opposite. That didn't happen more either. Using the macro macro to choose the optimal if you've got high performance on the test set sort of upper bound they can see that it's a

1:03:00

little bit lower than that is by excluding new stories actually required monthly results to occupy the maximum performance

1:03:15

so the logical thought process is more efficient, that I'll be proud of because prior to prior to Slack and slack and using Markov Chain Monte Carlo

1:03:31

also we actually see them 47

1:03:40

I think is the thing is that why they have an upper bound, right, how much practice you know, I don't know the higher the ability to trade you can only raise the higher part ability. So so in this case, 47 is more than point nine. So it's not that this is more like a very complicated product. I replace a console with 90% accuracy is better than point five or point five, you have a hard time controlling flipping coin you got 4545 right because your ROC curves don't follow the bag. So the area under the curve so whatever is the point but it's always better than rent. Any other questions? So that's the we found just to briefly summarize and talk about using causing fire and then using foster dogs and cars and have this full form solution for a lot faster. We use a lot of techniques to solve. So now we're going to talk about the last fire which is the one that we use in our next lighters. That is this light bulb. Let me first motivate that. So remember, we have this multiple requested by me here not only we want to obtain the facts, we also want to obtain the sort of the binary unit, but then in the company in terms of probability. So this Max has told us what's the probability that is caused. Sometimes you call him off theory or inclusive and the other record, so you want to have the answer involves what to do with value multiply. First, whether it's in this column, obviously in the middle row, so we have a probability so we weigh the impact that way the fact that like

1:06:10

so, so now what type of firewall do the most so that's the answer. Is one answer. So now second slot is you're gonna see that fires are even more subdued. So we have this a mixture of Gaussians and this delta function. So now the first part first term here is the SJ binary indicates whether this is competent facts is coming from Dalton is this come from Gaussian? Because we know that to come from Gaussian you can never be zero, but it is positive. Otherwise one minus one minus that says I'm using that the family from this adult actually, it is a properly defined distribution. It has only the nonzero density at 05 1000

1:07:11

And so in truth is,

1:07:13

so it's Gaussian.

1:07:21

So that's the prior. So now just the write down and we will be there with lots of data we write down the formula for the elements that we need. Firstly, prior to poverty, we find that we have first we define as an indicator as a Bernoulli Bernoulli assumed to call the indicator for individuals are independent. So this is an independent from that if you have products like coin toss, the past is one that's indicated. And that given condition now this isn't called and now the distribution for beta is times delta is probably one of my so that's the so that's why distribution sensitive and you have the two types of environments once beta one is the library is there just for convenience. So we can read franchise like, like what I saw in the dialog where it is in the Gaussian now you're missing this. Why here why fold it and so if X time as the artist location is otherwise what I have left so that's the mean that's your verdict and that's what we promised.

1:09:06

So, because of the this fire, there is no closed form solution, I suppose. What the talent there to go forward as opposed to refine to solve the posterior probability distribution once again, for the fact size J and the binary indicator. And now all we have to integrate out the rest of this there because this is a binary indicators or configuration over a numerator. And then the type of data in real value, right. So this is, in fact, the last that makes the Bayesian inference more difficult. Although the priors seem to be more natural, it's something we want to achieve, but the universe is more founded. Not always find it. We kind of saw this by Markov Chain Monte Carlo we can sample approximating the school popular music, but the other very popular variation, a very popular Bayesian proclamation is called the variation of users. Specifically, we're going to make use of this a museum. Museum means that we assume under the variation of doing something that we do propose, because it's convenient to solve the taxpayer under the variational. So that is, the version of the agreement that we're pulling is beta as your beta unless they are independent. So the few you know the variational although we know that their posterity are not independent, because LD, like under the various assuming they are and so that you have to, so that's an answer. That is the only assumption we don't even need to assume parametric form. So if we can solve this analytically, so the producer will reveal itself. So let's see how that works. So how many of you know

