Momentum (physics) Artificial Neural Networks Machine Learning

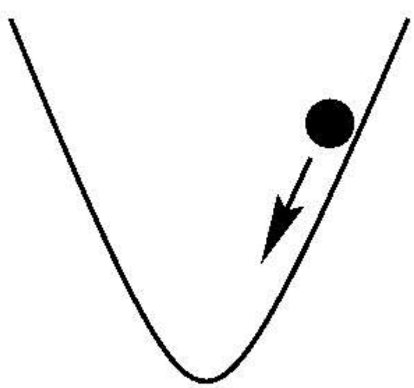
What does momentum mean in neural networks?

Answer Follow · 15 Request

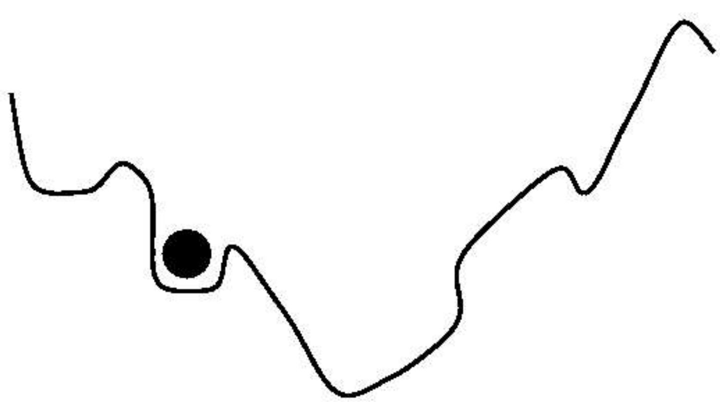
3 Answers

Shehroz Khan, ML Researcher, Postdoc @U of Toronto
Answered Sep 7, 2016

In neural networks, we use gradient descent optimization algorithm to minimize the error function to reach a global minima. In an ideal world the error function would look like this



So you are guaranteed to find the global optimum because there are no local minimum where your optimization can get stuck. However in real the error surface is more complex, may comprise of several local minima and may look like this



In this case, you can easily get stuck in a local minima and the algorithm may think you reach the global minima leading to sub-optimal results. To avoid this

Related Questions

- How momentum works in backpropagation in neural networks?
- Why does momentum help in training a neural network?
- Is momentum useful when training a neural network with batch normalization?
- What does weight mean in terms of neural networks?
- How are recurrent neural networks different from convolutional neural networks?
- What is new in deep neural networks?
- Can you train neural networks in parallel?
- How do convolutional neural networks work?
- What is Attention Mechanism in Neural Networks?
- What is a convolutional neural network?

+ Ask New Question

More Related Questions

Question Stats

14 Public Followers
29,835 Views
Last Asked Jul 22, 2016
Edits

by trying to jump from a local minima. If the momentum term is large then the learning rate should be kept smaller. A large value of momentum also means that the convergence will happen fast. But if both the momentum and learning rate are kept at large values, then you might skip the minimum with a huge step. A small value of momentum cannot reliably avoid local minima, and can also slow down the training of the system. Momentum also helps in smoothing out the variations, if the gradient keeps changing direction. A right value of momentum can be either learned by hit and tr willamette.edu ı cross-validation.

Pictures Source - Momentum and Learning Rate Adaptation

Robby Goetschalckx, 11+ years as researcher in Machine Learning.
Answered Jul 22, 2016

Let's say your first update to the weights is a vector θ_1. For the second update (which would be θ_2 without momentum) you update by $\theta_2 + \alpha\theta_1$. For the next one, you update by $\theta_3 + \alpha\theta_2 + \alpha^2\theta_1$, and so on. Here the parameter $0 \leq \alpha < 1$ indicates the amount of momentum we want.

The practical way of doing that is keeping an update vector v_i, and updating it as $v_{i+1} = \alpha v_i + \theta_{i+1}$.

The reason we do this is to avoid the algorithm getting stuck in a local minimum. Think of it as a marble rolling around on a curved surface. We want to get to the lowest point. The marble having momentum will allow it to avoid a lot of small dips and make it more likely to find a better local solution.

Having momentum too high means you will be more likely to overshoot (the marble goes through the local minimum but the momentum carries it back upwards for a bit). This will lead to longer learning times. Finding the correct value of the momentum will depend on the particular problem: the smoothness of the function, how many local minima you expect, how “deep” the sub-optimal local minima are expected to be, etc.

Brando Miranda, member of MIT CBMM (Center for Brains Minds and Machines)

Answered Jul 22, 2016

Its an optimization algorithm used to train the NN: