IIT Madras

ONLINE DEGREE

Computational Thinking
Professor Madhavan Mukund
Department of Computer Science
Chennai Mathematics Institute
Professor G. Venkatesh
Indian Institute of Technology, Madras
Construction of a Predictive Analysis Algorithm

Professor Madhavan Mukund: So, we were doing this bottom-up approach, so we were trying to find out in the previous case and trying to identify the student. So, we trying to identity one of the categories on the card, basically the card has a name, so we are trying to identify the name based on some other things.

Professor G. Venkatesh: It is a kind of classification such.

Professor Madhavan Mukund: It is a kind of classification. So, is it the student or not? I will sort it. So, but we could also apply it, we were using that time the marks to determine this, but may be a more relevant thing which people the students or the teachers might be interested in is predicting the marks.

Professor G. Venkatesh: Predicting the marks.

Professor Madhavan Mukund: So, maybe you want to look at the marks in some subjects and figure out what is this.

Professor G. Venkatesh: We tried that earlier, we saw that we cannot predict physics marks.

Professor Madhavan Mukund: We could not predict physics marks, maybe supposing you take two of the subjects.

Professor G. Venkatesh: And predict what?

Professor Madhavan Mukund: Say can I estimate what the total will be because the two subjects will add up to part of the total.

Professor G. Venkatesh: Would take chemistry this time.

Professor Madhavan Mukund: Let us, take math and chemistry, maybe.

Professor G. Venkatesh: Maths and chemistry and then from math and chemistry.

Professor Madhavan Mukund: So, maths and chemistry maths is out of 100 marks, chemistry is also out of 100 marks, so if I add up I get total out of 200 marks.

Professor G. Venkatesh: So, multiply 1 by ..

Professor Madhavan Mukund: But I want to know the total of 300, so I can predict that it is to start with I can predict that it will be 1.5 times 100.

Professor G. Venkatesh: You take the sum.

Professor Madhavan Mukund: To add these things and take the sum.

Professor G. Venkatesh: I would not give you the marks I mean this giving the marks is too much, I will give you only so.

Professor Madhavan Mukund; So, maybe you can give just here some block.

Professor G. Venkatesh: Round off to the nearest 10.

Professor Madhavan Mukund: Round of too nearest 10. So, let us say.

Professor G. Venkatesh: So, let us maths in this case for example to take Tauseef, maths 90 and chemistry 40.

(Refer Slide Time: 01:51)


IIT Madras
ONLINE DEGREE
Name Tauseef
Gender M
Date of Birth 30 Dec
Town/City Trichy
Mathematics 87
Physics 86
Chemistry 43
TOTAL 216
Name Aditya
Gender M
Date of Birth 15 Mar
Town/City Vellore
Mathematics 84
Physics 92
Chemistry 76
TOTAL 252
TAUSEEF
SAMEER
ADITYA


IIT Madras
ONLINE DEGREE
Name Tauseef
Gender M
Date of Birth 30 Dec
Town/City Trichy
Mathematics 87
Physics 86
Chemistry 43
TOTAL 216
Name Aditya
Gender M
Date of Birth 15 Mar
Town/City Vellore
Name Harish
Gender M
Date of Birth 3 June
Town/City Salem
Mathematics 62
Physics 45
Chemistry 91
TOTAL 198
TAUSEEF
SAMEER
ADITYA
HARISH


IIT Madras
ONLINE DEGREE
Name Surya
Name Tauseef
Gender M
Date of Birth 30 Dec
Town/City Trichy
Mathematics 87
Physics 86
Chemistry 43
TOTAL 216
Name Rida
Gender F
Date of Birth 5 May
Town/City Chennai
Mathematics 42
Physics 53
Chemistry 78
TOTAL 173
Name Harish
Gender M
Mathematics
Physics
Chemistry
TOTAL


IIT Madras
ONLINE DEGREE
Name Surya
Name Tauseef
Gender M
Date of Birth 30 Dec
Town/City Trichy
Mathematics 87
Physics 86
Chemistry 43
TOTAL 216
Name Rida
Gender F
Date of Birth 5 May
Town/City Chennai
Name Harish
Gender M
Name Srinidhi
Gender F
Date of Birth 14 Jan
Town/City Chennai
Mathematics 52
Physics 6
Chemistry 71
TOTAL 187

Professor Madhavan Mukund: So, we will take 90 and 40. So, this is what I will say. So, mister Tauseef, so here is 90 here its 40, so this total is 130.

