



Search Quora



Add Question or Link

Momentum (physics) Artificial Neural Networks Machine Learning

What does momentum mean in neural networks?

[Answer](#) [Follow · 15](#) [Request](#)

Ad by DigitalOcean

Build, test and deploy something new. Free for 60 days.

Try our optimized plans with dedicated hyper-threads on best-in-class CPUs, now with more RAM and SSD.

[Learn more at try.digitalocean.com](https://try.digitalocean.com)

3 Answers

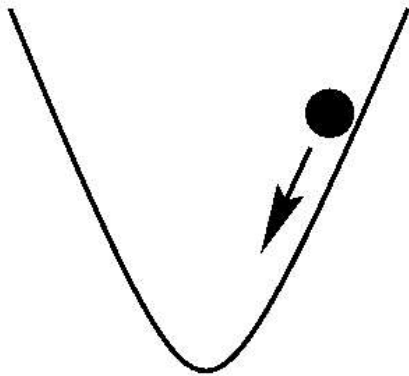


Shehroz Khan, ML Researcher, Postdoc @U of Toronto

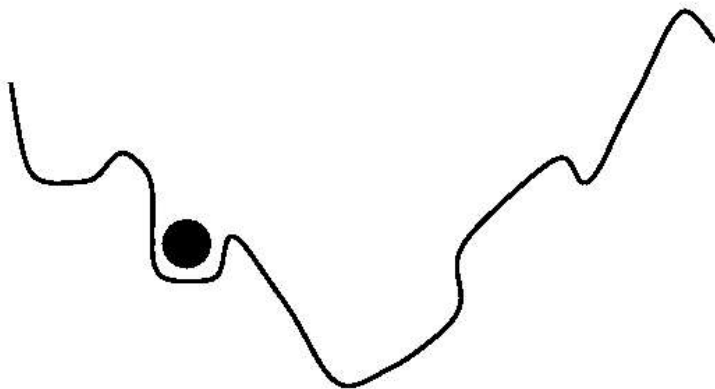
Answered Sep 7, 2016



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

Home¹

Answer



Spaces



Notifications

Search Quora



Add Question or Link

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 trial or by cross-validation.

Pictures Source - [Momentum and Learning Rate Adaptation](#)

26.2k Views · View Upvoters

Your feedback is private.

Is this answer still relevant and up to date?

Upvote · 91

Share



Recommended All

Sponsored by Elastic

Elasticsearch Service: Deploy hosted Elasticsearch and Kibana.

Deploy with the latest versions, powerful features, and optimized templates for your use case.

Free trial at elastic.co

...



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.

15.8k Views · View Upvoters · Not for Reproduction

Your feedback is private.

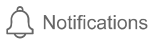
Is this answer still relevant and up to date?

Home¹

Answer



Spaces



Notifications

Search Quora



Add Question or Link



Add a comment...

Comment

Recommended All

Sponsored by Dataquest

We teach data science using real-world Python and R projects.

Start learning data science for free by building projects in your browser using real datasets.

Start now at dataquest.io

...



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:

11.1k Views · View Upvoters

Your feedback is private.

Is this answer still relevant and up to date?

Yes

No

Upvote · 10

Share



Add a comment...

Comment

Recommended All

Top Stories from Your Feed

Answer written · Python · 21h

In layman's terms, how do programming languages actually work? For example, in Python, how does the computer know what "print" means in the first place?



Matthew Lai, Research Engineer at Google DeepMind (2015-present)

Answered 21h ago

Answer written · Computer Programming · 6h

Self Taught programmers, how many projects did you do before feeling good at programming and then selling your skills?



Vohid Karimjonov, Amateur writer of Quora

Answered 6h ago

Two projects, both are calculators. After having learned the basics around front-end web


Popular on Quora · 6h


I saw my neighbor sunbathing without any clothes on in her backyard. I have young children in my house. Should I call the police for indecent exposure in public?





William Farrell, former Retired Deputy Sheriff at Los Angeles County Sheriff's Department (1985-2016)


Answered 6h ago

 Home¹

 Answer

 Spaces

 Notifications



Add Question or Link