

[Open in app](#)

Search Medium



Epoch vs Batch Size vs Iterations

Know your code...

SAGAR SHARMA · [Follow](#)Published in [Towards Data Science](#)

5 min read · Sep 23, 2017

[Listen](#)[Share](#)[More](#)

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.

Epoch vs Batch Size vs Iterations

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*...

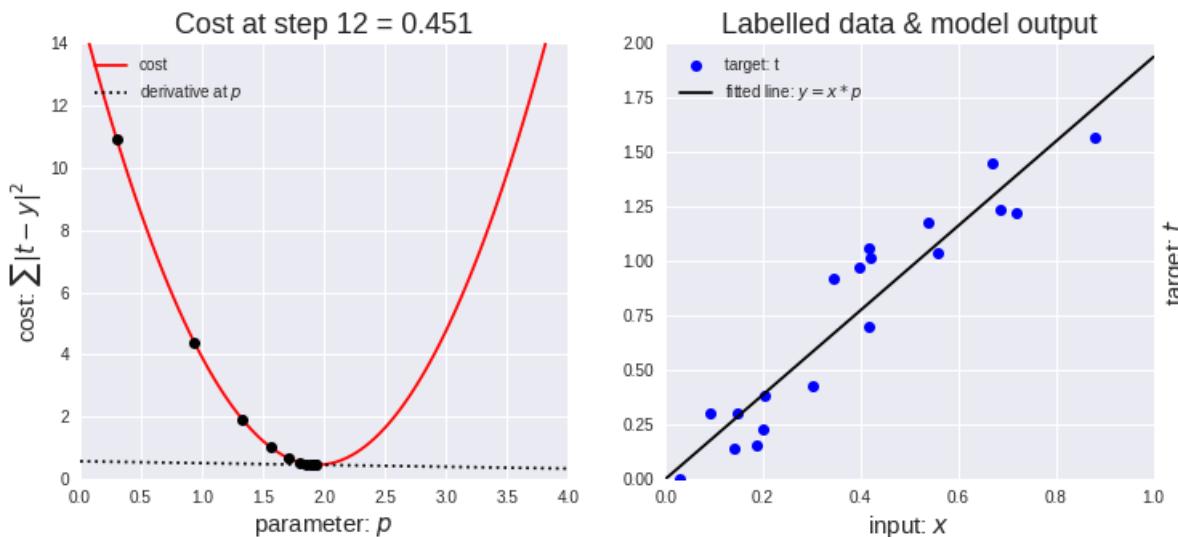
Gradient Descent

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.



[Source](#)

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

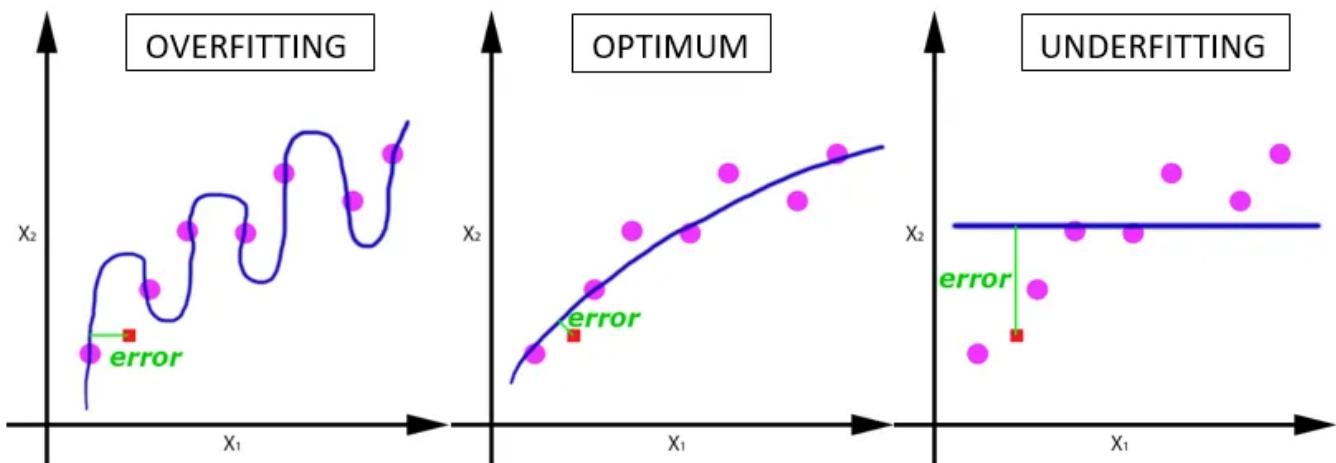
One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE.