Professor G. Venkatesh: And you will multiply 1.5.

Professor Madhavan Mukund: Then I will multiply by 1.5.

Professor G. Venkatesh: But this might this rule is we started with this rule, they hopefully change it, you did not need to say it, we will see.

Professor Madhavan Mukund: We will see. But 195 is quite far away from 216.

Professor G. Venkatesh: It stick far away, its 216 195 that is under estimating.

Professor Madhavan Mukund: Under estimating.

Professor G. Venkatesh: So, what do I change, do I change 1.5 or.

Professor Madhavan Mukund: So, obviously it is one possibilities

Professor G. Venkatesh: Or do I change, I do not.

Professor Madhavan Mukund: So, this is like saying I am taking 1.5 times this, so I am giving equal.

Professor G. Venkatesh: Way to.

Professor Madhavan Mukund: weightage to.

Professor G. Venkatesh: Maths and chemistry.

Professor Madhavan Mukund: Maths and chemistry, but maths is a higher mark, so this is maybe getting lower because math actually is more reflective of the total for the student and chemistry is. So, therefore the total is high compared to the math marks but the chemistry is giving you a low prediction.

Professor G. Venkatesh:

Professor Madhavan Mukund: So, we could adjust this 1.5 also saying maybe maths is more and.

Professor G. Venkatesh: 71.6.

Professor Madhavan Mukund: So, let us scale this, so maybe I make this.

Professor G. Venkatesh: 0.1 you add, which 0.1 of 90 is 9. So, 195 plus 9 is 204.

Professor Madhavan Mukund: But I should reduce one chemistry also, I cannot give all because then you will get everything will be very high.

Professor G. Venkatesh: That is true.

Professor Madhavan Mukund: So, maybe these two together should add up to 3.

Professor G. Venkatesh: You are dividing three point among, 300 basically you are dividing.

Professor Madhavan Mukund: 300, so.

Professor G. Venkatesh: 160 goes there 140 comes here, something like that. You do not have the physics mark.

Professor Madhavan Mukund: I do not have the physics marks so I am trying to …

Professor G. Venkatesh: This happened this year because of Corona, they did all that, they freed out the one subject marks with another subject marks. Some algorithm they use. So, they are trying that similar algorithm.

Professor Madhavan Mukund: So, if do this for instance, what did we say we get here we will get 56 1.4 times this, so 4 so I think we get 56 instead of 60 which is what we got 1.5, but then here

we will get an extra we get 99, sorry we get 130 plus 9, 139, sorry sorry is that right? So, 90 into one and half is 135.

Professor G. Venkatesh: 91 9 plus 13.

Professor Madhavan Mukund: Plus 9, so 144 we will get.

Professor G. Venkatesh: Correct 144.

Professor Madhavan Mukund: So, now in fact my estimate goes up a little to 200.

Professor G. Venkatesh: Still not good enough, 216 or

Professor Madhavan Mukund: Still not good enough. So, we could reduce it a little more.

Professor G. Venkatesh: Make this 1.3 and 1.7.

Professor Madhavan Mukund: So, let us say 1.7 and 1.3. So, essentially.

Professor G. Venkatesh: 97 63 91 9 plus 615 153.

Professor Madhavan Mukund: We will get 153 and 52, under 52.

Professor G. Venkatesh: 13 4s are 52.

Professor Madhavan Mukund; So, now I got 205, but this is all based on one card.

Professor G. Venkatesh: One card you should.

Professor Madhavan Mukund: So, now we look at the next some random other sample maybe just take down pull out some other card, we can see whether this make sense or not.

Professor G. Venkatesh; I have taken Sameer.

Professor Madhavan Mukund: So, Sameer.

Professor G. Venkatesh: Algorithm is 1.7-time maths plus 1.3 times chemistry.

