Epoch vs Batch Size vs Iterations

Know your code...

SAGAR SHARMA   Sep 23, 2017 · 5 min read

You must have had those times when you were looking at the screen and scratching your head wondering "Why I am typing these three terms in my code and what is the difference between them " because they all look so similar.

To find out the difference between these terms you need to know some of the machine learning terms like Gradient Descent to help you better understand.

Here is a short summary on *Gradient Descent*…

Gradient Descent

It is an **iterative** optimization algorithm used in machine learning to find the best results (minima of a curve).

Gradient means the *rate* of inclination or declination of a slope.

Descent means the instance of *descending*.

The algorithm is **iterative** means that we need to get the results multiple times to get the most optimal result. The iterative quality of the gradient descent helps a under-fitted graph to make the graph fit optimally to the data.
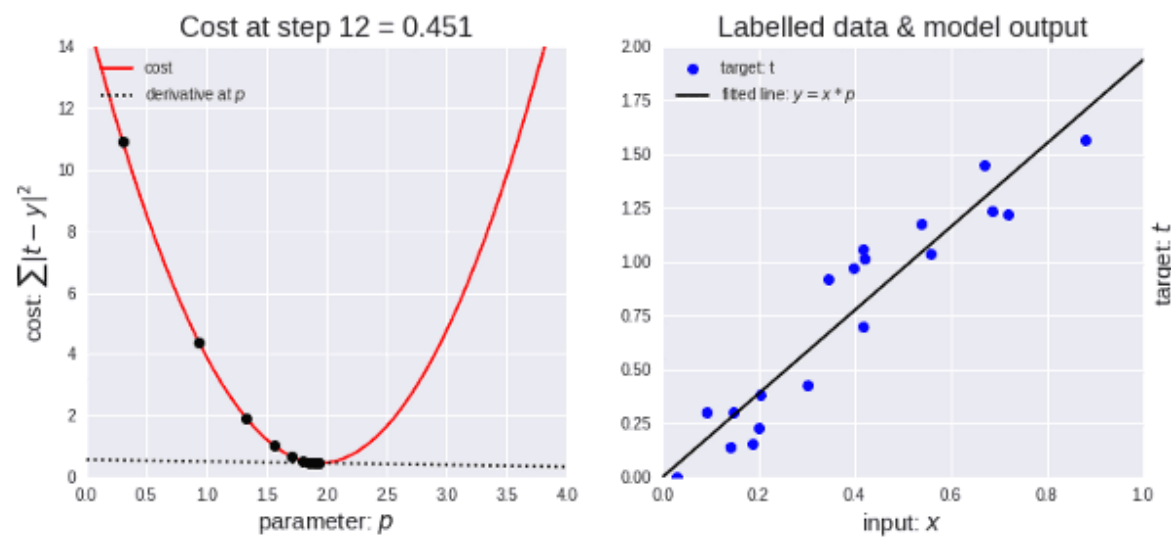
The Gradient descent has a parameter called **learning rate.** As you can see above (left), initially the steps are bigger that means the learning rate is higher and as the point goes down the learning rate becomes more smaller by the shorter size of steps. Also,the **Cost** Function is decreasing or the cost is decreasing .Sometimes you might see people saying that the **Loss** Function is decreasing or the loss is decreasing, *both **Cost** and **Loss** represent same thing* (btw it is a good thing that our loss/cost is decreasing).

We need terminologies like epochs, batch size, iterations only when the data is too big which happens all the time in machine learning and we can't pass all the data to the computer at once. So, to overcome this problem we need to divide the data into smaller sizes and give it to our computer one by one and update the weights of the neural networks at the end of every step to fit it to the data given.
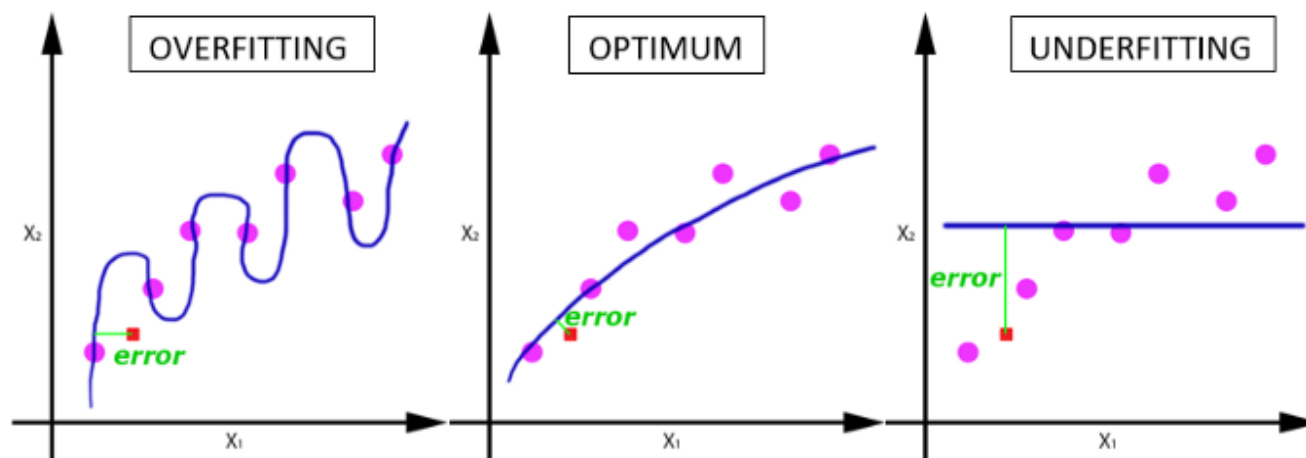
Epochs

One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE.

