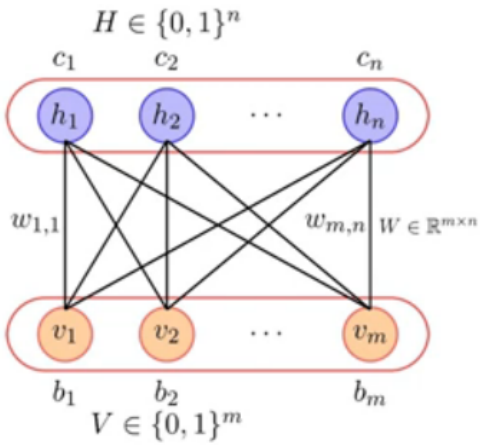
Lecture - 18.4
RBMs as Stochastic Neural Networks

Refer Slide Time :(0: 14)



$$E(V,H) = -\sum_i \sum_j w_{ij} v_i h_j - \sum_i b_i v_i - \sum_j c_j h_j$$

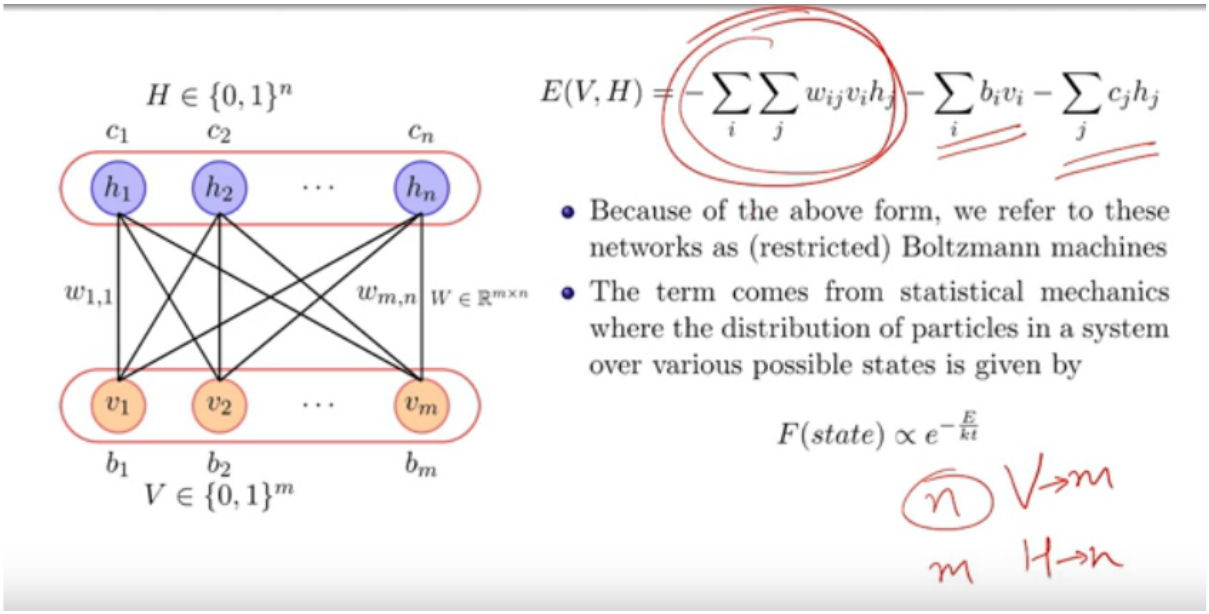
- Because of the above form, we refer to these networks as (restricted) Boltzmann machines
- The term comes from statistical mechanics where the distribution of particles in a system over various possible states is given by

$$F(state) \propto e^{-\frac{E}{kt}}$$

which is called the Boltzmann distribution or the Gibbs distribution

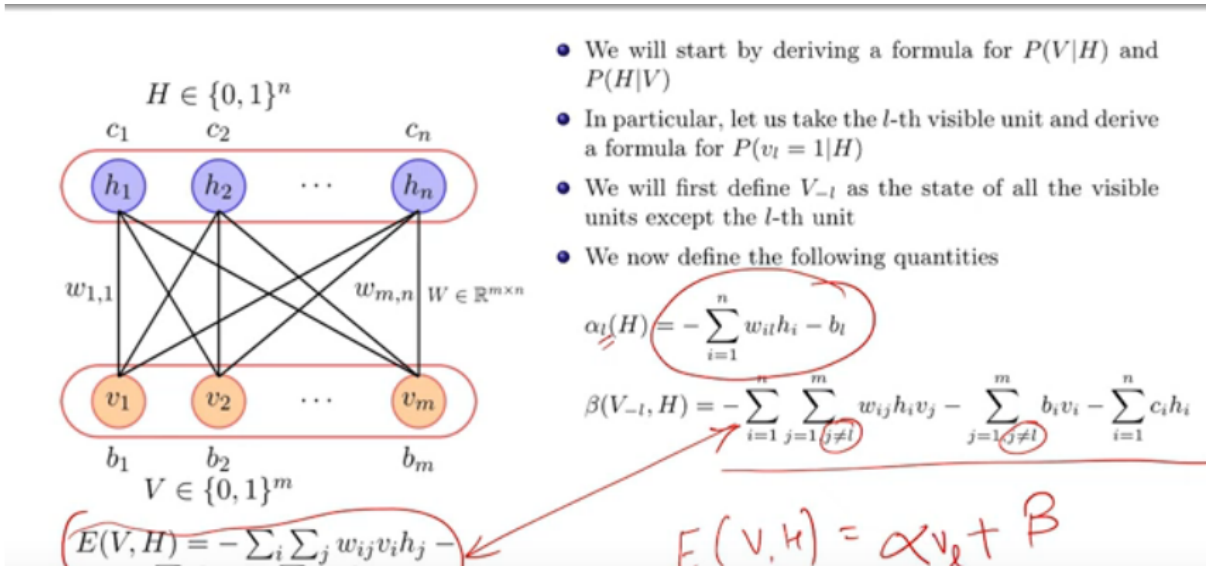
Okay, so welcome back so just before the quiz we will looking at restricted Boltzmann machines and we started with a motivation of a undirected graphical model and specifically we introduce the consent of a lantern variable and send an all the visible variables in stood of directly depending on each other, actually depend on each other through the lantern variables. So, this is the neural network that, this is the graphical model that we had, which you see on the left hand side and you have the visible variables and the hidden variables and for this discussion we had assume: that all of these are binary and what with the factors in this graphical model? What with the factors? Adjust between the visible variables and the hidden variables. Right? So, you had these, how many factors did you have? Encores in factors and what is the parametric form that we choose for the factors, we said that each factor depends on the corresponding visible and hidden variable, between which the factor is and there is weight associated with that. Right? So that's what we decided it? Vi hj those are visible hidden variables in the pay and Wij was the weight, associated with that and then addition we also, had factors for the visible variables and the hidden variables and we said all of these legal as long has been compute the partisan function so that the entire thing its normalize as a distribution. Okay?

Refer Slide Time :(1: 39)



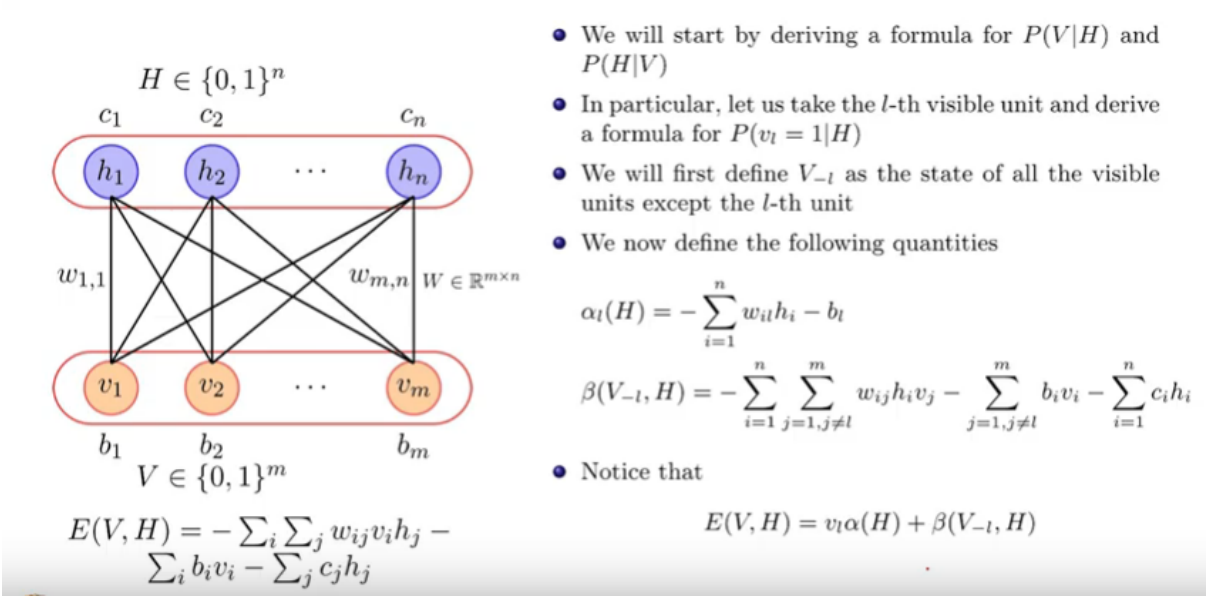
So, if I pick one, if I pick one visible variable, how many terms in this summation correspond that? Remember the this is a double summation by question is as I pick one visible variable, how many terms in this summation correspond that? Assuming there are n visible variables and m hidden variables, my question is if I take one of these guys, m of those and what about this how many some terms in this summation correspond to one and about this one. So, you just remember this we need this for deriving certain things here. So, there are m terms here corresponding to each, visible variable that we have is that fine? And I think yeah! Sorry, this is m and this is n just the confusion there in needed are the beginning are the hi. So there are m visible variables and n hidden variables so that means if I repeat my question, how many terms summation correspond to this? N. Okay? Fine. So, now with that back ground, what you want to do is? Justify that why we are teaching this in a course on deep learning. So, we have to show: that RBMs have some finishing to neural networks and that's what we will see in this modules, which is we will write to see, RBMs has Stochastic Neural Network. Okay?

Refer Slide Time :(3: 02)



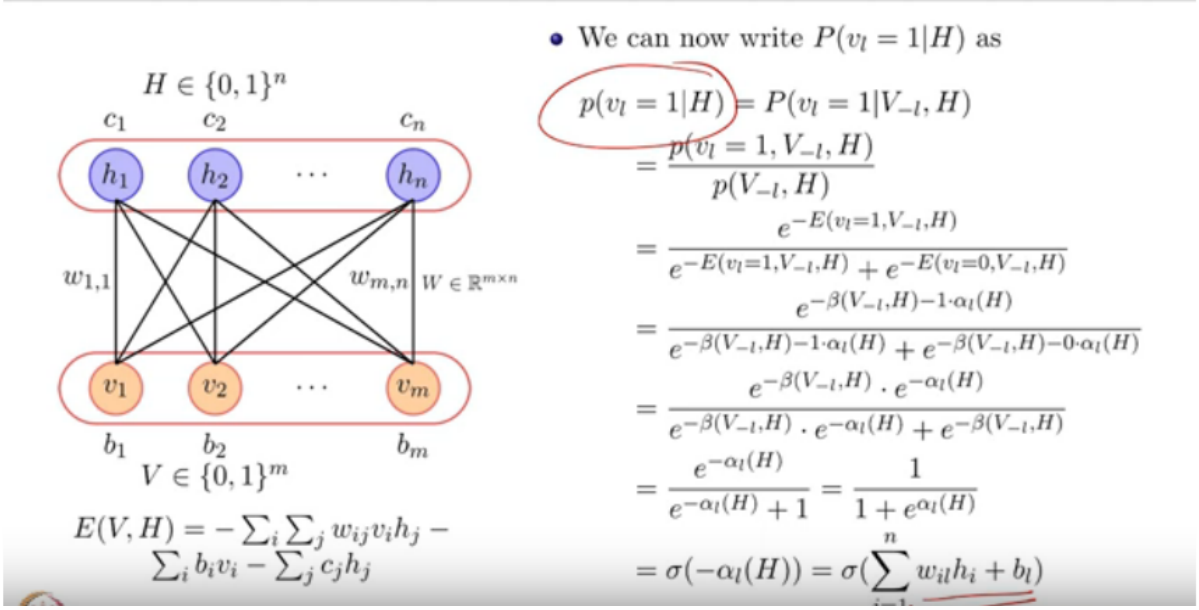
So, what will do is? Given all this definition and notation that we have, in particular this energy function that we have, will start to derived the formula for P of V given h and P of h given v can I do that? Why can I do that? Because once a have a join distribution I can compute all kind of conditions and marginal involving those random variables. Right? So, let's look at one of the visible unit, one of the visible variables. Okay? Let's look at one of the visible unit, variables and I am interest in getting a formula for that visible variable taking on the value one. Right? That's what I am interest. Okay? Now, what I define is I'll define a term units V subscribed minus l, which essentially to denote this state of all the variables, except this elite variables. So, when I say V minus l, at means all the other visible except this except this variable is that clear? Okay? The set of all visible variables except the one which an interested in. An I define two quantities the first one is alpha, which correspond to elite unit, what does alpha actually capture? It's relevant to the question which had ask you know, it is, in that in this, formula what does alpha l actually capture? All the terms corresponding to d elite unit. Okay? That's what it captures is that clear? How many if you get that? Please raise your hands okay. Now, what does beta capture, focus on this. Actually if you look at these two, these two equation look exactly the same, except that I have removed the terms from the summation which correspond to elite term. Okay? Now, in terms of alpha in beta, can you give me your formula for E? E of V comma h is equal to something, which is alpha in beta. Alpha plus beta, is that correct? Your saying alpha plus beta, alpha V l plus beta, how many if you agree with this? Okay that's straight forward to see.

