Computational Neuroscience
Dr. Sharba Bandyopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology Kharagpur
Week – 10
Lecture – 50

Lecture 50 : Synaptic Normalization

Welcome. So we have been discussing long term plasticity implementation in computational models using different kinds of rules of update of synaptic weights or actually we have been talking about how the change in synaptic weights take place with using the derivative of the elements of the weight matrix with respect to time and that is some function of the input activity u which is multiple number of inputs and the single output variable v. And so we have these rules that we have considered are the correlation based rule, the covariance based rule and the BCM learning rule. And so we saw that the BCM learning rule overcomes most of the I mean rather all the problems that have that are associated with the Hebbian based correlation based learning rule or covariance based learning rule and appears to work very well. However there is one issue that comes in is that the weight updates stability that is the stability of the weight from or the length of the weight not growing unbounded that is indirectly controlled through the output activity. So in other forms there are normalization models, normalization based rules that actually are constrained the weights directly and so we have two kinds of normalization based models one is the subtractive normalization and the other is multiplicative or in fact it is actually divisive multiplicative or divisive normalization.

And this is also called Ojah's rule. However we must also say that these rules that we will talk about the subtractive and multiplicative rules they are not totally biologically grounded as we will see the subtractive rule somehow requires information from all the synapses to each of the synapses which is maybe possible but there is very I mean no evidence in support of that. And similarly the multiplicative normalization is more a mathematical way to normalize it rather than how actual sort of normalization works in terms of homeostatic plasticity where the overall weight is overall response activity or output activity is over long periods of time is that average is maintained. However so since these are used in a number of computational models that include plasticity and these are some of the basic rules we will be covering them in this course in particularly in this lecture.

So what do we mean by the subtractive learning rule essentially we introduce this idea that our the sum of our weights that is W_1 plus W_2 up to our W_{NB} where there are W_{NB} weights this is constrained to be constant. So in other words if we define a vector N that is all unit all elements are unity then $N \cdot W$ is what this sum is the dot product of N and W. So our learning rule that is our $\tau_W \frac{dW}{dT}$ is given by our VU as we always have and that restriction is brought in by VNQ times N divided or N vector N times NB. So this is the correlation based increasing apart that is the weights would be increasing so that is countered by this particular term which is based on again the both the output activity and the input activity and it is it is made to be of this particular form. So that ultimately our sum of W this NW does not change.

So for the $N \cdot W$ for not changing the derivative of $N \cdot W$ must be 0 and so let us see what we have so $D(N \cdot W)$ which is the this is the sum of all the weights this will we require this to be constant and that is why that additional term has been introduced and so this is nothing but the N vector dot $\frac{dW}{dT}$ and our $\frac{dW}{dT}$ from our previous expression we have $\frac{dW}{dT}$ equals V sorry we have VU minus V times $N \cdot U$ times N divided by NB. So now we simply plug this in into this term here $\frac{dW}{dT}$ and so we have N and dot product with VU minus $VN \cdot U$ and N divided by NB. So now if we take the N in here what we will get is V times $N \cdot U$ in the first term. So this becomes $VN \cdot U$ because V is a scalar we can take it out and N is multiplied by U and we also have the same $VN \cdot U$ here and we can plug in the N inside times $N \cdot N$ divided by NB and so if we take $VN \cdot U$ as a factor out what we get is $1 - N \cdot N$ divided by NB. So what is our $D(N \cdot W)$ that is the sum of the weights coming out to be here as you see if we multiply N with N $N \cdot N$ we are essentially having 1 1 1 this vector multiplied by 1 1 1 this vector.

So it is basically sum of NB ones because this number of elements is NB and here also number of elements is NB so this becomes 1 so sorry this becomes NB. So what we have is if this becomes NB then this whole term goes to 0 and so what we have is our $N \cdot W$ is a constant. So there cannot be any unbounded growth here and we still have a competition among the synapses that is when one keeps on increasing or few of them keep on increasing the others are bound to reduce and go down to keep the overall sum of weights constant. And while this is when thinking of just the mathematical treatment this is very suitable kind of rule there is hardly much evidence to justify this clearly although this rule can be used to explain certain phenomena in networks but this again brings up the issue of the biological relevance and how we want to implement certain things. So if the objective is to solve a particular problem we and if this rule helps us to solve that problem that is fine but if the objective is rather to understand and build on the rules that are biologically based and see how they are different then such a rule actually would not be that much useful because while we do understand the correlation based increase in weights which is sort of the Hebbian rule and we have discussed the biological relevance of that as well this subtraction this normalization subtractive normalization is not entirely physically true because as you see the for weights to decrease or increase the information about all the other weights on the neuron is required at a particular synapse.