1:11:25

so various different you have this and you first need to derive just as a multidisciplinary problem you need to derive a authentic function. So this alternative function sometimes you want to optimize with respect to your practice. In our case, parameters are made up nice. So the alternative function in the variational inference is called the evidence lower bound word elbow. So all the words in the bottom now you think you're a lot about the marginal like, that's it that's a p of y. That's the marginal likelihood. So you're looking around you have integrity, love your brand. They're desperate and stretching, not flattered. And location I leave out as so if there hasn't been. So now I'm very happy that you came along. here because this is the margin all you need to integrate over there this is integrity over fit theory how to fly. Well, now you have your you do this a trick to do the trick. So you're having to join, divided by two beta. And then finally, the weather there's all kinds of so you might wonder why would you do that? While we do that because we now we're going to use a family's Jensen's inequality. So this, these two are strictly not strictly equal. But this is the lower bound. But that's the name of this lower bound. So now the reason we do this is because it makes things more convenient so that we can put the log inside of it. But in practice, this is not this. But here we have we have worked on off the log in here. But now you have two terms. So that first term here have a few beta time models. Now you can say the log of the ratio, it's basically a lump sum together. And now you have this killer beta time log of the likelihood y and the beta. But this is a normal either the firm that you want to maximize and all the other firm is going to have a tough time law. So we call this and then I have this confirms. Well, that's the evidence lower bound. But you might wonder why do I maximize it? But what's what's the point? You also need to know about another quantity is called the KL divergence. Right? About a divergence is so you have the difference between your proposal and the tool healthier to have a difference between the overhauls. You want them to be as similar to how they are you don't know. But now you want to. You want to kind of estimate this in terms of so now check out every metric the difference between so in theory how big I think there are two beta time log of beta divided by. So now here, in this form here you have you take an expectation Q also the beta respect to the PC gamers feel better. So notice the popularity and relax in terms of it like attracts me to flat out here essentially by writing down the basis. There are minus signs that's why I was plus so now you can see here, I have this forum. So that's measuring the difference between microphones up here and not having the knowledge to add it all together. You see what happens that you clearly have a minus sign here. Add them up. So this term and this term, and then this term, this term payoff plus a row. And now what's left is this. Right? So that's the marginal likelihood. Why? So this we also assume this is a constant return in your while beta because you already have it already. So now what this means is that by increasing the elbow, because you haven't maintained the concept, the concept and now you are now divergence, because this is something that we don't know how to read because we don't have a forum where the last area but this is something that we want to optimize. So we want it in other words, we want to press this oval in closed form, so that we can optimize it, but that by optimizing that with respect to beta under variation, we are actually useful the difference between the 1000s So that's the question. So when I was still so far, you still need to worry about how to take advantage of the writing program. You had to write it down

1:16:40

into our evasion so

1:16:46

what will you what we're what I'm trying to tell you is that maybe I'm going to give you a new toy, this is an offline so you already have an elbow. So what I really want to show you anytime you want to maximize your own under music, that means you know as I mentioned, my dad so here you have this product rentals are independent. They're independent and you can just be popular. That's the post I want. Right now. I'm interested in estimating the fact that because you're under my rights I can take out you know, what's the rest of