Refer Slide Time :(5: 22)



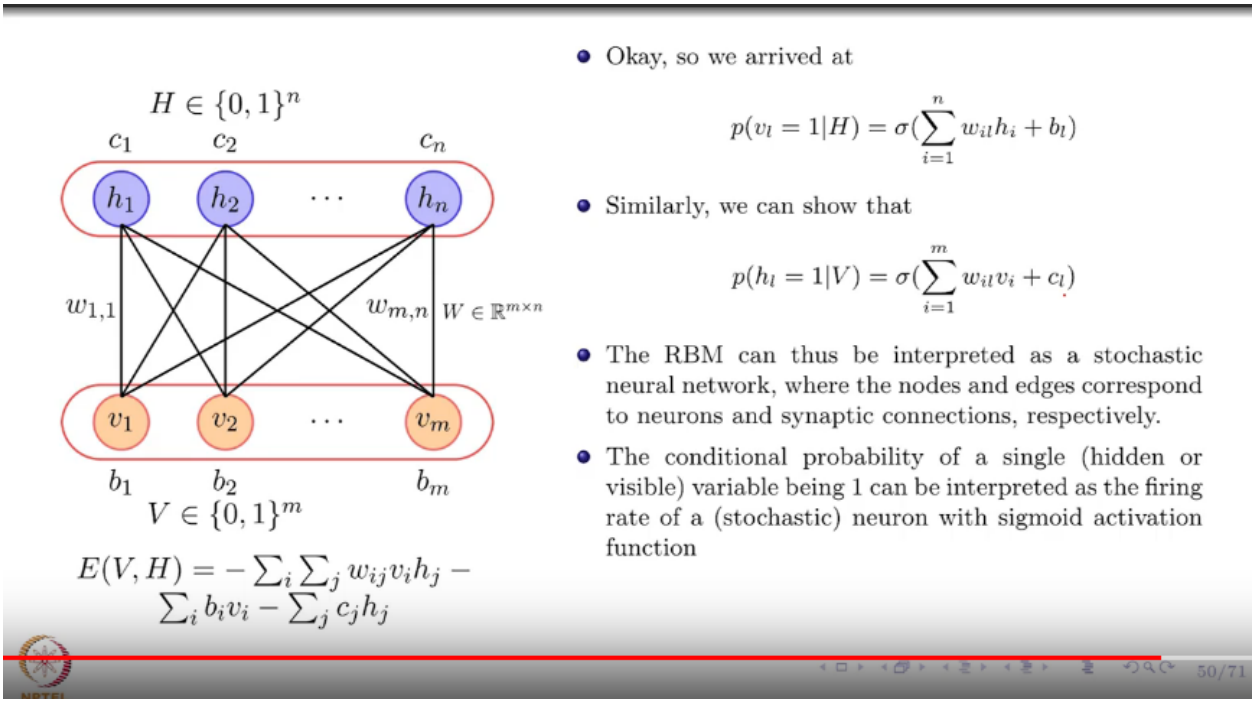
So that's what E of V comma h right, its V l terms alpha plus beta. Right? And I am just retain this particular broken it down this means so that, later on then we deriving something, it would be easy to cancel out some terms and so on. Right? So, with definition,