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.



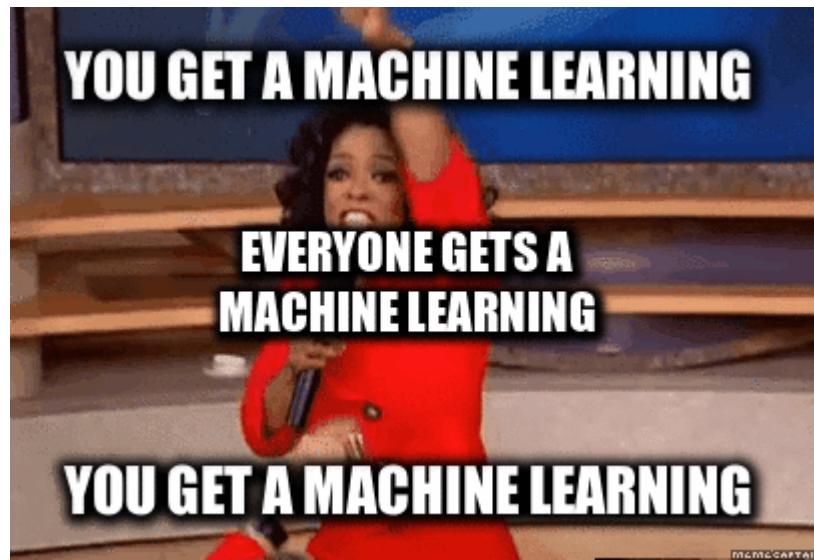
Iterations is the number of batches needed to complete one epoch.

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](https://towardsdatascience.com/mcts-for-every-data-science-enthusiast-101f3a2a2a1)

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](https://towardsdatascience.com/policy-networks-vs-value-networks-in-reinforcement-learning-101f3a2a2a1)

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](https://medium.com/@mohitgupta1993/design-professional-mails-using-flask-101f3a2a2a1)

TensorFlow Image Recognition Python API Tutorial

On CPU with Inception-v3(in seconds)

[towardsdatascience.com](https://towardsdatascience.com/tensorflow-image-recognition-python-api-tutorial-101f3a2a2a1)