1:17:24

the recipe. So

1:17:28

now you can call that a call, whatever it involves. And then the rest. I'm not going to play around with it with a map. So new arrived by this is a different town divergence from the previous slide. So this how to hurt in the sense of is the difference between specifically for a shoe beta, and then the expectation of the joint with respect to the restaurant because it's amazing what it means if you condition on the estimates for the recipe when you're estimating so this might remind you something that we just talked about, which is the loss of corn and cents. So when you when you estimate a loss all you do is cycle one step at a time and you condition on the reference, the fact that the estimates also quite similar, except for a theater or operating the variational inference or the logical case where a maximum value so in other words, to maximize this also here at a conference are just all the firm's that doesn't involve essentially replicating our concept to us because we're conditioned on them and their immune sales. Therefore they are independent from from sitting under the vacuum. So we want to we want to minimize this, right the patch er is extremely positive. So we have negative sign in front use this is zero. We have achieved the maximum of health condition on the residency. How do we make a deal? Well, I made this report that was now that urgency from zero. So the question is, how do I solve this? So that's the basic difference, right? So how do I make that? They're not we're living longer, but they'll be like, Hey, can I arrive at a closed form solution using the application over the rest of the pipeline can arrive at that is it. So that's the sort of takeaway. Yeah, folks, we can solve this. We don't have to do breathing topics. Either way more after deep learning research, everything is splitting by progress. But now you can kind of solve it. You don't have to be great enough. So that means is more efficient. So there are a lot of the armies that go into this. So we don't have fun because we're not aware of this is called a corne set of losses before they resign because you have a loss, right and you want to minimize your last boss. Your penalty for that. So you maximize, maximize. So now the solution is the following. So now if you carefully thought about firstly, condition, assessment is not what is the fact that so now about that today. It's about priorities. It's causal. We know that and follow the Gaussian. Follow this that. Now if that is the case, you can solve it and for those who haven't been really very wealthy, where to plug your immune and the various bits we find in the denominator okay at this, here, this is also quite similar to the last okay. So that is, you, you basically when you compute in fact, 530 in a day, you forget to take into account the environmental synthesis, right, it's all nice correlation with the phenotype. And now you can see the correlation between Smith and so this is a way they buy their own facts. So that is for history. I add the primary indicator, and then the source here again, we're sort of back in February the flu that date is already happening. It's already estimated, where we're trying to estimate application for siblings. So if this this is this quantity, what this means is that you have a large impact for the rest of adult children and also for the person. That means this might be likely to be a passenger's death why you Pamela, graduated from Barbados. That sense. On the other hand, if you in the coronation day are very low or normal, in fact, regardless of the fact that you've been backed by progressives are also so this is small, but then you have a larger margin and I mean, this might be more likely to deposit so that's the value of the numerous Americans instead of spending, how much variation now that's accommodation, if this is called. And now you're going to solve the other. Right? So that is what the problem is. So that's the variable the period includes in probabilities. So here, it turns out the formula we arrive at, first of all to solve this, you want to compare two phases right one is what is called or what is known. So now this together, they kind of arrive at this a sigmoid function logistic, the legit, the uj. Essentially the log of q as J equal one over to SJ, what do you think so therefore, is written in it from here, it applies to your prior belief for what the fractional policy recompiles verify. So that means it's likely that you this net is likely to be causal, sort of prior so that's probably one of the rules, right? So like, what's the weight of type one? And then this is over one minus sc, and all the other firms. So in this, you can see that there's a star there, definitely something that we already saw over here. Right now the new stars also the type of artist that has wire up, that's shorter the magnitude of the defects. So if this is verified, ratified, so the probability to me was

1:24:13

around the same probability for sniffing causal salsa. The idea that basically this is just That's it, that's the solution.

