|  |  |
| --- | --- |
| http://sphotos.ak.fbcdn.net/photos-ak-snc1/v2219/119/28/51288393663/n51288393663_1327303_6197.jpg | old_logo_color |

ASSIGNMENT Week 2-3

Improving Deep Neural Networks Hyperparameters

By

Ahmed Usama Khalifa

Submitted to

Dr. Omar Nasr

Electronics and Electrical Communications Engineering Department

FACULTY OF ENGINEERING, CAIRO UNIVERSITY

GIZA, EGYPT

### Optimization algorithms

**Mini**-**batch**

### **Mini**-**batch gradient descent** is a variation of the **gradient descent** algorithm that splits the training dataset into small **batches** that are used to calculate model error and update model coefficients. ... It is the most common implementation of **gradient descent** used in the field of deep learning.

Benefits

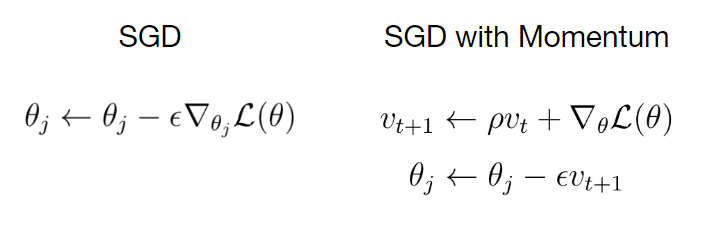
* Easily fits in the memory.
* It is computationally efficient.
* Benefit from vectorization.
* If stuck in local minimums, some noisy steps can lead the way out of them.
* Average of the training samples produces stable error gradients and convergence.

Exponentially weighted averages

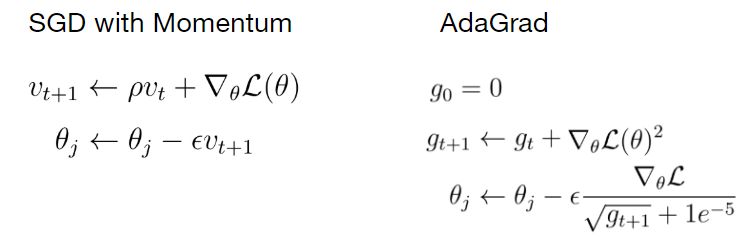
Exponentially Weighted Average for Deep Neural Networks. This algorithm is one of the most important algorithms currently in usage. From financial time series, signal processing to neural networks, it is being used quite extensively. Basically, any data that is in a sequence.



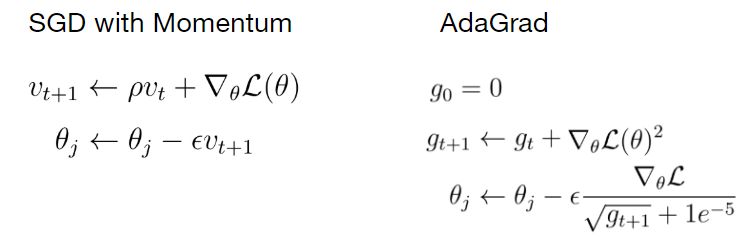
Stochastic Gradient Descent with Momentum



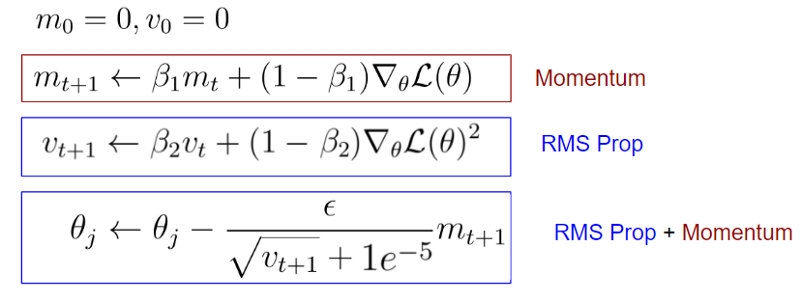
AdaGrad



RMSProp



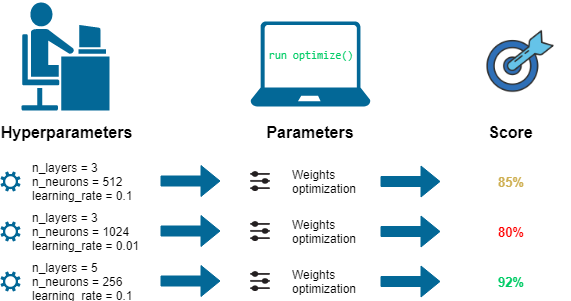
Adam Optimizer



### Hyperparameter tuning

**Hyperparameters in Machine** /**Deep Learning**. ... Model **Hyperparameters** are instead properties that govern the entire training process. They include variables which determines the network structure (for example, Number of Hidden Units) and the variables which determine how the network is trained (for example, **Learning** Rate).

What is **the importance** of **hyperparameter tuning**? **Hyperparameters** are crucial as they control the overall behaviour of a machine learning model. The ultimate goal is to find an optimal combination of **hyperparameters** that minimizes a predefined loss function to give better results.



### Batch Normalization

### As the learning progresses, the distribution of layer inputs changes due to parameter updates. This can result in most inputs being in nonlinear regime of the activation function and slow down learning. Batch normalization is a technique to reduce this effect.

### Multi-class classification

