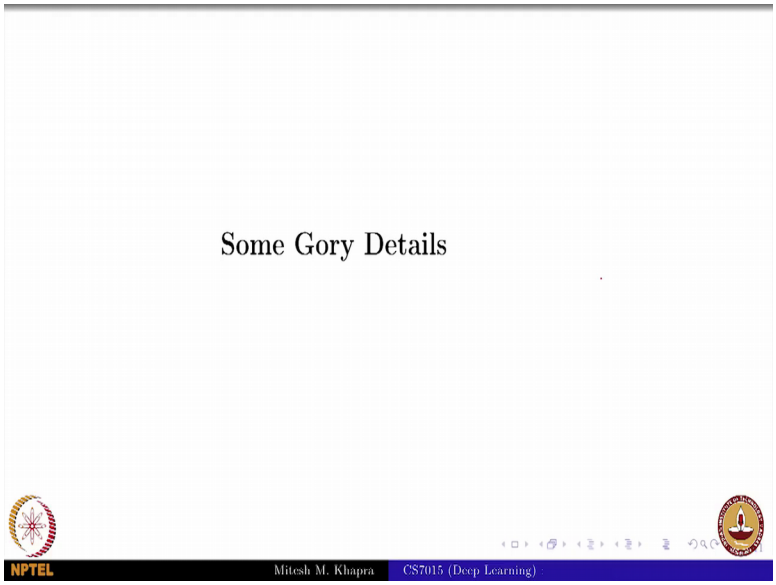
Deep Learning
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

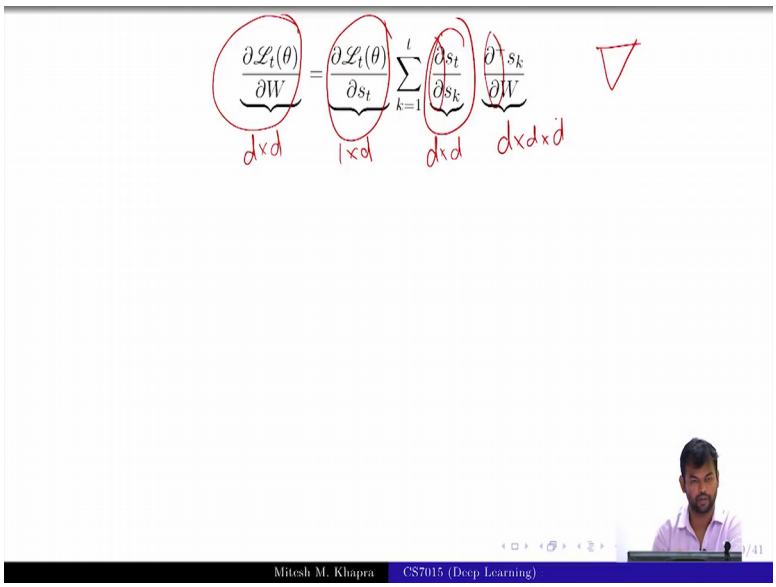
Lecture – 107
Some Gory Details

(Refer Slide Time: 00:11)



Before that I will just go into Some more Gory Details about the math, this is not to scare, but just to make you more comfortable that we can actually deal with something, which is not very straightforward or very neat as compared to what you are seen so far.

(Refer Slide Time: 00:25)



So, I just go back to the formula, which I had for computing the gradient of the loss function with respect to W, and I cannot repeat enough times that, all these notations are actually a bit of abuse of notation, because these are gradients and not partial derivatives. So, I should actually be using this notation, but for ease of explanation, I use I stick to my original notations ok.

So now, let us look at each of these quantities here and tell me the dimensions of these quantities? Let us start with the left hand side, what is the dimension of this? W was what matrix? What is I am talking about the circle entity, what is the what is the dimension of that? K cross d.

Student: (Refer Time: 01:09).

D cross d.

Student: (Refer Time: 01:11).

D cross d someone n cross d.

Student: (Refer Time: 01:13).

N cross d and n cross k are the two options, which are left. W is the recurrent weight. So, W is what dimension?

Student: (Refer Time: 01:20).

D cross d.

So, what is this gradient, d cross d? Ok what about this fast? S t what the hidden representation, so that was d dimensional. So, what is this d cross 1? Ok what about this? Why do you guys still struggle with this?