Machine Learning

TensorFlow

Data Science

Programming

Tutorial

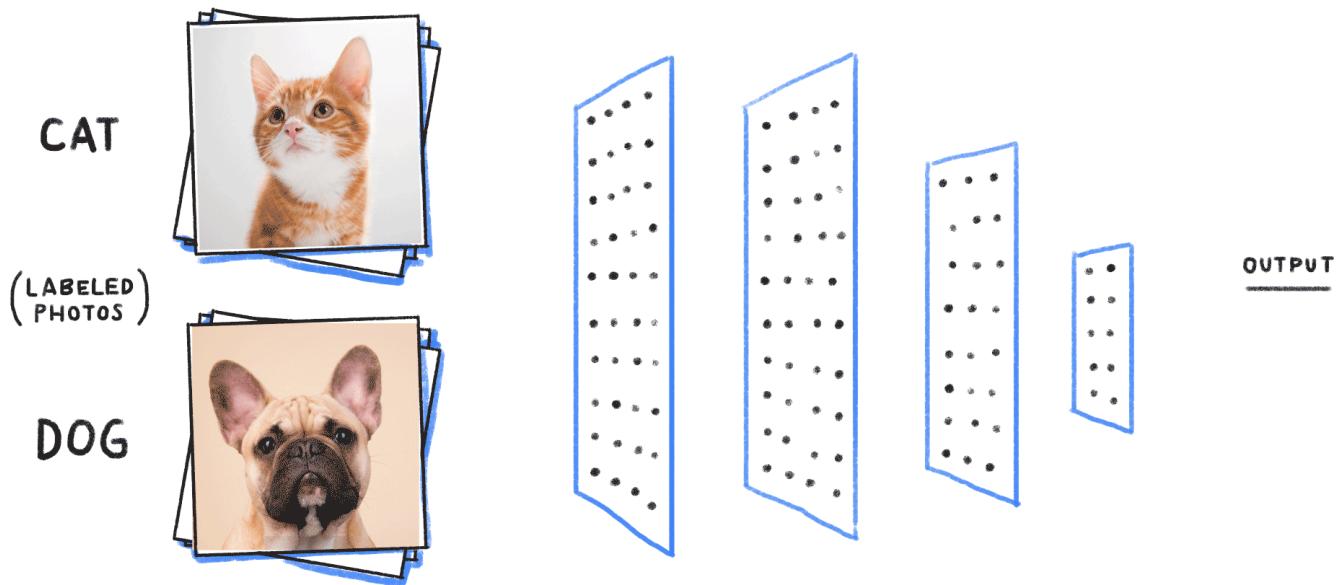
[Follow](#)

Written by SAGAR SHARMA

4K Followers · Writer for Towards Data Science

Freelance Writer. React developer. Deep learning/AI Electronics sagarsharma4244@gmail.com

More from SAGAR SHARMA and Towards Data Science



SAGAR SHARMA in Towards Data Science

What the Hell is Perceptron?