Professor Madhavan Mukund: So, first you have to round off, so we take his maths is 80.

Professor G. Venkatesh: 80.

Professor Madhavan Mukund: And say chemistry is 90, so I take 1.7 times that so that 80 and then 56, so 136.

Professor G. Venkatesh: 136.

Professor Madhavan Mukund: And 1.3 times that is 90 plus 117.

Professor G. Venkatesh: 117. Not bad

Professor Madhavan Mukund: 253.

Professor G. Venkatesh: 253, It is the better estimate for Sameer then for Tauseef actually. Let us, take one more.

Professor Madhavan Mukund; Let us, take one more and see we have stumbled upon a very good relationship, Aditya. So, again we round off some mathematics comes to 80.

Professor G. Venkatesh: 80.

Professor Madhavan Mukund: Chemistry goes to 80, so 1.7 again we have 136 and here you have a 104.

Professor G. Venkatesh: 240.

Professor Madhavan Mukund: 240 less.

Professor G. Venkatesh; it is alright.

Professor Madhavan Mukund: But not too bad.

Professor G. Venkatesh: Not too bad. Plus minus 10 I think it is okay somewhere around that, let us try this guy, Harish.

Professor Madhavan Mukund: So, here there is a big variation in the mark, so there might be some variation in our outcome also, because the maths is.

Professor G. Venkatesh: So far, this guy underestimated, this guy 253.

Professor Madhavan Mukund: Was overestimated, so maybe I should write that, so the actual thing here was.

Professor G. Venkatesh: Marginally over estimated.

Professor Madhavan Mukund: 215 actual thing here is 250 and the actual thing here was do you remember.

Professor G. Venkatesh: Aditya.

Professor Madhavan Mukund; Aditya is 250, no Aditya is 252. So, here we were quite off. So, Harish now, let us see. So, 60.

Professor G. Venkatesh: So, we underestimated this guy and this guy, but slit overestimate here. So, actually we are still airing on the.

Professor Madhavan Mukund; Lower side only.

Professor G. Venkatesh: Lower side only.

Professor Madhavan Mukund: Let us, see what we have get with Harish, so 6 so 6 7s are 42 so we should get 102 here and here we get 117, so this tells us 219.

Professor G. Venkatesh: Overestimated.

Professor Madhavan Mukund: Which is a fairly big over estimate compared to 198.

Professor G. Venkatesh: So, we should because he is got high in chemistry, so for chemistry, chemistry one more.

Professor Madhavan Mukund: But.

Professor G. Venkatesh: Looks like.

Professor Madhavan Mukund: Looks like.

Professor G. Venkatesh: 1.2 and 1.8 maybe better actually.

Professor Madhavan Mukund; So, then we will have to go back and calculate all of this case again.

Professor G. Venkatesh: Can we do that? I mean just. Just for the ... Let us just do it.

Professor Madhavan Mukund: This is to 1.2 and 1.8, so we have 90 plus 81.

Professor G. Venkatesh: 98 72.

Professor Madhavan Mukund: Sorry 162.

Professor G. Venkatesh: 162 and 48.

Professor Madhavan Mukund: 48 so we will come to 210, so now we are.

Professor G. Venkatesh: Closer much closer, much closer the to 215.

Professor Madhavan Mukund: So, here we have now 64, so 144 and 108.

Professor G. Venkatesh: 108.

Professor Madhavan Mukund: So, here 252.

Professor G. Venkatesh; Very close.

Professor Madhavan Mukund: So, this slightly gone down, but closer to that. Here it is not going to change because they have 80 80.

Professor G. Venkatesh: 80 80.

Professor Madhavan Mukund: But anyway, so we can say 144.

Professor G. Venkatesh: Plus 96.

Professor Madhavan Mukund; 96 this is going to be the same, I guess.

Professor G. Venkatesh: So, it did not change it, so it did not make it better.

Professor Madhavan Mukund: Because the two marks are the same, so it does not matter how we combine them. And here it is going to hopefully change, so 60 into 1.8 now, so this now becomes 108.

Professor G. Venkatesh: 108

Professor Madhavan Mukund: And 90 into this thing will come.

Professor G. Venkatesh: 108 again.

Professor Madhavan Mukund: 108 again.

Professor G. Venkatesh: 216

Professor Madhavan Mukund: So, it is slightly deduced.

Professor G. Venkatesh: But did not make it much better. Let us, try one more. Rida.

Professor Madhavan Mukund: Very lower maths and quite high in chemistry. So, now which will go with the most recent one 1.8.

Professor G. Venkatesh: One do that.

Professor Madhavan Mukund: So, that 72 here I think and here it is 96.

Professor G. Venkatesh: 168, not bad.

Professor Madhavan Mukund: Surprising.

Professor G. Venkatesh: Not at all bad. So, we can try on one more, let us see. Srinidhi. So, these algorithms work, take 1.2 times math and 1.8 times math and 1.2 times chemistry and add it up.

Professor Madhavan Mukund: You get something close to the good to actual total, so the. But after doing some rounding to start with so we will ignore know some, so 50 into 1.8 so that is 90, 70 into 1.2 is 84.

Professor G. Venkatesh: 174.

Professor Madhavan Mukund: We get 174.

Professor G. Venkatesh: Under estimate.

Professor Madhavan Mukund: which is a.

Professor G. Venkatesh: So, far we only under estimating, so if at all.

Professor Madhavan Mukund: So, in rare cases like in Harish case because the chemistry mark was very high mark compared to the total.

Professor G. Venkatesh: In that cases you over estimated.

Professor Madhavan Mukund: But most cases they getting an underestimated.

Professor G. Venkatesh: So, you should you even do 1.8, should even go higher, something like that.

Professor Madhavan Mukund: Maybe.

Professor G. Venkatesh: Okay let us try Geeta. So, here is another kind of algorithm, the building bottom up. What are we doing here?

Professor Madhavan Mukund: So, here we are trying to fit these.

Professor G. Venkatesh: Fit doing a fit.

Professor Madhavan Mukund: Fit these coefficients.

Professor G. Venkatesh: fit these coefficients?

Professor Madhavan Mukund: So, we have multiplying something, so we know with is linear thing that ax plus by.

Professor G. Venkatesh: ax plus by.

Professor Madhavan Mukund: So, we are trying what is a good value of a, what is a good value of b, so that ax plus by.

Professor G. Venkatesh: So, that ax plus by line, will go near.

Professor Madhavan Mukund: Near the total.

Professor G. Venkatesh: Total.

Professor Madhavan Mukund: That we want.

Professor G. Venkatesh: So, totals are plotted somewhere to it two dimensional diagram, x and y, totals are plotted on this diagram, you tried draw it, draw it for these points. See what its look like a 90 90.

Professor Madhavan Mukund: So, 90 90 is going to be I guess.

Professor G. Venkatesh: This is going to be out of range.

Professor Madhavan Mukund: Because I am going to get now.

Professor G. Venkatesh: Does not matter what it is.

Professor Madhavan Mukund: But it is going to go to 270 I think ultimately because.

Professor G. Venkatesh: 180 into 1.5, 270. So, overestimated.

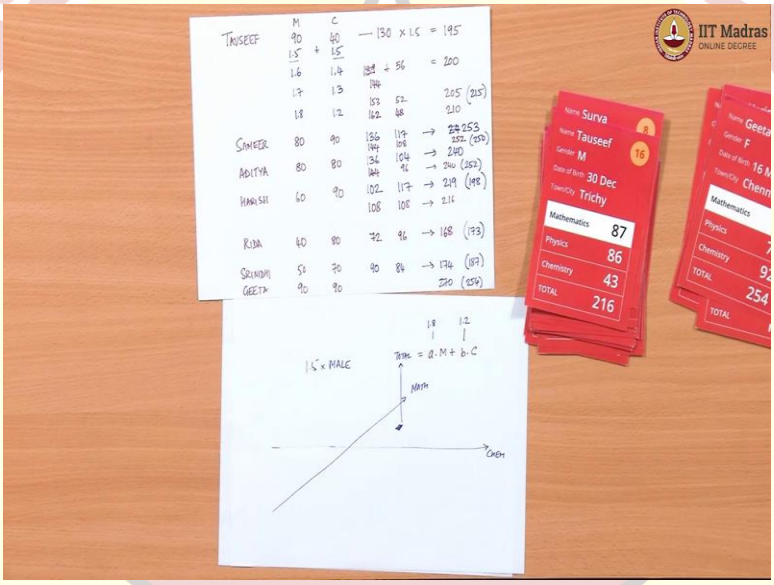
Professor Madhavan Mukund: So, I am going to and that's.