Refer Slide Time :(5: 39)



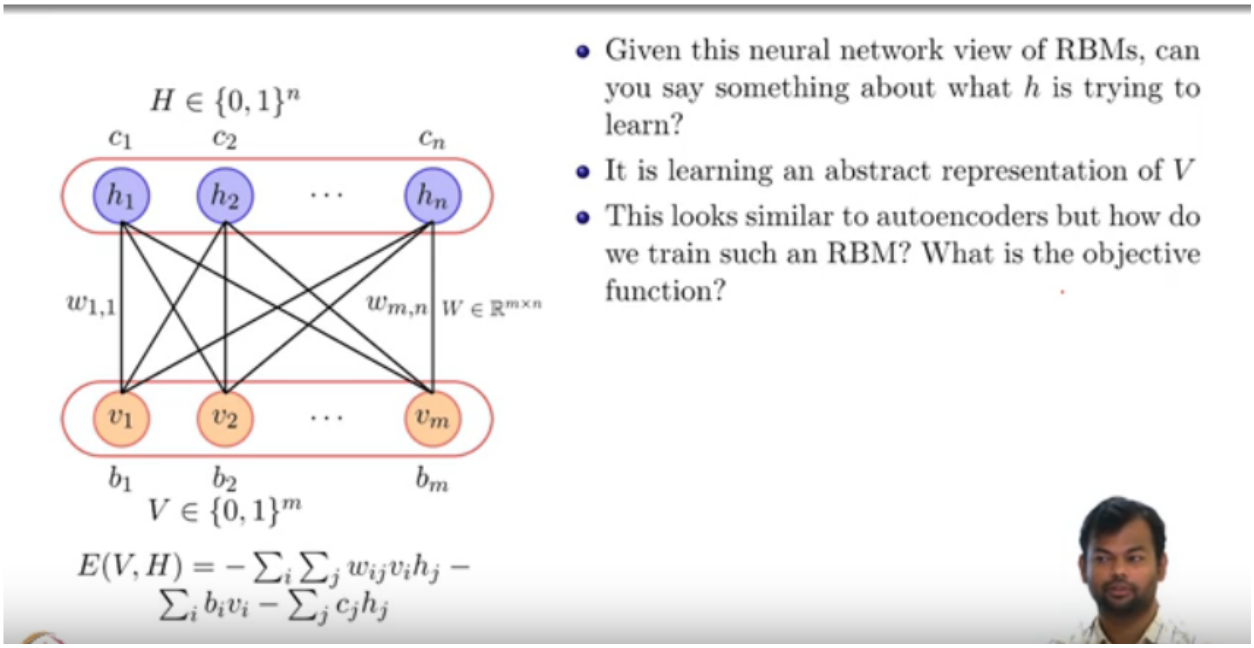
Now, let straight to compute, a formula for V of l equal to one, given h. And why don't I have V comma h here given Actually put the, in this actually a not ask but at last put in the that leaned. Is that fine? Because v are independent, given the H. Right? Is simple that. Okay? So, I can all way introducing these variable doesn't matter is that fine, because these to are equivalent. Now I can write as this. Okay? Which Okay? First tell me denominator I know p of V comma H, I know p of V comma H how do you I get p of V minus L comma H I have to, how many elements dose is, is how many random values is, is as m plus n, how many random values is, is as have m minus one plus n. So, now what do to have do to? Have do to operation starts with m and n, which ready marginalize over elite unit that means some over all possible values, of the elite unit. Is that fine? Okay? And that is exactly what the denominator is don't focus on the numerate just focus denominator. Right? A have taken both values of VL and compete probability and this same over that is that fine denominator just marginalize, over the elite unit ends. How many get that? Okay? How many get numerate that simple the numerate as all the emblems variables and in particular I know the values fun of the variable. So, just plug that into the energy function. Is that fine? Right? So, remember energy function as these vi hj. Right? So, I know value of the vi so I am going to plugging that particular value for other vi. Is that fine? Okay? Now base on the Alberta business that we just that I am going to just rewrite some of these, thinks remember E was Alpha into VL plus beta. Okay? So, this numerate clear let's focus on numerate first I know that VL equal to one subject substitute one there I am beta have it is as it is how many get the numerate please raise your hand. Okay? And denominator is again, some think variable similar alpha plus beta; I know the alpha value of VL is one in the case and VL is zero in the case. Is that fine? Now what you do next? What is common factor here? E bar minus beta factors common. Right? Okay? Okay this of cores since zero it at be e raise to zero, which is one so, just beta part remains here. Now if take a e raise minus beta factor common what will get, we have seen the ever in the life sigma of I what the alpha LH? Minus of this quantity. Right? Is that fine? So, the probability of the elite unit benign one is actually is sigmoid of the weight some of the all the incoming connection doesn't make since how many if get what I set the probability of the elite unit benign one is as actually the of the weight some of the all the input connected to it. And what are input connected to the visible unit all the elite units. Right? Such a weight some that.

