Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-15

So, let us kind of look at dou L, you know what have what are we indicating this by, so here if you see right, so we are calling it as B i L right at any L th layer. So, some suppose I take B i 3 ok, let me say that right I need dou B i 3 ok. Now, dou L by dou B i 3, so what is your what is your L now, so L you had as let us just go and look at what your L was can you can somebody tell me what was it. So, L L was 1 by 2 summation Y i hat minus Y i square right this is what you had and therefore, what will happen. So, if you do dou L by dou B i 3 right, so this will become, so 2 will cancel off we will get Y i hat minus Y i and then you will get dou Y i hat by dou B i 3 ok. And Y i hat is nothing but ok, but there is I think a simpler form let me just see ok, sometimes right people kind of get away with a simple form, but let us see.

So, Y i hat is what Y i hat is A i of what is it just go back to that figure Y i hat is A i 4 right. So, Y i hat is A i 1 minute that is see that is the notation we use know A i 4 know. So, A i 4 and A i 4 itself yeah, so A i 4 itself yeah I think. So, what you do is right you play this trick that well it is simple right.

So, what you do is you write this as dou A i 4 by dou Z i 4 yeah right and then you write this as dou Z i 4 by because you need a direct relation between Z i and B i right. So, you write this as dou B i 3 because I mean eventually you want this with respect to B i 3 right this is what you want, but Y i hat is A i 4 and A i 4 I mean and your relation is with B i 3 is directly related to Z i 4 right. I mean that comes from that equation right Z i 4 is in terms of W i j 3's and B i 3's right and therefore, right what happens. So, this is nothing, but F dash of Z i 4 and Z i 4 right is in terms of there is just an addition right plus B i 3 therefore, that will be just 1 right this last term is just 1 and therefore, right. So, you can sort of write this dou L by dou B i 3 right to be equal to Y hat minus Y i and then dou Y i hat by dou B i 3 which is F dash of Z 4 Z i 4.

So, again right this is something that you can evaluate right which means that which means that you can change your B i 3. So, how will you change your B i 3? You will say that B i 3 right in the next iteration right and the next iteration will be B i 3 old and then minus some step size times dou L by dou B i 3 that is how you will use this right this gradient right in order to update. So, you just have to go back to that to the gradient descent

right. So, theta n plus 1 is equal to theta n minus alpha times dou L by dou theta evaluated at theta equal to theta n right this is how this is that equation. So, just that theta can be weight theta can be bias and that the it is the gradient that you need to be able to evaluate and theta equal to theta it means that whatever was whatever is the current value of B that you have for you know within the say network plug that in.

Once you plug that in you will get values for all of this and which will then mean that this will be finally, a number for you right and that you multiply and then you update your B i right and then this is how you you look but of course, you know there are there are certain things like you know how you do the training and so on it is not like you push all the examples at one go okay. So, that we will come to okay once once we get there but for the time being right it is just enough that in that we understand this now let us just look at what do you. So, yeah so I think this let us just fill this up here it is Y i hat minus Y i F dash of Z i 4 what about let us say we also need dou L by dou B i 2 okay and dou L by dou okay well 1 minute right. So, see I told you right so yeah so there is there is actually a simpler form see okay right this is also correct that is why I said it is no there is no unique way to do this but then right look at the simpler form there because you want everything in terms of this right delta. So, what you do is you do write this is dou L by dou Z i 4 to dou Z i 4 I thought that there was a simpler form dou B i 3.

So, you go back to the diagram right. So, you see looking at looking at right dou L by dou L by you see dou B i 3 right and it is simply dou L by dou Z i 4 right into dou Z i 4 by dou dou you see right right B i 3 and.. Yeah exactly no the point is that what I am saying is you do this because you want everything in terms of the delta I mean the other thing is also correct there is nothing wrong with that I mean that is the way I would impulsively proceed but then the whole idea is that you want to boil everything down to a delta rule right. So, what you do is instead of so this is correct but then this is not what you actually do right.

So, what you do is this and this you know is right delta i 4 and this you saw here right dou dou Z i 4 by dou B i 3 was 1 and therefore, it is simply delta i 4 and this you can apply everywhere right even at see B i 2 also see if I do dou L by dou B i 2 right I can come straight away from this right. So, if I want let us say B i 2 then what will I do I will say I will say dou L by dou Z i 3 and dou Z i 3 by into dou Z i 3 by dou by dou C B i 2 right which then means that I will again write say delta whatever right. So, I will get you know delta 3 and then the other term is again 1 and therefore, right in general. So, in general right what you have is dou L by dou B i L will then be what dou i L plus 1 see right we do this one. So, that you know we know as to how this works ok that is all I mean not that right not that you know this is when during implementation only you do this but I thought it is worth right doing it just once ok.

