Quantum Draft

June 8, 2017

1 Simulated Annealing and Learning

Finding the minimum or maximum of some function is an essential tool in many learning algorithms. Most problems of interest can be described in terms of some loss function we seek to minimize given our training data and model parameters. Finding parameters to minimize this function is the heart of machine learning, and it is an arduous task. Non-trivial problems have dizzingly complex feature spaces that are nigh-impossible to navigate. The number of possible solutions grows exponentially with the number of parameters we use, and we cannot be sure any proposed solution is optimal. Exhaustive search is not viable in the least, so there exists a plethora of methods to estimate good parameters, among which include simulated annealing.

Simulated annealing is a random statistical process that allows us to move around a model's parameter space in search of a global optimum. Broadly speaking, the process involves starting with some set of parameters, making semi-random alterations to those parameters, evaluating the new parameters, choosing whether to stay with the old or new parameters, then repeating all previous steps. The "annealing" part of the method comes from our probability of accepting the new parameters: the higher the "heat" of our process, the higher chance we have of accepting "bad" parameter alterations in the hopes we move out of local minima/maxima. The start of annealing has high heat so we explore larger portions of the feature space whereas the end of annealing has low heat, giving us better precision and local search. Although simulated annealing has proven useful for optimization problems, there always remains the problem of getting trapped in local optima. Here, quantum devices may prove useful.

Quantum tunneling is a phenomenon where an annealing process on a quantum device can escape local optima [1] whereas a classical annealing process may remain stuck. We may get better parameters using quantum annealing, or at least we may get a different set of optimal parameters. Devices like the D-Wave promise such quantum effects and it is in our interest to investigate how we might exploit these reported benefits for use in deep learning algorithms.

2 Quantum Annealing and Deep Learning

The following attempts to draw a clearer connection between deep learning algorithms and quantum annealing on the D-Wave device. A deep learning architecture is most simply described as a neural network with many stacked layers, and each layer's neural units are connected to the neural units of the layer above. The deep learning problem is to find a set of connection weights J and unit biases h that will produce the correct output when given a particular input, where each neural unit $\sigma_i \in \{0,1\}$ is either "on" or "off". It so happens that such a connection weight and bias description is easily converted into an Ising model with the following energy expression:

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i$$

Happily for us, this is also exactly the type of optimization problem D-Wave is set up to solve. We can translate our deep learning architecture into a form D-Wave solves and vice versa.

What may not be obvious at first glance is why this could be a potential benefit to deep learning aside from the previously mentioned quantum tunneling effect, which might give different or better answers to the same optimization problems. Normally, deep learning architectures only allow connections between units of adjacent layers; units within the same layer or units in distant layers do not interact. This is done to make computation feasible, otherwise the Ising model is an NP-complete problem [2]. However, what D-Wave offers is the ability to help optimize deep learning architectures that allow connections between neural units in the same layer. This expands our possible network architecture choices and configurations which may not be feasibly computed on classical machines. Some preliminary work suggests that inclusion of extra connections between intralayer units may produce better results for some learning problems [insert something about our own ongoing work?].

However, there are limitations to what is currently possible. The D-Wave hardware architecture is limited by physical space, so each qubit can only sport six connections to other qubits. Real problems are often solved by networks with thousands of connections, and the human brain has over 100 trillion connections. Either we constrain ourselves to toy problems or attempt ot approximate real networks using only 6 connections per qubit. Introducing additional connection capacity in future quantum devices would prove a boon to deep learning development using quantum annealing.

Another issue is hardware precision. The physical nature of the D-Wave quantum device means we often have noise in our J and σ parameters and results. Additionally, the range of parameters, themselves, is constrained; we cannot physically tune them to an arbitrary magnitude. These phenomena, along with the qubit connection limitation, restricts what types of problems we can realistically tackle. Quantum devices are already iterating, however, with more qubits and precision being added with time, so we may yet see improvement in hardware that enables additional research into exploiting quanutm annealing for deep learning.

References

[1] Boixo, Sergio, et al. "Evidence for quantum annealing with more than one hundred qubits." Nature Physics 10.3 (2014): 218-224.

[2] Cipra, Barry A. "The Ising model is NP-complete." SIAM News 33.6 (2000): 1-3.