



DataCamp

Introduction to Cyclical Learning Rates for training Neural Nets

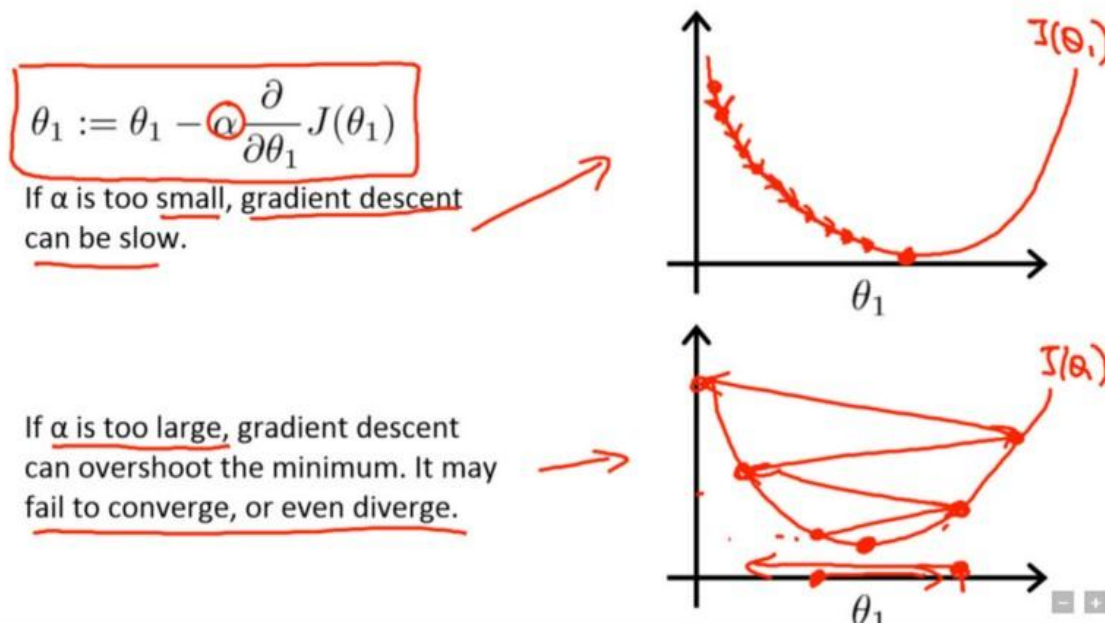
Sayak Paul

Project Instructor @ DataCamp
(GDG DevFest, Kolkata, India)

2018-11-03

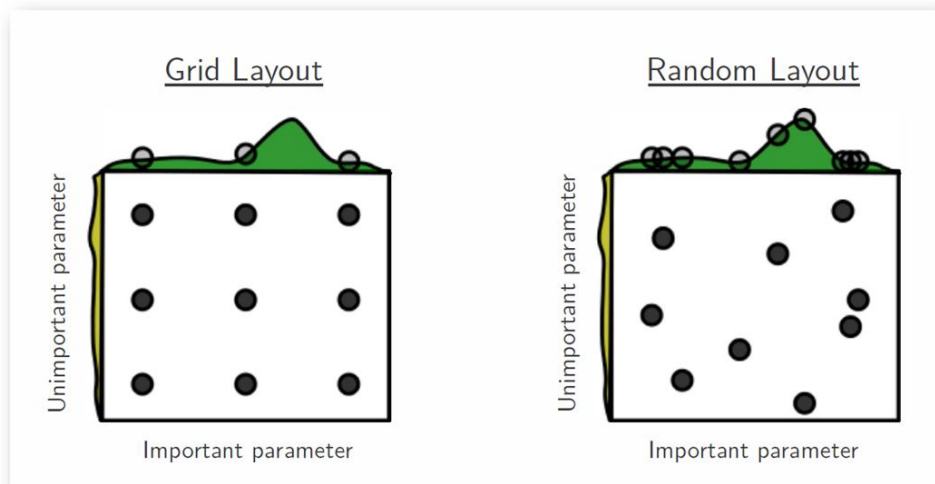
- Why are *learning rates* used?
- Some existing approaches for choosing the right learning rate
- What are the *shortcomings of these approaches*?
- Need of a systematic approach for setting the learning rate –
Cyclical Learning Rates (CLR)
- What is CLR?
- Some *amazing results* shown by CLR
- Conclusion

Learning is an important **hyperparameter** for adjusting the weights of a network with respect to the loss gradient.