1:24:27

So once you have that, you call this when you hear it isn't the new frame that they're introducing for the day and the acid, right? So that's the actual application. So in the Bayesian, we have this as our case and maximization, well, that's the DISA use that we're solving the contribution, but now we can see that we also have this parameter. Called epsilon. Now, we also have to tie ties also know most of you don't know what's the causal fraction. We also have this beta which is universal very disclaimer first, but those are three hyper parameters. So the three hyper parameter we can solve through this empirical base, empirical basis, and once you have those expectations, you're popping back into your elbow, not going to maximize your elbow with respect to obey them. But it turns out many times when you saw that you take a part of the rotors off of your elbow, respect each other and then instead of zero, because it's often Shall we have a close look so? So the first one you can see that it essentially is the prediction error, sum of squares or mini square error, separate applications. You can use it as a beta. That's where the previous slide can tell you how often times have confused the environment where and now the data is very plain, and then the fraction you can see that it is also kind of intuitive. So the solution platform solution for five essentially sunny over all the things divided by the total number of snips. So that is a performance measure for the total number of solute solution. itself. So, so the formula is what I like to watch Saturday nights and you want a different author out his mask, you kind of write down the solution on one night. So that's about what we do here regularly that if you start it's a little too late and implement pretty much, maybe five or six lines of code. So that is you have a visa is that you compute for this year, then you compute the variance and then and then for for SJ, you compute the city for those three lines. And you find out you have an offer is that not to decide what to do with three separate branches, right if hybrid branches are very important because they affect all of your activity, provisioning. So in here, we're basically maximizing the elbow we subtracted but we can also sue them through cybercrime research using a validation set.

1:27:31

Here we all have the ultimate untold numbers are covered as measured by El Sol so when you do a rise, you're always gonna you're gonna see a very smooth although nutrients and yes, you are maximizing your your although very gracefully, but you have made a mistake with your maths or you make a mistake and you have any contagion you almost feel guilty. That means there's a blockage so that's once you have that as reality

1:28:15

Did you just mentioned that we could solve the Gaussian prior or the lattice power by using law so or arrive at regression? So I was wondering that the primary proposta spiking prior like whether if there's a frequentist interpretations, like for lasso is equivalent to we want to minimize the sum of squares by subject to the the sum of l r one norm less than some, some value and for REG is subject to some of the most

1:28:53

isolated I don't know whether it

1:28:57

was elastic net.

1:29:00

Disaster that's probably the last the last night you have lost for one hour to one hour. And now you have lambda two lambda. One lambda is controlling the territory for our one more lambda controlling the penalty for eligible.

1:29:31

All right. So I think

1:29:38

you're here I want to briefly talk about simulation. I think Saudi is going to talk about this very extensively, though ones, you're going to learn this right. So this is actually the assignments number three in my course, a lot. What you do is that you refer to take the or Talentino samples you have access to 100 minutes, you want your simulation to be small enough so you can debug it. Now, you select three columns out of one, random and now you say you're hired million, I say point five, it's not 500 but for debugging purposes, decided well beta from your so here's a sample of a call you back for three colleges. And now you have your ratio, because you want to figure out what's the variance for the environments. So now you have a square over here better than squeezing. So that's the good that mentioned the naughty how to square it, saw this graph before sigma square alpha. So now you know how much noise is given the heritability set to be certain that the mounting here you slap your your model perform accurate, your artwork or four days and this will be closing. Because at this point data, right so you're simulating the way that is similar to the model. The model is super cheap, something like that. So that's how you would have seen by designing the simulation you know that now you want to have a sample or you can have a sample from this Gaussian, which only condition on this 3d column only for for all the columns that you use later Later in the training. Has there been a night where maybe your training was plentiful? Fantastic. You're separated. And now you can come here your marginal summary statistics of the same data and then the LD that we're using the example how do you just Google the same training data? So these are your summary statistics? That's the input to your spiking snap. When an author is going to model your first few training, imagine in each of the EM iteration. So here we explain the model for adding to the rituals. You can see the model commerce offers. second iteration is already much closer to convergence. Because of the chlorinate instead of because you don't have to do with updates. These updates close work. That's the benefit of the Yeah. So here is the fall from the ballet. You're covered with using the full hour. So if you see this is nice. And now you can validate your prediction. And then you're the boss. So you hear it is it on the training. on the training side, you have my son on the test set. You have two recordings of all the one actually it's a little bit higher than the actual carpet and you have someone who's tired of it if you just sat in this simulator, we just play flat. So if I see that difference for your production coordinator, because I didn't know call is correct. So you all comers, not you and your products, and so there's only one 100 snips that's a lot of work on going to when you're trying to scale this up community.