Professor G. Venkatesh: So, what we are doing, so x is basically the chemistry marks I have to say, y is the math mark. It is a 10 by 10 grid, something like that.

Professor Madhavan Mukund: so we actually.

Professor G. Venkatesh: Yeah

(Refer Slide Time: 12:19)



Professor Madhavan Mukund: So, it is like a, so it is more like a three-dimensional picture I guess so we have to draw it carefully. So, we have like, so we take chemistry here and you take say math there and then at every point you pick a combination of math and chemistry here then at this point you have some total. So, this is what we are trying to estimate. So, this is some a times maths plus b times chemistry. And so our current estimate is that we should use 1.8 for M and 1.2 for C.

And at each point we have an actual total, so we can take the actual total and compare it to that and kind of estimate whether we have got a good number or not. So, initially we gave equal, we say take 1.5 math 1.5 chemistry and that was actually not doing so well, if you go back 1.5 here for instance, you will get 1, so you may get a better thing here, here you get 180.

Professor G. Venkatesh: 1.5 50 plus 70 is 120 into 1.25 180.

Professor Madhavan Mukund: So, there is a better estimate then that.

Professor G. Venkatesh: 1.5 have been better.

Professor Madhavan Mukund: Here for instance if you take the same again it is 120 into 1.5, you will go over shoot, so in some 173 you will go high.

Professor G. Venkatesh: So, this 1.6 1.4 may have been better. So, what might be a way to do basically is that for your sample set itself, we will take various weight combinations w1 w2.

Professor Madhavan Mukund: And then for each of them you will check for that particular sample and then you would take maybe because so we saw some will overshoot some will under shoot.

Professor G. Venkatesh: Actually what you should do is some distance calculation.

Professor Madhavan Mukund: Because as long as you are close you do not care whether you are plus 2 or minus 2 they are equally bad or good, compare to plus 10 and minus 10, not uniformly plus 10.

Professor G. Venkatesh: So, you should take the difference and square it.

Professor Madhavan Mukund: We could do that so as make it a positive number, you can square that difference.

Professor G. Venkatesh: Also square sum of the distance x square plus y square square root of this distance, so you take the difference.

Professor Madhavan Mukund: So, here you would say.

Professor G. Venkatesh: 10 is a difference.

Professor Madhavan Mukund: 10 is difference so 10 square or here is 3 or 2 whichever one, so the last thing about is 5 actually, so 5 is 1.8 .12 the distance is 5 square 25, here the distance is 2 square 4 here, here it is 8 square so on. So, we will keep doing.

Professor G. Venkatesh: Square of the distances.

Professor Madhavan Mukund: And then we will add it up and add it up.

Professor G. Venkatesh: And see here which of these combinations that is.

Professor Madhavan Mukund: For each combinations we will get some in some sense this errors square errors sum of the squared errors, so you square the difference each time add it up, so you get the sum of the squared errors and you want to find that combination for which that some is.

Professor G. Venkatesh: The least.

Professor Madhavan Mukund: The least. So, actually we have been of course picking numbers which are like one point there is nothing to stop us from 1.75.

Professor G. Venkatesh: 1.75.

Professor Madhavan Mukund: And also the we have arbitrary decided that they should add up to 3, it could be that.

Professor G. Venkatesh: did not be 3.

Professor Madhavan Mukund: Did not be 3 or it could be something else. So, there is a more complex search for these things I guess to.