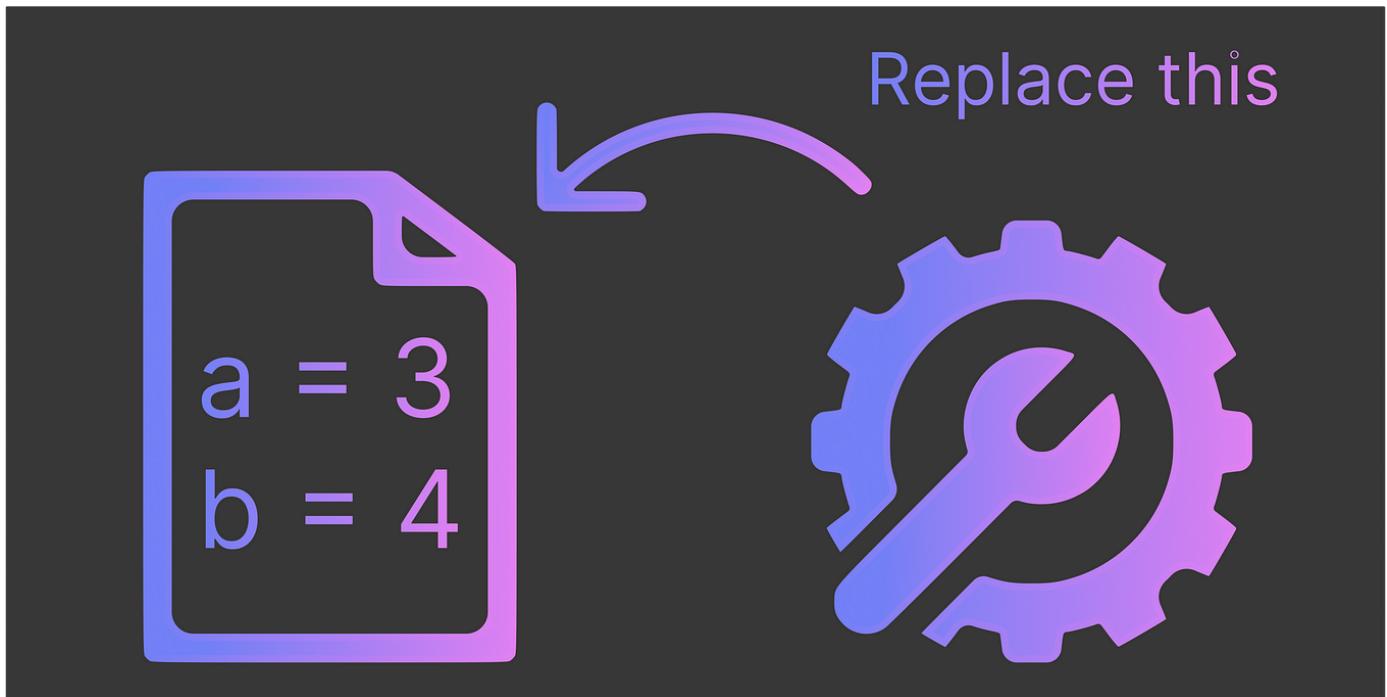
The Fundamentals of Neural Networks

3 min read · Sep 9, 2017

3.1K

8

...



Khuyen Tran in Towards Data Science

Stop Hard Coding in a Data Science Project—Use Config Files Instead

And How to Efficiently Interact with Config Files in Python

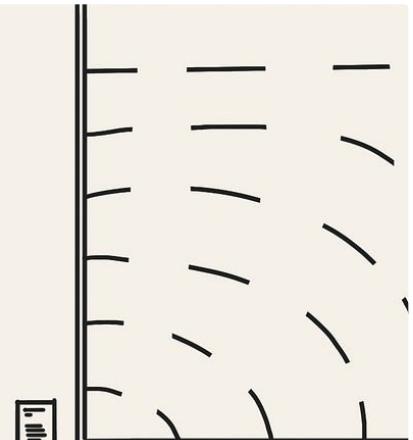
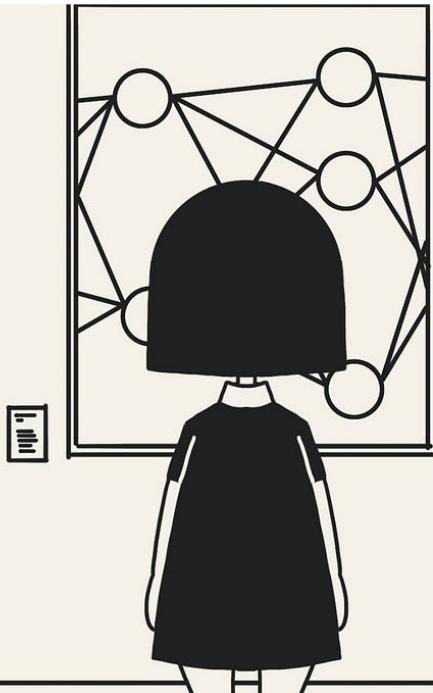
· 6 min read · May 26

1.5K

19

...

```
000100010001000100010001111  
001110011110000011111010000  
01111000101101110011110000/  
101001010001000100010001000  
1101010101110111010101010  
1010110101010111101110001  
011010101010001000001000100  
010002000001000100010001000  
100011110011101111000001111  
101000001111100010110111001  
1110000/10100101000100010001  
0001000110101011101101111  
010101010101010101010101110  
111000101101010101000100000  
100010010002000001000100010  
00100010001110011101111000.
```



Leonie Monigatti in Towards Data Science

10 Exciting Project Ideas Using Large Language Models (LLMs) for Your Portfolio

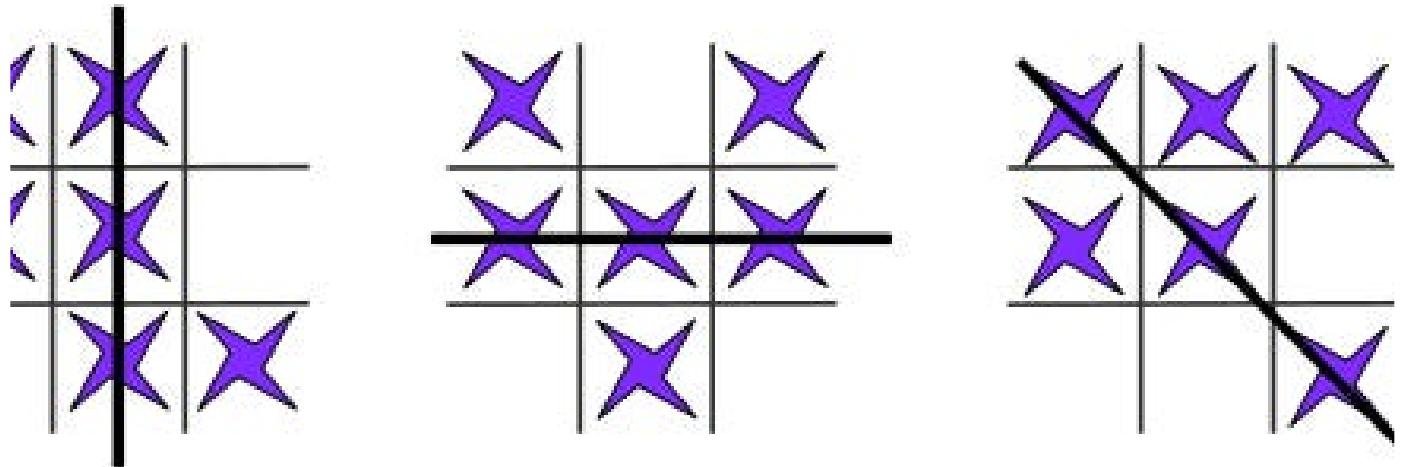
Learn how to build apps and showcase your skills with large language models (LLMs). Get started today!

· 11 min read · May 15

1.4K 9



...