So, you have like dou i dou i L plus 1 ok that is for dou L by dou B i L. Now let us I thought we will also do maybe the matrix form of this. So, which one shall we do first. So, we have this equation right what do you have you have delta i where is that delta i L it is equal to summation j is equal to 1 to some what is we write as L plus 1 then we wrote delta j L plus 1 and what is that W j i L and then into f dash of Z i of L right this is what you had know. So, now suppose you wanted to suppose you wanted to find out right delta 1 let us say 3 ok I mean I am just doing it for just one case ok.

Suppose you wanted to wanted to do it for you know delta 1 3 that means you go back to this diagram always right always you know keep this in mind therefore right you are kind of looking at this delta 1 3 because assuming that you know delta 4 we already know ok. So, then i equal to 1 right and therefore, so how many how many neurons here right. So, your summation will go from j equal to 1 to 2 you are going to the fourth layer right L plus 1 therefore, 2 right you are going to 1 to 2 and then delta j ok. So, then right what will what will this be like. So, right let me just write this down delta 1 3 is equal to summation j equal to 1 to 2 right 1 to 2 and then you will have delta j L plus 1 is 4 right and then W then you have j and then I am looking at i 1 then L is 3 and then F dash of Z i is 1 L is 3 and therefore, if I expand this what will you get you get delta 1 4 all these are scalars ok.

So, therefore, it is ok to multiply whichever in whichever order ok. So, W 1 1 3 plus just let us put in a bracket plus delta 2 4 W 2 1 3 into F dash of Z 1 of 3 correct. Now, if you try to do what try to write down I mean write. So, also how many how many do we have like this delta in the I mean third layer you got 1 2 3 right therefore, therefore, let us also write the next one. So, can we write you know delta 2 3 what do you think what do you think right that will be can somebody tell quickly delta 2 3 will be delta 1 4 W 2 1 3 plus delta 2 4 W 2 2 3 right into F dash of Z 2 3 right and then we can have delta 3 delta sorry what happened the delta j equal to 1 to 2 j is 2 oh no no 1 minute 1 minute I think no no ok no this should be other way ok W i is 2 right therefore, this will become W 1 2 3 and this will become W 2 2 3 correct because j i you know.

So, so see j is inside. So, j is summing over 1 and 2 that is standard, but i is what is changing right from here to here what is changed is i therefore, it will become W W 1 2 3 and then W 2 2 3 this is right and then and then what will happen to you know delta 3 3 going in the same way we will have delta 1 4 and then j is 1 first and then 1 3 3 plus delta 2 4 W 2 3 3 into F dash of Z 3 3 right I mean this just follow right whatever we have done. So, what this means is that right I mean you know I could actually update the whole thing right together I mean if I mean I could actually instead of writing them you know individually like this what I could do is I could actually think about writing this as a vector right where I have because each of these is a number right. So, delta 1 3 delta 2 3 and then

let me say I have delta 3 3 and therefore, right I need I need this to express in terms of a matrix and then right see if you go back and then see right what are what are the terms right. So, what are the terms that you have you have actually delta 1 4 and you see right delta 2 4 right.

So, let us say I have delta 1 4 and delta 2 4 and and then right you should have have and you see element wise a multiplication because this is a vector see here here everything is is a scalar multiplication right. This is tricky see here right. So, what you have is into F dash into into F dash of Z 2 right. So, it is equivalent to getting this matrix vector product get a get a vector and multiply it right element wise multiply the two vectors element wise. So, what will that mean? So, see so your first.

So, you have like what delta 1 3 should be w 1 1 3 w 2 3 ok w 1 1 3 delta ok. So, so what will be the first two entries w 1 3 and w 2 1 3 right this will be the first two first two entries right w 1 1 3 w 2 1 3 what is the what is the next guy delta 2 3 is w 1 2 and w 2 2 3 yeah w 1 2 3 w 2 2 3 and then this will be w 1 3 3 w 2 3 3 right and here you will have a vector. So, this is this is element wise multiplication ok this is element wise multiplication and this will be like what is it what was that F dash of Z i Z 1 3 Z 2 3 and Z 3 3 right. So, then so see the the standard notation for for a for a weight matrix right let us say suppose I call this is w 3 I mean in that layer right this is we are looking at the third right we are looking del equal to 3 you know. So, this standard notation would have been like w 1 1 3 w 1 2 3 w 1 3 3 this is how you would write it in a standard form like w 2 1 3 w 2 2 3 wait a minute w 2 2 3 and then w 2 3 3 this is how you would have written.

Therefore right you can think of the matrix above as a kind of a transpose of this right. So, this is like 3 cross 2 this is how you would have written the standard weight matrix and all your unknowns are these w's ok. So, so then write this if I call as delta 3 as a vector right then I have like delta 3 right is equal to w 3 I will write this as a transpose right w 3 transpose and then multiplying and and this multiplication is as a standard matrix vector multiplication ok not element wise. And then I can write this as this is kind of what will this be delta 2 right this sorry delta 4 this is like delta 4 see this is delta 4 this is in the fourth layer and then you can just think of this as one vector that you get by multiplying and then you have this one I think let us call this as f dash of what is this z 3. And all the all all these places we have to know as to what it contains right I mean that let us let us assume that we know how many elements are sitting there and you know and the and the size and all right the dimensions and all are all correct right.