Since one epoch is too big to feed to the computer at once we divide it in several smaller batches.

Why we use more than one Epoch?

I know it doesn't make sense in the starting that — passing the entire dataset through a neural network is not enough. And we need to pass the full dataset multiple times to the same neural network. But keep in mind that we are using a limited dataset and to optimise the learning and the graph we are using **Gradient Descent** which is an *iterative* process. So, *updating the weights with single pass or one epoch is not enough.*

> *One epoch leads to underfitting of the curve in the graph (below).*

As the number of epochs increases, more number of times the weight are changed in the neural network and the curve goes from **underfitting** to **optimal** to **overfitting** curve.

So, what is the right numbers of epochs?

Unfortunately, there is no right answer to this question. The answer is different for different datasets but you can say that the numbers of epochs is related to how diverse your data is… just an example - Do you have only black cats in your dataset or is it much more diverse dataset?

Batch Size

> Total number of training examples present in a single batch.

> *Note: Batch size and number of batches are two different things.*

But What is a Batch?

As I said, you can't pass the entire dataset into the neural net at once. So, you **divide dataset into Number of Batches or sets or parts.**

Just like you divide a big article into multiple sets/batches/parts like Introduction, Gradient descent, Epoch, Batch size and Iterations which makes it easy to read the entire article for the reader and understand it. 😄

Iterations

To get the iterations you just need to know multiplication tables or have a calculator. 😃

Iterations is the number of batches needed to complete one epoch.

> **Note:** *The number of batches is equal to number of iterations for one epoch.*

Let's say we have 2000 training examples that we are going to use .

We can divide the dataset of 2000 examples into batches of 500 then it will take 4 iterations to complete 1 epoch.

Where Batch Size is 500 and Iterations is 4, for 1 complete epoch.
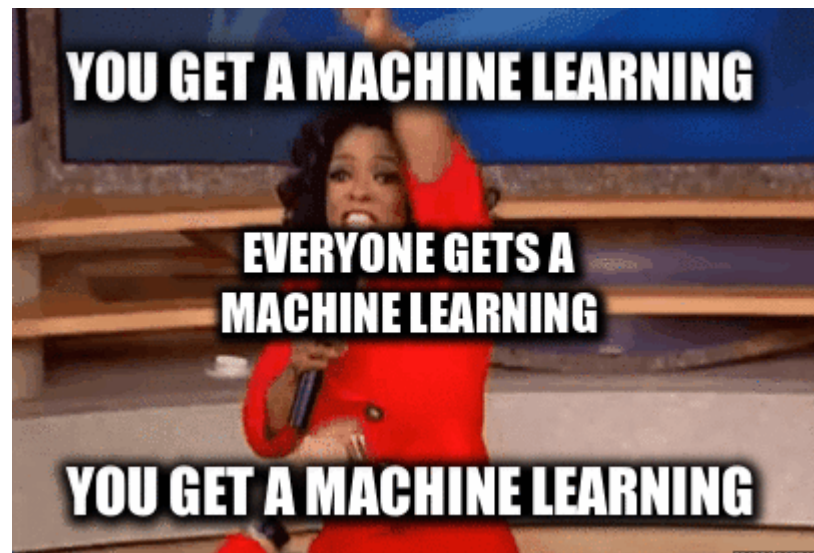
Follow me on Medium to get similar posts.

Contact me on Facebook, Twitter, LinkedIn, Google+

Any comments or if you have any question, **write it in the comment.**

Clap it! Share it! Follow Me !

Happy to be helpful. kudos…..

Previous stories you will love:

Monte Carlo Tree Search

MCTS For Every Data Science Enthusiast
towardsdatascience.com

Policy Networks vs Value Networks in Reinforcement Learning

In Reinforcement Learning, the agents take random decisions in their environment and learns on selecting the right one...

towardsdatascience.com

Activation Functions in Neural Networks and Its Types - theffork

It is a curve (sigmoid, tanH, ReLU) which is used to map the values of the network between bounded values. This is done...

theffork.com

How to Send Emails using Python

Design Professional Mails using Flask!

medium.com

TensorFlow Image Recognition Python API Tutorial

On CPU with Inception-v3(In seconds)

towardsdatascience.com

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. Take a look.

Get this newsletter

Machine Learning   TensorFlow   Data Science   Programming   Tutorial