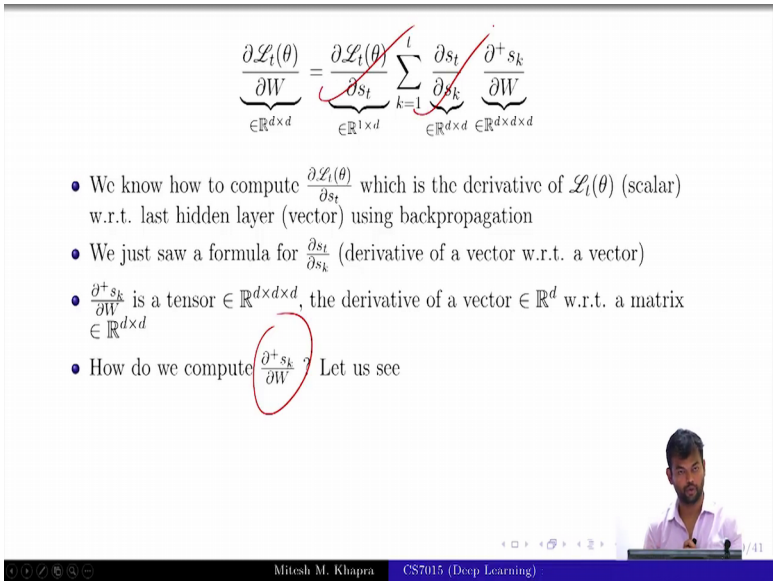
Student: (Refer Time: 01:42).

D cross d and this d cross

Student: (Refer Time: 01:44).

D cross it is very straightforward right, what is the dimension of numerator? what is the dimension of denominator that is all right.

(Refer Slide Time: 01:53)



So, you see the kind of multiplication that you are doing here. Say of d cross, d 1 cross d, d cross d and then d cross, d cross d ok. Let us look at each of these quantities and see if you are actually comfortable in implementing these. Are you comfortable with this? The loss function with respect to the hidden representation, we have done this enough times in back propagation, what about this? We just saw a formula for this right. So, we know how to compute this quantity, we have seen this in back propagation, this is the derivative of a scalar with respect to a vector and we are very comfortable in computing this.

This was slightly tricky, but we just derive this formula on the previous slides, everyone with that, what about this? This is a tensor, how do we compute this tensor? what is our standard recipe focus on?

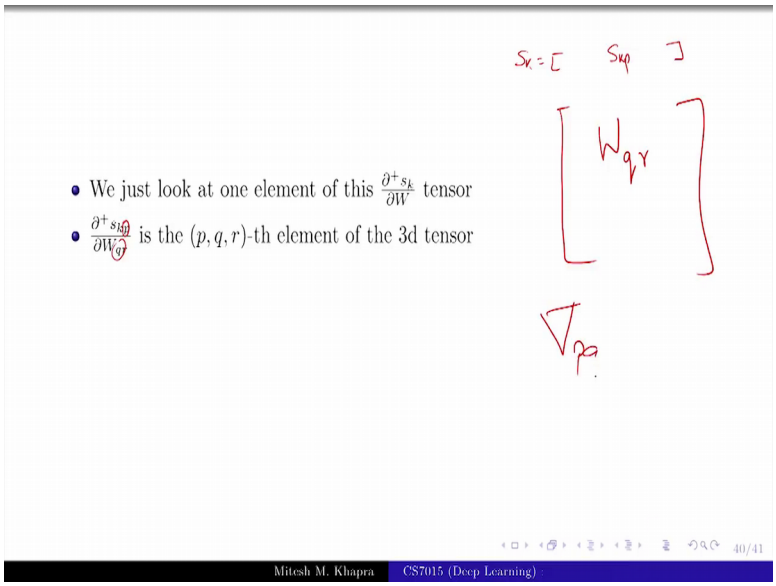
Student: (Refer Time: 02:38).

The little guy, 1 element of this tensor and then you can generalize somewhere right. So, this is the tensor and we will just see that, this just to make you all comfortable is this not like just to intimated you with all these large sized tensors, but I am just trying to show

that, this is all easy, this is not hard ok. So how do we compute this? All the other terms are covered.

This is the only one, that we do not know.

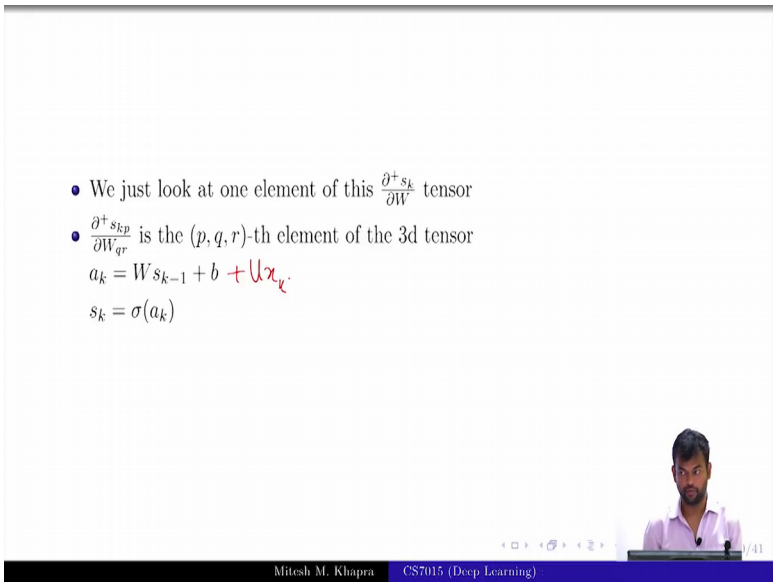
(Refer Slide Time: 03:03)