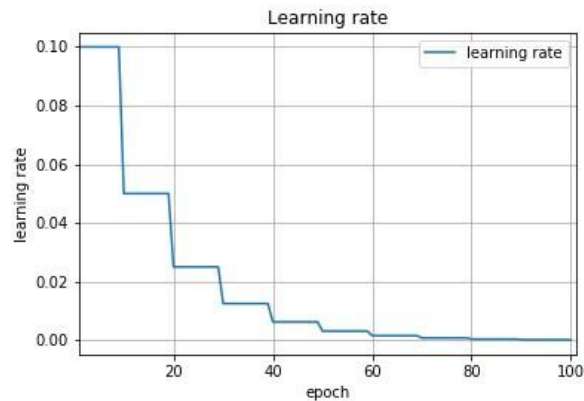


Source: Andrew Ng's lecture notes from Coursera

- Trying out different learning rates for a problem.
- **Grid-searching/Random-searching.**
- [Adaptive Learning Rates](#) / Learning Rate Schedules.



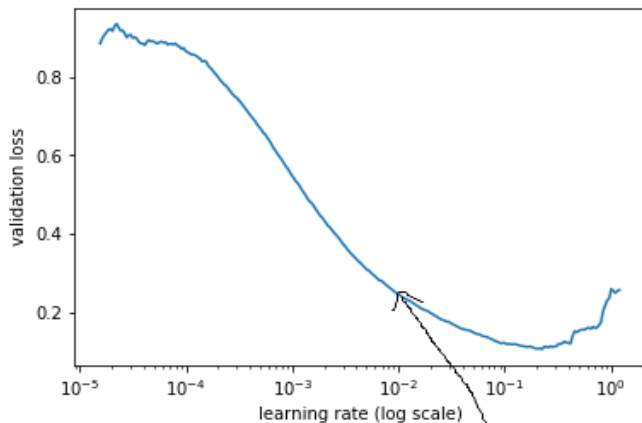
Grid and Random layout of parameters



Step Decay Schedule

- Computationally costly.
- Gives no early clue if at all the result would get better.

- Proposed by **Leslie N. Smith** in his paper entitled “[Cyclical Learning Rates for Training Neural Networks](#)” in 2015.
- The idea is to simply keep increasing the learning rate from a very small value, until the loss stops decreasing.



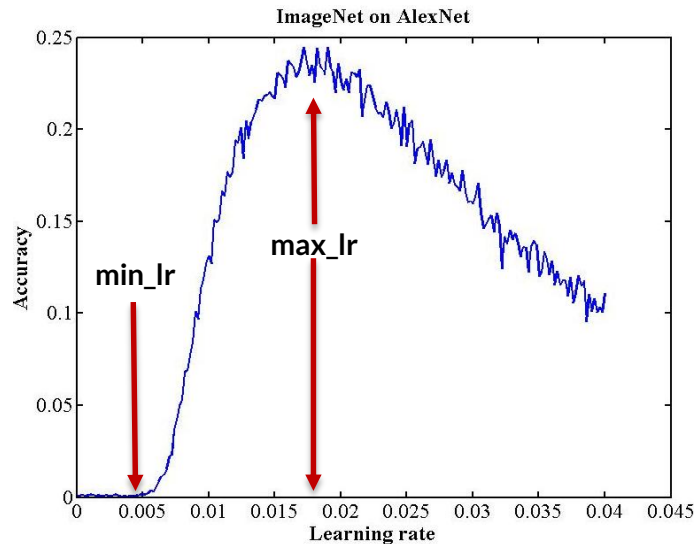
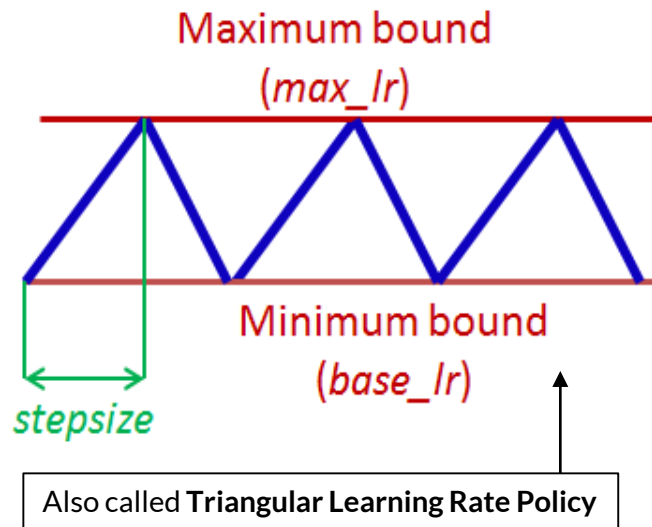
The sweet spot!