 SAGAR SHARMA in Towards Data Science

Monte Carlo Tree Search

MCTS For Every Data Science Enthusiast

6 min read · Aug 1, 2018

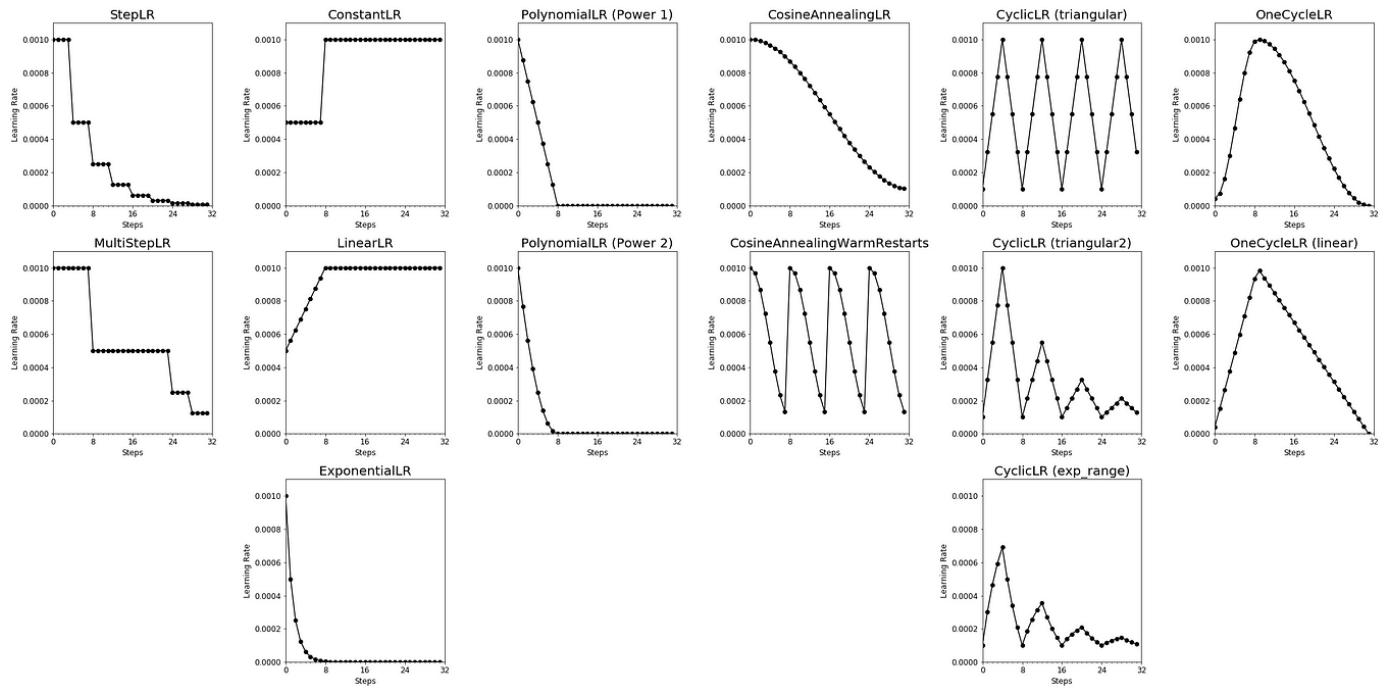
 851  3

See all from SAGAR SHARMA

See all from Towards Data Science

Recommended from Medium



Leonie Monigatti in Towards Data Science

A Visual Guide to Learning Rate Schedulers in PyTorch

LR decay and annealing strategies for Deep Learning in Python

★ · 9 min read · Dec 6, 2022

1K 6





Clément Delteil in Towards AI

Mastering the Fundamentals: Differences Between Sample, Batch, Iteration, And Epoch

A Beginner's Guide to Mastering the Fundamentals of Machine Learning

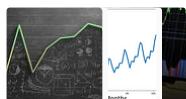
◆ · 6 min read · Feb 3

51



...

Lists



Predictive Modeling w/ Python

18 stories · 7 saves



Practical Guides to Machine Learning

10 stories · 19 saves



General Coding Knowledge

20 stories · 6 saves



Natural Language Processing

341 stories · 5 saves



Rukshan Pramoditha in Data Science 365

Determining the Right Batch Size for a Neural Network to Get Better and Faster Results

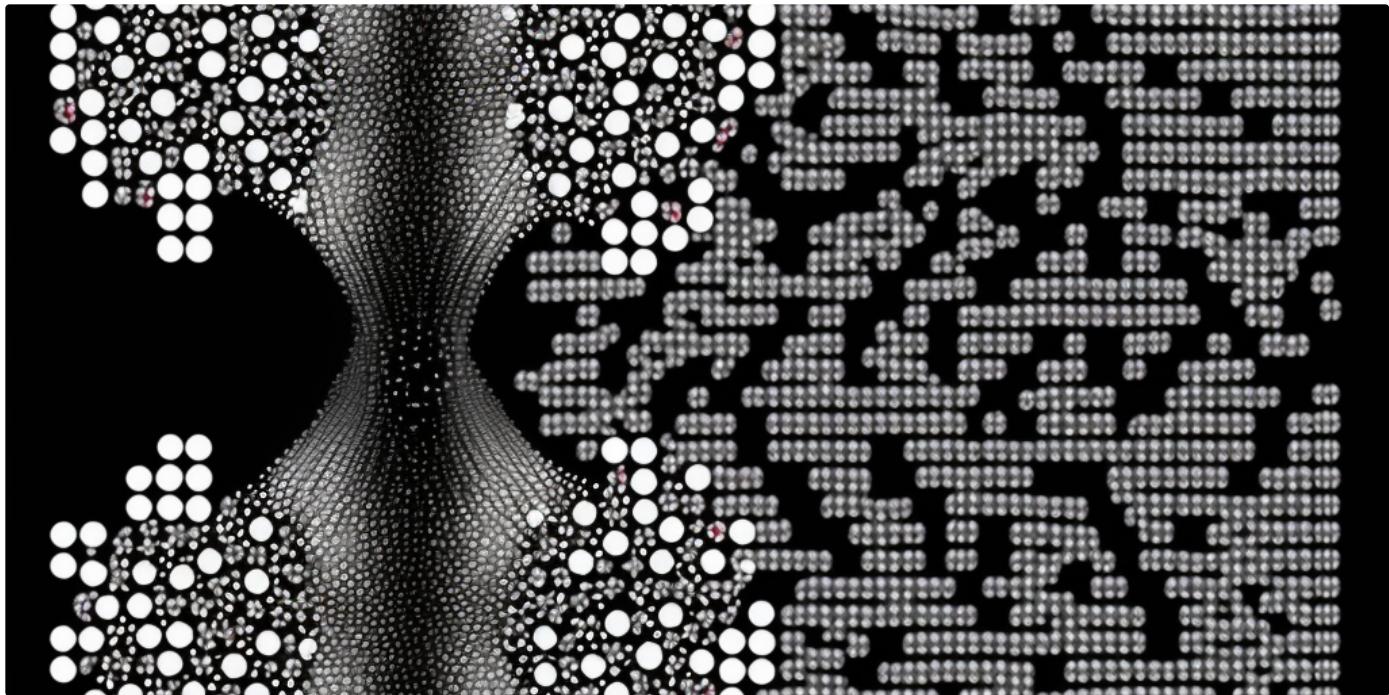
Guidelines for choosing the right batch size to maintain optimal training speed and accuracy while saving computer resources