So, so in a way right so in general ok this I will just leave it to you right I mean you can do it for any layer for that matter. So, we can write this as delta L right. So, we can write this as w L transpose delta L plus 1 right and the whole thing into f dash of z L right. So,

so in a way write in one shot you know because finally, write during implementation all people always like to have a matrix vector form because so in one shot right when you can get you can get at everything out. It is all one and the same, but you know sometimes implementation wise write it matters I mean how you how you write these things right.

So, therefore, this is this is a this is a matrix vector form ok and this called you see right delta delta learning rule and the word back you know propagation right when you say back propagation say what you really mean is. So, the so let me let me just give you the exact ok somebody said right that day that this came in 1986 right. So, the exact this is the principle right let me just write down when somebody ask you what exactly I mean this is all the math part, but what exactly I trying to do is a principled method ok involving chain rule as you saw to update the weights and the biases ok. When they say weights I mean the biases also to update the weights by back propagating the error by back propagating the error. So, that is how the network error reduces error reduces right this in a sense is what is you know this the kind of summary of what actually back propagation is all about and how you implement right is what we have seen ok.

Then I will just leave it to you as an as maybe right as an exercise or maybe right we will just we will just find out right right right. So, what do you think some of these are pretty straight forward what do you think I mean if I had let us say what do I have here dou J dou L by dou B L where B L is again right I mean. So, if you were to stack the B i L straight at that at that place right as a vector what do you think this should be equal to we already saw know what was delta L delta L or L plus 1 no is it L plus 1 no delta L plus 1 not delta L delta L plus 1, but as a this one vector ok. So, the bias part right is very simple if you write it in a in a vector form I will just leave this to you right as a small exercise ok.

Let me just ask you know. So, how would you write let us say W let me see right somebody would be able to answer this let us say dou J by dou C W what do I have here ok alright. So, for this example it is a that we had suppose I did dou J by C right dou C W 3 ok. Let me just write down. So, what do you what I mean I will just write down this expression right which you have already found out by the way right. So, W i J 3 was actually delta i 4 a J 3 right this is what we had know right I mean no no right in the start right when we did dou J by dou dou W i J 3 we got delta i 4 right a J 3 and therefore, right I mean if I try to see dou J by dou C W 1 1 3 again right depending upon what my i and J are running from let me just write down i is running from 1 to 2 j is running from 1 to 3 in this case for that layer that is how it is running.

Therefore, you will have a overall 6 terms right W 1 1 3 W 1 2 3 W whatever right for the for the i j combinations and therefore, right if you had to write it down and dou J by dou W 3 ok. So, W 3 is actually matrix right this is that matrix that I wrote down which

was a 2 cross 3 matrix right and therefore, J is a scalar by the way right L right we are using L right. So, L is a scalar therefore, right this will be this will be a matrix again right what will what will that how will you express it in terms of delta and a J? . Outer product very good right.

So, outer product. So, this will be delta L plus 1 into a L transpose ok this is just just when you see outer product right. So, you have something like this right. So, you have this vector and you know column vector and then another row vector right. So, and in general right this you can write down in general as dou J no no ok this is for L in fact, that is for L. If it was 3 it would have been delta 4 whatever right I mean a 3 transpose that is what it have been, but in general right you can write it down and these are all vectors by the way a L on top L is always coming on top right.

So, a L transpose. So, this is an outer product ok. So, so what is right what this kind of essentially means is that all these updates can be done ok you can have you can have as many layers as you want and you can always back propagate and and you know and then sense what is the error right that you are making right now and therefore, how should the how should the weights be updated. So, that the new weights are such that the overall error if you think of that cost or landscape right that we had for the cost function right you are trying to traverse on that and you are trying to traverse in a sort of a direction that will that will reduce the error and which way to go is coming out of the out of these equations right and therefore, and you can imagine right people are doing it over you see millions of millions of weights and biases. So, so that way it is very powerful for us right we just took a small example just to just to have some insights into how this works works, but think about the way it is all this is been implemented right people have like a million parameters 60 million, 70 million parameters out there and the whole thing works. So, very smoothly right and there is a lot of work that is gone into actually building those packages and all ok.

So, so that you know you can do this quickly and you know make make use of you know all kinds of you know matrix properties and all in order to be able to you know do this in a fast manner ok. Now, yeah so is there anything else that I wanted to say in this ok and then by the way right. So, what I will do is when we come to convolutional networks which we will come very shortly right we would not we would not again solve all of this right this back prop and I will just leave it to you ok because once you do it for MLP I mean you know that becomes in fact a simpler case of this right because I mean you know you have convolution and all going on. Therefore, it we do not want to sort of spend too much time right do I just thought we will do this once. So, that right for those of you right who have not seen back prop right I mean you will at least have seen it once right in your life ok. So, I think we will stop here.