[Source](#)

* [Cyclical Learning Rates for Training Neural Networks](#) – Leslie N. Smith

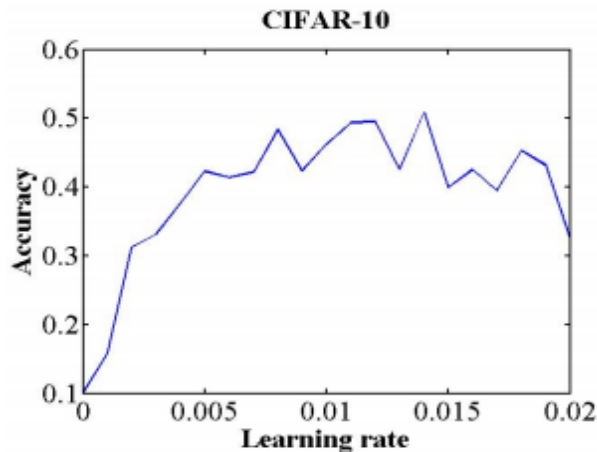
- The main idea behind CLR *is varying learning rates* between min and max values.
- ***LR_Range_Test()*** is conducted for fixing the min and max values of learning rate.

- One step of increasing learning rate.



Source: [Cyclical Learning Rates for Training Neural Networks](#) – Leslie N. Smith

- Run the model for several epochs while letting the learning rate increase linearly (use triangular learning rate policy) between low and high learning rate values.
- Next, plot the **accuracy versus learning rate** curve.
- Note the learning rate value when the accuracy starts to increase and when the accuracy slows, becomes ragged, or starts to fall. These two learning rates are good choices for defining the range of the learning rates.



Source: [Cyclical Learning Rates for Training Neural Networks](#) – Leslie N. Smith

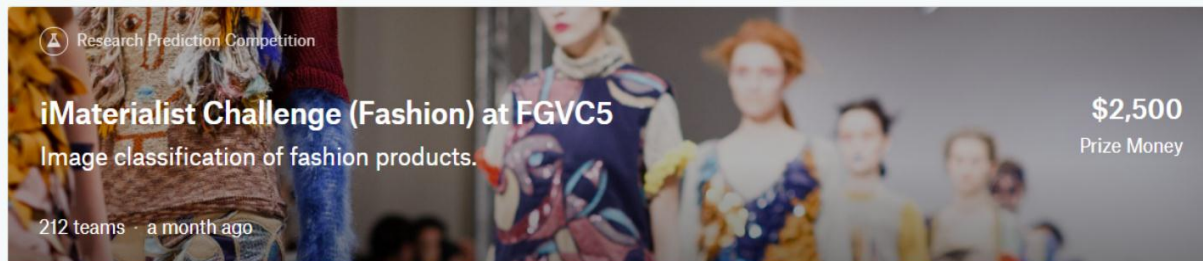



[As a Keras callback](#)

fast.ai

Making neural nets
uncool again

[As lr_find\(\) method](#)



 Research Prediction Competition

iMaterialist Challenge (Fashion) at FGVC5

Image classification of fashion products.