Refer slide time :(09:49)



So, I can write it as this, similarly if I want to compute, the property of the elite it an unit taking on the value one by the kind of derivation I will get this form. Right? That is again sigmoid weight some of the all the inputs connected to it. What are input connect to it all the visible units? Is that fine? Okay? And now you can tell me why did me have the distance factors the biosis. Okay? So, now can you think of this as a neural network? Right? We have same form as we as seen neural network, it can be interpreted as a stochastic neural network, which tells as what is the probability of the hidden unit firing, given curtail certain inputs or alternative we were interesting both abstraction and generation. So, I am doing abstraction what my interesting computing what given what everyone hidden given the observe variable, I am taking generation, what my interesting computing, visible units given the hidden variables. Right? So in both of these things, I can think of it as, finding how what is the probability of hidden unit being on, given a configuration of visible units. And alternately, what is the probability of visible units being on, given a particular configuration of the hidden units, is that fine. So both of these can be interpreted as the stochastic neural network, where I'm getting certain probabilities, from a computation which looks very much like a, neural network computation. Okay?

Refer slide time: (11:23)



So, given this neural network interpretation of RBMs, and given the answer which I have already given you, what does h actually compute? Your ability disappointing at dash representation, I thought says start with A and the D. Okay? It is learning an abstract representation of V. So this is very smellier to what? What kind of neural network? Order encoder, why do you say order encoder, we do the same in multi layer perception also right, where given input, we have multiple layers and then compute every layer computes, some representation of the input. What will we the difference between the training data for an auto encoder and multi-layer perception and supervise? In the case of multi-layer perception, we had X comma Y are objective function was some squander are loss different on the, defined on Y or some cross enterprise loss defined on Y. Right? That was supervising learning, we will learning task specific abstract representation, in the case of auto encoders what was are training data? Just X. Right? Just the visible variables, no hidden variables of course and no Y's also. In the case of RBMs what are training data? Same as what we have into order, just the visible units, we don't have label associated with them and of course, hidden variables are hard of the question right, when these are R imagination no once give as that. Right? So hence, we are comparing this to hand or auto encoder. Now what should be the objective function? Now, I bolted down to the neural network, once we bring out to the neural network what are favorite learning algorithm? Okay, back propagation okay fine. So that's are favorite algorithm that's going to be, so data I told you, model is RBMs, parameter are obvious, learning algorithm is define, what's remaining, what should be the objective function? Next lecture means, next module. So all that we will see it soon, we will see it soon, we will see it soon, all that will now that today's soon. So what is the objective function? You want to maximize something, what you want to maximize? Profits. What is the past few lectures being about? Linear algebra or maximize scale evidence. Okay? You will get it, before I tell you. Maximize something, what can I maximize? Someone side, likely hood. Okay?