Cauchy Machines

Mariana Ávalos Arce   Daniel Heráclito Pérez Díaz

Faculty of Engineering
Universidad Panamericana

0197495@up.edu.mx

0190575@up.edu.mx

May 4, 2022

Content

Introduction: Problem

- When the classical energy/cost function $C(\vec{x})$ has a **single minimum**, any method of gradient descent can approach the minimum and optimize such Cost Function.
- However, when $C(\vec{x})$ has **multiple extrema**, an optimization technique that allows **hill-climbing** from local minima is required.
- A **Boltzmann Machine** represents one solution to this, but to speed its convergence, **Cauchy Machines** were defined by Harold Szu.

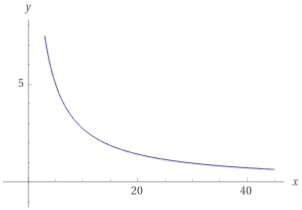
Cauchy Machines: Model

They use 5 steps for training:

1. Set an initially *high temperature* and randomize all network weights.
2. Apply an input vector and compute the objective/cost function with current weights.
3. Change each weight randomly as determined by the **Cauchy/Lorentzian Distribution**:
$$P(x) = \frac{T}{t^2 + x^2} \tag{1}$$
4. Recompute the objective/cost function: if it is reduced, make the weight change permanent (the new state of the machine); if it is increased (worse), make the change become the new state with the probability determined by Eq. (1). Otherwise, the old state is retained.
5. Compute a new value for *temperature* as follows:
$$T(n+1) = \frac{T(0)}{1+T}; n \geq 1 \tag{2}$$
6. Repeat Steps 3 though 6.

Temperature Concept

- Researchers have used simulated annealing to solve the problem of **multiple extrema**: by analogy, **networks are compared to thermodynamic systems having a temperature T**.
- When **annealing a metal**, the metal is heated to a temperature high enough so that the **atoms have sufficient energy to move about freely**, and entropy causes atoms to seek **a minimum energy configuration**.
- As we gradually reduce the temperature T, atomic movility is more limited and atoms can line themselves up in more and more perfect **minimum-energy states**.
- From the Eq. (2), we can see that the function $f(x) = \frac{temperature}{1+x}$ is the one that makes this gradual reduction. Say *temperature* $= 30$:



Cost Fuction

- The Cauchy machine extends the **discrete synchronous Hopfield model**. The Hopfield models have an *Energy (E) Function*
- It is defined as energy since it is proved that if the values of states of the neurons are chosen **randomly (using the Cauchy Distribution in this case)** for the update of the network values, the network will converge to states that are the local minima points of the Energy Function, which altogether reach the global minimum of E.
- Thus, **the energy function or cost function** to be optimized is:

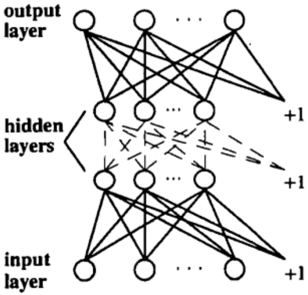
$$E(\vec{s}) = -\frac{1}{2}\sum_{i,j}^{n,n} w_{ij}s_i s_j - \sum_{i}^{n} \theta_i s_i \tag{3}$$

where s_i and s_j are the states of the *neurons* that are updated using the Cauchy Distribution mentioned. During training, the net input of a unit u_i is updated using the equation:

$$\frac{du_i(t)}{dt} = -\frac{\delta E(\vec{s})}{\delta s_i} \tag{4}$$

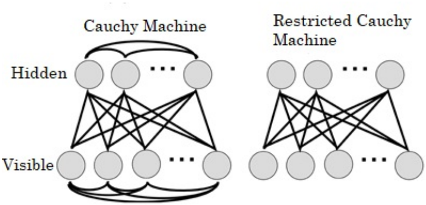
Architecture: BP ANNs'

The Backpropagation (BP) network is one of the most used ANNs' model, usually paired with a typical gradient descent technique. Thus, **it is common that a Backpropagation Neural Network Architecture uses the Cauchy Machine** training method for weight adjustment, in order to avoid the multiple minima problem. Such topology is below:



Architecture: Boltzmann

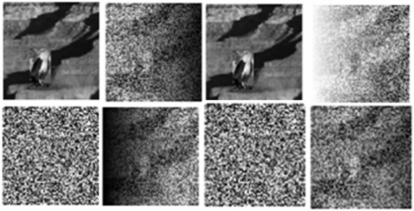
- The only difference between the statistical methods of a Boltzmann Machine and a Cauchy Machine is the distribution used in choosing the random weights' variations and the artificial temperature decreasing law, with the Cauchy converging faster.
- Thus, both machines use an architecture in which **every node is connected to every other node**. Unlike the ANNs', this architecture consists of **bidirectional connections**.
- There can be two types of nodes: **visible** nodes and **hidden** nodes, which are a similar idea to hidden and visible layers of a BP ANN. Both types of architectures described are shown below: to the left, a bidirectional Cauchy Machine NN; to the right, a BP ANN or Restricted Cauchy Machine.



Usage

The Cauchy Machine represents a possible solution to the local minima problem encountered with every other neural network training algorithm.

- Cauchy Machine NNs' are part of Associative Memory architectures, and thus are used whenever the storage of associative data is needed for predictions.
- One of the differences with the Boltzmann Machine is that a Cauchy Machine is **completely parallelizable** and is widely used for multiprocessor DL optimization techniques.
- Cauchy Machines are widely used in matrix optimization problems: therefore, **image recovery from noise mixtures** algorithms using NN often use the Cauchy method.



References

[1] CATALIN CONDURACHE and DAVID J. EVANS. "Implementation of a backpropagation-cauchy neural network training method on a parallel computer system". In: *Parallel Algorithms and Applications* 5.1-2 (1995), pp. 69–77. URL: https://doi.org/10.1080/10637199508915476.

[2] Ivica Kopriva and Harold H Szu. "Blind inversion in nonlinear space-variant imaging by using cauchy machine". In: *Independent Component Analyses, Wavelets, and Neural Networks*. Vol. 5102. SPIE. 2003, pp. 5–16.

[3] Michail G. Lagoudakis. "Neural Networks and Optimization Problems". In: *IEEE USL* (2000), pp. 2–5.

[4] Harold Szu and Ralph Hartley. "Fast simulated annealing". In: *Physics letters A* 122.3-4 (1987), pp. 157–162.

[5] Philip D Wasserman and Tom Schwartz. "Neural networks. II. What are they and why is everybody so interested in them now?" In: *IEEE expert* 3.1 (1988), pp. 10–15.