Parameters vs Hyperparameters

(DESCRIPTION)
Text, Deep Neural Networks. Parameters versus Hyperparameters. Website, deep learning, dot, A.I.
(SPEECH)
being effective in developing your deep neural Nets requires that you not only organize your parameters well but also your hyper parameters so what are hyper parameters let's take a look
(DESCRIPTION)
New slide, What are hyperparameters? Text, Parameters are W to the 1, B to the 1, W to the 2, B to the 2, W to the 3, B to the 3, and so on.
(SPEECH)
so the parameters your model are W and B and there are other things you need to tell your learning algorithm such as the learning rate alpha because on we need to set alpha and that in turn will determine how your parameters evolve or maybe the number of iterations of gradient descent you carry out your learning algorithm has other you know numbers that you need to set such as the number of hidden layers so we call that capital L or the number of hidden units right such as zero and one and two and so on and then you also have the choice of activation function do you want to use a rel you or ten age or a sigma little something especially in the hidden layers and so all of these things are things that you need to tell your learning algorithm and so these are parameters that control the ultimate parameters W and B and so we call all of these things below hyper parameters because these things like alpha the learning rate the number of iterations number of hidden layers and so on these are all parameters that control W and B so we call these things hyper parameters because it is the hyper parameters that you know somehow determine the final value of the parameters W and B that you end up with in fact deep learning has a lot of different hyper parameters later in the later course we'll see other hyper parameters as well such as the momentum term the mini batch size various forms of regularization parameters and so on and if none of these terms at the bottom make sense yet don't worry about it we'll talk about them in the second course because deep learning has so many hyper parameters in contrast to earlier errors of machine learning I'm going to try to be very consistent in calling the learning rate alpha a hyper parameter rather than calling the parameter I think in earlier eras of machine learning when we didn't have so many hyper parameters most of us used to be a bit slow up here and just call alpha a parameter and technically alpha is a parameter but is a parameter that determines the real parameters our childhood consistent in calling these things like alpha the number of iterations and so on hyper parameters so when you're training a deep net for your own application you find that there may be a lot of possible settings for the hyper parameters that you need to just try out so apply
(DESCRIPTION)
New slide, Applied deep learning is a very empirical process.
(SPEECH)
deep learning today is a very imperiled process where often you might have an idea for example you might have an idea for the best value for the learning rate you might say well maybe alpha equals 0.01 I want to try that then you implemented try it out and then see how that works and then based on that outcome you might say you know what I've changed online I want to

increase the learning rate to 0.05 and so if you're not sure what's the best value for the learning ready-to-use you might try one value of the learning rate alpha and see their cost function j go down like this then you might try a larger value for the learning rate alpha and see the cost function blow up and diverge then you might try another version and see it go down really fast it's inverse to higher value you might try another version and see it you know see the cost function J do that then I'll be China so the values you might say okay looks like this the value of alpha gives me a pretty fast learning and allows me to converge to a lower cost function jennice I'm going to use this value of alpha you saw in a previous slide that there are a lot of different hybrid parameters and it turns out that when you're starting on the new application I should find it very difficult to know in advance exactly what's the best value of the hyper parameters so what often happen is you just have to try out many different values and go around this cycle your trial some value really try five hidden layers with this many number of hidden units implement that see if it works and then iterate so the title of this slide is that apply deep learning is very empirical process and empirical process is maybe a fancy way of saying you just have to try a lot of things and see what works another effect I've seen is that deep learning today is applied to so many problems ranging from computer vision to speech recognition to natural language processing to a lot of structured data applications such as maybe a online advertising or web search or product recommendations and so on and what I've seen is that first I've seen researchers from one discipline any one of these try to go to a different one and sometimes the intuitions about hyper parameters carries over and sometimes it doesn't so I often advise people especially when starting on a new problem to just try out a range of values and see what works and then mix course we'll see a systematic way we'll see some systematic ways for trying out a range of values all right and second even if you're working on one application for a long time you know maybe you're working on online advertising as you make progress on the problem is quite possible there the best value for the learning rate a number of hidden units and so on might change so even if you tune your system to the best value of hyper parameters to daily as possible you find that the best value might change a year from now maybe because the computer infrastructure I'd be it you know CPUs or the type of GPU running on or something has changed but so maybe one rule of thumb is you know every now and then maybe every few months if you're working on a problem for an extended period of time for many years just try a few values for the hyper parameters and double check if there's a better value for the hyper parameters and as you do so you slowly gain intuition as well about the hyper parameters that work best for your problems and I know that this might seem like an unsatisfying part of deep learning that you just have to try on all the values for these hyper parameters but maybe this is one area where deep learning research is still advancing and maybe over time we'll be able to give better guidance for the best hyper parameters to use but it's also possible that because CPUs and GPUs and networks and data says are all changing and it is possible that the guidance won't to converge for some time and you just need to keep trying out different values and evaluate them on a hold on cross-validation set or something and pick the value that works for your problems so that was a brief discussion of hyper parameters in the second course we'll also give some suggestions for how to systematically explore the space of hyper parameters but by now you actually have pretty much all the tools you need to do their programming exercise before you do that adjust or share view one more set of ideas which is I often ask what does deep learning have to do the human brain