★ · 4 min read · Sep 26, 2022

38



...



 Shreya U. Patil in Geek Culture

Loss Functions and Optimizers in ML models

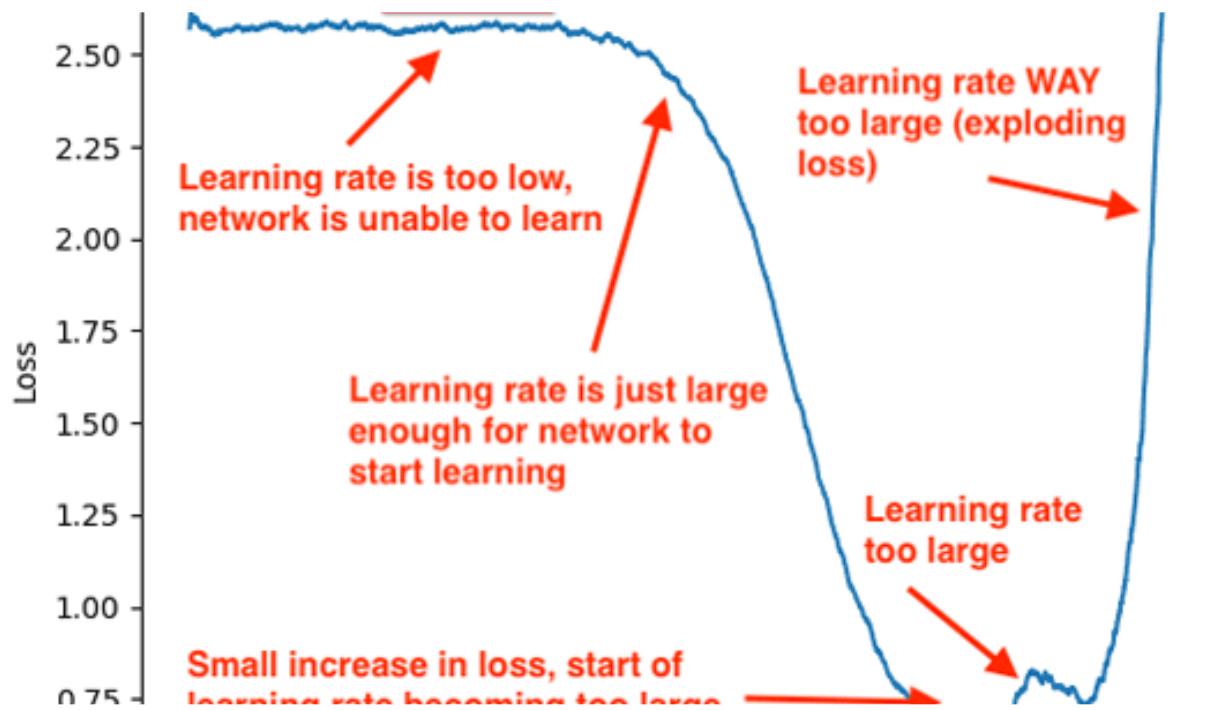
In machine learning, a loss function and an optimizer are two essential components that help to improve the performance of a model. A loss...