Professor G. Venkatesh: But if you do that is same as various a math method is there of doing it. If you do that, then you will get some coefficients w1 and w2 wn of it. Based on samples again and based on that you create this function, so you will get this aM plus bC and it is kind of predicts, some error is there but the error is kept minimum as possible and hopefully that after some number of samples are seen you can apply it now for any other card predicting the total. So, that do not get it exact, more or less approximately you can get it correct.

Professor Madhavan Mukund: So, as you said like in this board exam situation, if somebody misses one exam and you want to fill it in.

Professor G. Venkatesh: Fill it in.

Professor Madhavan Mukund: Then if you know the score and two exams that they have taken then you can make some based on this pattern you can say okay I will give you this.

Professor G. Venkatesh: About this.

Professor Madhavan Mukund: About this total to you and therefore this is implicitly this is the physics marks that you should have got. Now, of course that student happen to be bad in physics that would be good for them because they would get a higher score, if they were very good in physics and the physics marks was boosting the total then it will be a disadvantage, but for most people hopefully this will be correct because this is the best fit. So, that is one way you can use this.

Professor G. Venkatesh: So, both these methods what we saw basically making the tree decision tree to categorize or classify or whatever it is the item as well as this which is prediction.

Professor Madhavan Mukund: The numerical prediction.

Professor G. Venkatesh; Numerical prediction type stuff here, but you need only numbers right, numerical predication is only prediction numbers.

Professor Madhavan Mukund: so, it would not make any sense, we cannot put we cannot say 1.5 times male.

Professor G. Venkatesh: Male.

Professor Madhavan Mukund: It does not any sense so this will not make any sense, about 2.5 any numbers.

Professor G. Venkatesh: Are you going to convert M into a number? Meaningless it will be.

Professor Madhavan Mukund: It will do not make sense.

Professor G. Venkatesh: Unless I know the average for males.

Professor Madhavan Mukund: Yeah

Professor G. Venkatesh: So, you get average for male.

Professor Madhavan Mukund: So, you actually do some splitting, you can say for males you have one equation for females you have another.

Professor G. Venkatesh: Males have an average and females have and average, I can make two equations.

Professor Madhavan Mukund: You can find a different a and b for male and different a and b for female and also we can do. So, you can.

Professor G. Venkatesh: So, you can do part of this and part of that, partly tree, first you do tree first you do mf.

Professor Madhavan Mukund: So, you group them into two groups and in each group you find.

Professor G. Venkatesh: The one a and b,

Professor Madhavan Mukund: But because perhaps the performance is different based on that now it may be that the correct grouping is gender it may be the correct grouping is city.

Professor G. Venkatesh: Chennai matters code very high, so you do not know.

Professor Madhavan Mukund: So, but by grouping you might get for a that group you might get a better estimate than for the thing as a whole.

Professor G. Venkatesh: So, in all these things whether it is this that or a combination now you have seen a combination also, what you do basically is a slightly different method from what we have done earlier, you take a sample bunch of sample cards and then you from that you try to build this what built it up

Professor Madhavan Mukund: Your code in some sense. This is also in some sense this is a procedure.

Professor G. Venkatesh: Coming out as procedure.

Professor Madhavan Mukund: This is a procedure your procedure is.

Professor G. Venkatesh: But you did not write the procedure.

Professor Madhavan Mukund: Procedure is take two parameters M and C and return back aM plus bC.

Professor G. Venkatesh: aM plus bC.

Professor Madhavan Mukund: Where aM I do not know in advance.

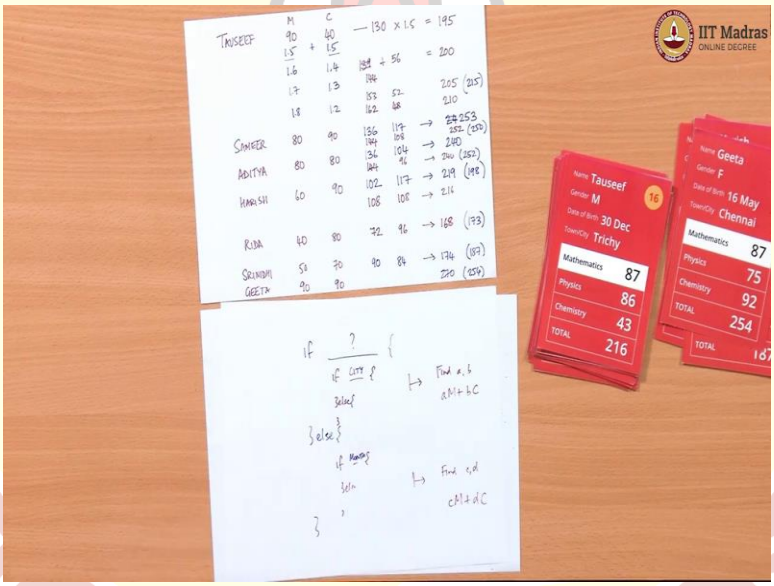
Professor G. Venkatesh: a and b are in advance