So, we will just look at 1 element of this tensor and it is going to be S k p by W q r.

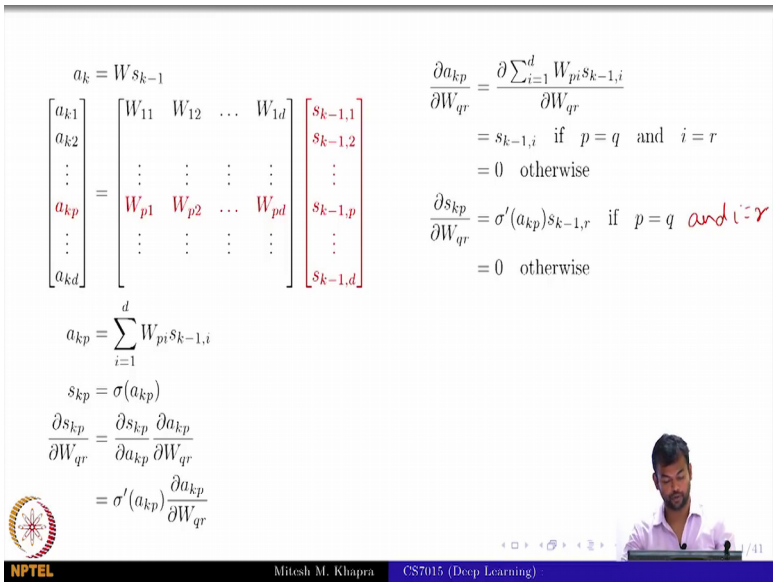
So, let us just see that you have S k as this vector and you have W as this matrix. So, I am considering one such weight, which is, W p comma q and one such element from here, which is S k sorry. So, q r and I am considering one element from this, which is S k p. So, I am trying to compute the derivative of one element of the vector with respect to one element of the matrix. So, this is going to give me one entry, in my tensor and that entry is going to be what? P q r, how many of you are fine with this? Ok fine.

(Refer Slide Time: 03:53)



So, now recall that a k was equal to W into S k minus 1 plus b and S k was sigmoid of a k. I think again, I have miss that U into x k, but that will not matter because, that is not there in the derivative ok. You are fine with so far?

(Refer Slide Time: 04:11)



So now, let us look at this, because the other 2 terms do not matter. So, I just look at a k is equal to W into S k minus 1. So, this is the matrix way of writing it now I am looking at one of these elements which actually comes, from the multiplication of a row and a column the highlighted row and the column everyone gets this ok.

Now so, I can write it as a k p is actually equal to this summation, which is nothing, but the dot product of this row with this column ok. Now S k p is just the sigmoid of that. So now, if I want to compute S k p with respect to W q r, I can just write the chain rule that S k p with respect to a k p which is straightforward and then a k p with respect to W q r and I already have a formula for a k p, how many of you are fine so far? Please raise your hands high up, if you are fine ok. So, what is the first term going to be? Sigma.

Student: (Refer Time: 05:10).

Sigma prime of a k p and what is the second term going to be? This is what, the second term is now, what this is lot of terms here, which of these terms would actually remain? Only the once where, only the terms where i is equal to.

Student: (Refer Time: 05:28).

R and.

Student: (Refer Time: 05:30).

P is equal to q. So, only that term will remain. In that case, it would be this right and in the other cases going to be 0 right. So, now, you have one element of this tensor and you have it as a very generic formula, you can just fill in all the elements of the tensor right. So, what does this tensor look like? It is a very.

Student: (Refer Time: 05:52).

Sparse tensor right that is all I wanted to convey ok. So, this is again the same thing right, is that fine? So, even though it is a nasty looking tensor, if we just break it down to one element, it is going to be very easy and now from this element, you can just reconstruct the entire tensor. Do not worry, I am not going to ask you to implement this, but if someone were to, maybe at some point, then you should be able to do it right, that is where we will end today.

So, we have finished recurrent neural networks and the next thing that, we are going to look at is LSTMs and gated recurrent (Refer Time: 06:25) ok.

Thank you.