212 teams · a month ago


\$2,500


Prize Money

[Overview](#) [Data](#) [Kernels](#) [Discussion](#) [Leaderboard](#) [Rules](#)

[Public Leaderboard](#) [Private Leaderboard](#)

The private leaderboard is calculated with approximately 70% of the test data.
This competition has completed. This leaderboard reflects the final standings.

 Refresh

 In the money








#	Δ pub	Team Name	Kernel	Team Members	Score 	Entries	Last
1	—	radek			0.72483	64	1mo
2	—	cybercore.co.jp		   	0.71433	79	1mo
3	—	Jazielinho			0.70669	109	1mo

Image Classification on CIFAR10

Training Time

[All Submissions](#)

Objective: Time taken to train an image classification model to a test accuracy of 94% or greater on CIFAR10.

Rank	Time to 94% Accuracy	Model	Framework	Hardware
1 	0:02:54	Custom Wide Resnet <i>fast.ai + students team: Jeremy Howard, Andrew Shaw, Brett Koonce, Sylvain Gugger source</i>	fastai / pytorch	8 * V100 (AWS p3.16xlarge)
2 	0:05:41	Resnet18 + minor modifications <i>bkj source</i>	pytorch 0.3.1.post2	V100 (AWS p3.2xlarge)
3 	0:06:45	Custom Wide Resnet <i>fast.ai + students team: Jeremy Howard, Andrew Shaw, Brett Koonce, Sylvain Gugger source</i>	fastai / pytorch	Paperspace Volta (V100)

Training Cost

[All Submissions](#)

Objective: Total cost for public cloud instances to train an image classification model to a test accuracy of 94% or greater on CIFAR10.

Rank	Cost (USD)	Model	Framework	Hardware
1 	\$0.26	Custom Wide Resnet <i>fast.ai + students team: Jeremy Howard, Andrew Shaw, Brett Koonce, Sylvain Gugger source</i>	fastai / pytorch	Paperspace Volta (V100)
2 	\$0.29	Resnet18 + minor modifications <i>bkj source</i>	pytorch 0.3.1.post2	V100 (AWS p3.2xlarge)
3 	\$1.18	Custom Wide Resnet <i>fast.ai + students team: Jeremy Howard, Andrew Shaw, Brett Koonce, Sylvain Gugger source</i>	fastai / pytorch	8 * V100 (AWS p3.16xlarge)

[DAWN Bench Challenge](#) Leaderboard and Leader's specs

- Limited applicability.
- Seems to work only for **Cifar-10** and **resnets**.
- But definitely provides a more systematic way for choosing learning rate than the earlier approaches.

- Learning rate annealing (SDGR).
- Differential Learning Rates.

Cyclical Learning Rates for Training Neural Networks

Leslie N. Smith

U.S. Naval Research Laboratory, Code 5514
4555 Overlook Ave., SW., Washington, D.C. 20375

leslie.smith@nrl.navy.mil

[Original CLR Paper](#)



Sayak Paul
October 15th, 2018

NEURAL NETWORKS +1

Introduction to Cyclical Learning Rates

Learn what cyclical learning rate policy is and how it can improve the training of a neural network.

[DataCamp tutorial covering CLR](#)



sayakpaul [Merge pull request #3](#) from sayakpaul/sayakpaul-patch-3 ...

Latest commit 61fc822 9 hours ago

Introduction to Cyclical Learning Rates for training Neural Nets.pdf	Minor fixes	9 hours ago
LICENSE	Initial commit	5 days ago
README.md	Update README.md	5 days ago

[Slides available on my Github \(Username: sayakpaul\)](#)



DataCamp

Thank you!

Sayak Paul
Project Instructor
spsayakpaul@gmail.com