Professor Madhavan Mukund: When I wrote the procedure, I did not know the a and b, but I found them out by looking at a bunch of samples. So, that is a crucial thing.

Professor G. Venkatesh: And in the other case also we said that if there are two procedures, suppose one in which you start with grades, then after that you do gender another we start with town and then do something, something else it is a male, then you basically try all these three different procedure, try this, try this, try this and see which one works out.

Professor Madhavan Mukund: So, because there if you look at the procedure, it is basically saying something like I mean we have in a way what we are doing so fixing a skeleton, things we are saying if then we have some condition to be asked and then and then we have else. So, we know that our first question comes here, but we do not know what questions to ask here.

(Refer Slide Time: 19:26)



So, we try grade and then we say if we say grade will the procedure work better, if we say city will it work better, if we say gender will it work better and so on. And then within this again there is going to be some once we have decided on that then again inside there is going to be another second question.

Professor G. Venkatesh: Or it could be a numerically equation like this.

Professor Madhavan Mukund: Or it could be an equation. So, we could now do that also and so. So, here are the other option is here instead we could say find a b for and here we find a different.

Professor G. Venkatesh: Different pair.

Professor Madhavan Mukund: Maybe some c d, so will get C M plus and these two may not be the same, because at this top level when we.

Professor G. Venkatesh: There we saw, actually when we are going on the tree different question could be asked depending on which branch of tree. In one case you could have gender other case you could ask town.

Professor Madhavan Mukund: So, here if you are in this case then we would have if we had another if here, then this question could be some so this could be something else this could be city and this could turn out to be whatever month, it is possible.

Professor G. Venkatesh: So, unlike a normal algorithm here actually this criterion what criterion to check itself is an emerging from data.

Professor Madhavan Mukund: So, what we are fixing is kind of the.

Professor G. Venkatesh: Skeleton.

Professor Madhavan Mukund: Skeleton the structure of what the program is supposed to look like, what type of decisions is try to make, but what are the actual checks that it makes for those decisions we are not trying to.

Professor G. Venkatesh: Fix.

Professor Madhavan Mukund: We do not know. And the same thing might be different even if we change to a different class or a different group different school, so they so this pattern that we derive for one group.

Professor G. Venkatesh: It is not work is class.

Professor Madhavan Mukund: It may not work in another class, we might know in general that there is a relationship between.

Professor G. Venkatesh: But this I will get that method is skeleton can be carried there and applied to the data there.

Professor Madhavan Mukund; And even in the same class it I mean the same school it might vary year to year, so over a period of time, so if we applied this year we might get some number next year if you apply to the classroom I get a different number. So, that also.

Professor G. Venkatesh: So, you are expecting that what you should do first is you to take for any given situation you want to do this method, you take some samples from that.

Professor Madhavan Mukund: So, you must make sure that the set samples.

Professor G. Venkatesh: Do not take make samples on somewhere else and try to apply.

Professor Madhavan Mukund: Samples must be.

Professor G. Venkatesh: From that.

Professor Madhavan Mukund: Form the they should represent the type.

Professor G. Venkatesh: Represent that kind.

Professor Madhavan Mukund: Otherwise, you are not going to make any sensible prediction. so you cannot.

Professor G. Venkatesh: You take samples from there, but skeleton is there, then use that skeleton basically to populate, once you got that then you use that to do whatever you want to do classification prediction whatever.