5 min read · Jan 30

 63



...



 Moklesur Rahman

Learning Rate Scheduler in Keras

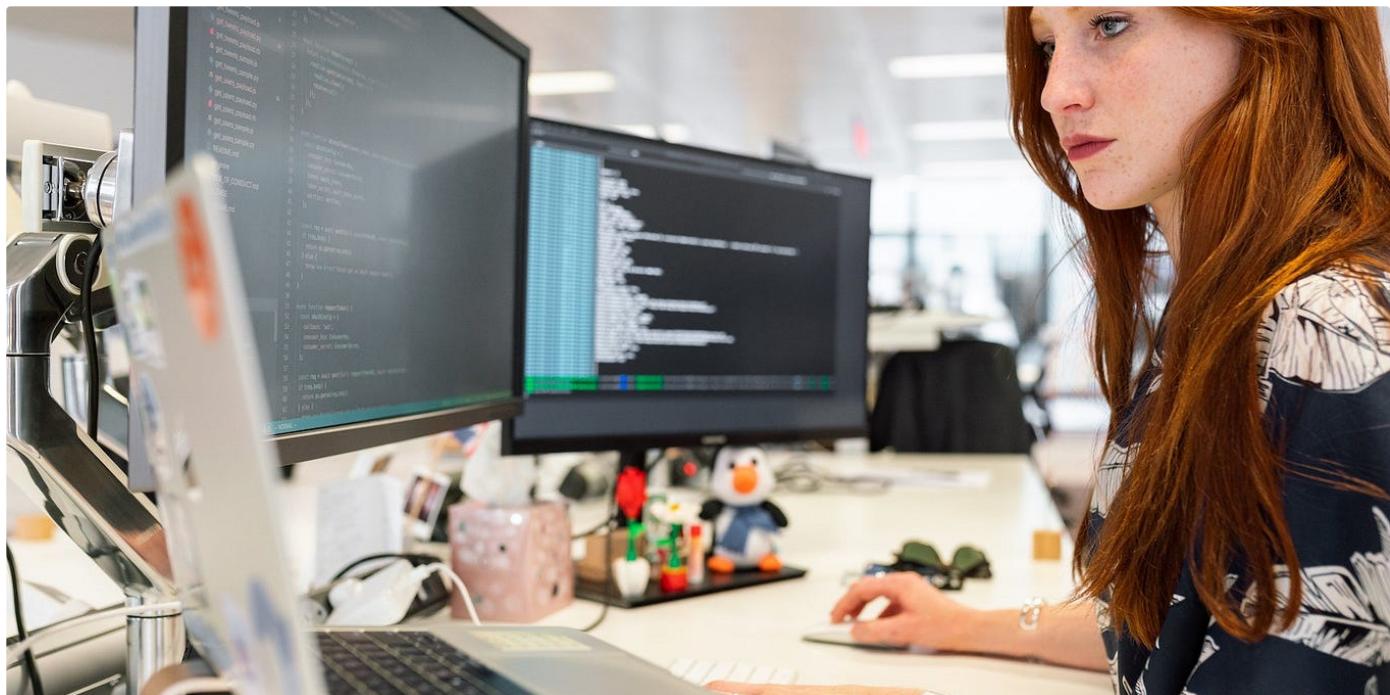
The learning rate is considered one of the most important hyperparameters for training deep learning models, but choosing it can be quite...

★ · 7 min read · Jan 23

 33



...



The Coding Diaries in The Coding Diaries

Why Experienced Programmers Fail Coding Interviews

A friend of mine recently joined a FAANG company as an engineering manager, and found themselves in the position of recruiting for...

★ · 5 min read · Nov 2, 2022

4.1K

86



...

See more recommendations