So that becomes a difficult proposition while it is it may be possible through intracellular signaling in some particular way or through even activity in some particular way but so far there is not much evidence in favor of this rule being true in real neuron. So in the next rule what we will be studying is as we have said the multiplicative learning rule or the multiplicative normalization in this case again so there we had a subtractive term that is subtracted here also we will have a term that is subtracted but it will be proportional to the weight itself and so what is this is the Ojah's rule what we again have is $\tau_W \frac{dW}{dT}$ and this is again we have our term vu and then we have our αv^2 and this term multiplication by w. So since we the update is proportional to this weight w this is what we call for this is what we mean by the multiplicative normalization the normalization that is brought in by a factor of the weight itself. So why is this I mean what does it achieve in terms of the normalization so here if we look at the norm of the weight so again if we look at our norm the length of the vector and we look at the derivative of it then and multiply it by τ_W then what we have again is τ w times $2w$ and vu minus $\alpha v^2 w$. So τ_W can be dropped because this is this is our τ_W into $\frac{dW}{dT}$ so then what we have here is twice our v into $w \cdot u$ minus αv^2 into $w \cdot w$ which is simply the norm of w.

So what is our $w \cdot u$ $w \cdot u$ again is our v as we have always had so we have 2 times v^2 minus αv^2 norm of w^2. So we have twice v^2 times $1 - \alpha w^2$ or actually we can say twice v^2 by α sorry α $1 - \alpha - w$ norm square. So what we have the derivative of the norm norm of the weight that is the Euclidean here in this case Euclidean length of the weight vector that is the dynamics of it that is the derivative of it in steady state that is when there will be no change finally it will always relax to $1/\alpha$ because ultimately when the system stops that is when the when we reach no change in weights the norm of $\frac{dW}{dT}$ being 0 the our w norm will approach $1/\alpha$ and so essentially that introduces a constraint on the length of the vector and so here also we see that there is no unconstrained growth of the weights because ultimately the w reaches a fixed value. So both these normalization rules are useful for many computational models however in terms of real neurons they have many elements that are not really supported by evidence so far. So for us the main rule turns out to be essentially the STDP based learning rule and that we will say will be representing by in this case τ_W then $\frac{dW}{dT}$ and so remember if we have a learning window that is based on let us say this particular learning rule that we have let us say this versus this let us say this is any function $H(\tau)$ on this axis.

So this is the pre minus post or the timing difference. So if we assume that over time the all of them all of the spike timing is actually getting implemented all the spike time differences are actually getting implemented and there is a continuous spiking going on in terms of the probability of spiking occurring and so lot of the ensemble average then these $\frac{dW}{dT}$ can be represented by a sum of 0 to infinity $H(\tau)$ times $v(t)$ and $u(t-\tau)$ this term we have $d\tau$ and plus $H(-\tau)v(t-\tau)u(t+\tau)$. So essentially here this particular term is all the LTP that is occurring and this is the contribution of LTD. So here you see that our previous v is correlated with the current u and weighted by the H in the backward direction in the negative axis because we are integrating over 0 to infinity and here we are correlating $v(t)$ and $u(t-\ \tau)$ where is that is the past u is being correlated with the current output which is 2 and is weighted by this $H(\tau)$ depending on τ this correlation will be changing and this particular term is for the right hand side of the learning rule. So here it is the change in weight that we had discussed.

So the LTP and LTD are both included and the difference from what we had seen earlier in terms of the experimental observation of spike time independent plasticity we have changed it in the sense that we are now basing it on average rate responses assuming that there is a probability of spiking associated with that output rate and so we can indeed set up this model to explain the change in weight based on the spike timing dependent plasticity. And now with variety of $H(\tau)$s or even various depending on how the different systems different neurons and synapses work we can modify the use in a manner that is based on observation or even the reads that is based on observation like we were talking about bursting activity and so on. We can have a general sort of model for change or update of weights. So with these discussions of synaptic weight update implementations we come to the conclusion of the modeling of long term plasticity and next we will be talking about how these are implemented in networks what are the consequences of both short term and long term plasticity ah in our coming lectures. Thank you.