

Fastai Study Group



Cleveland AI Group (CAIG)
October 8, 2018 - Week 2

Event Hosts



Michael Kudlaty
@Michael Kudlaty



Jason Mancuso
@jvmancuso



Michał Wojczulis
@Michał

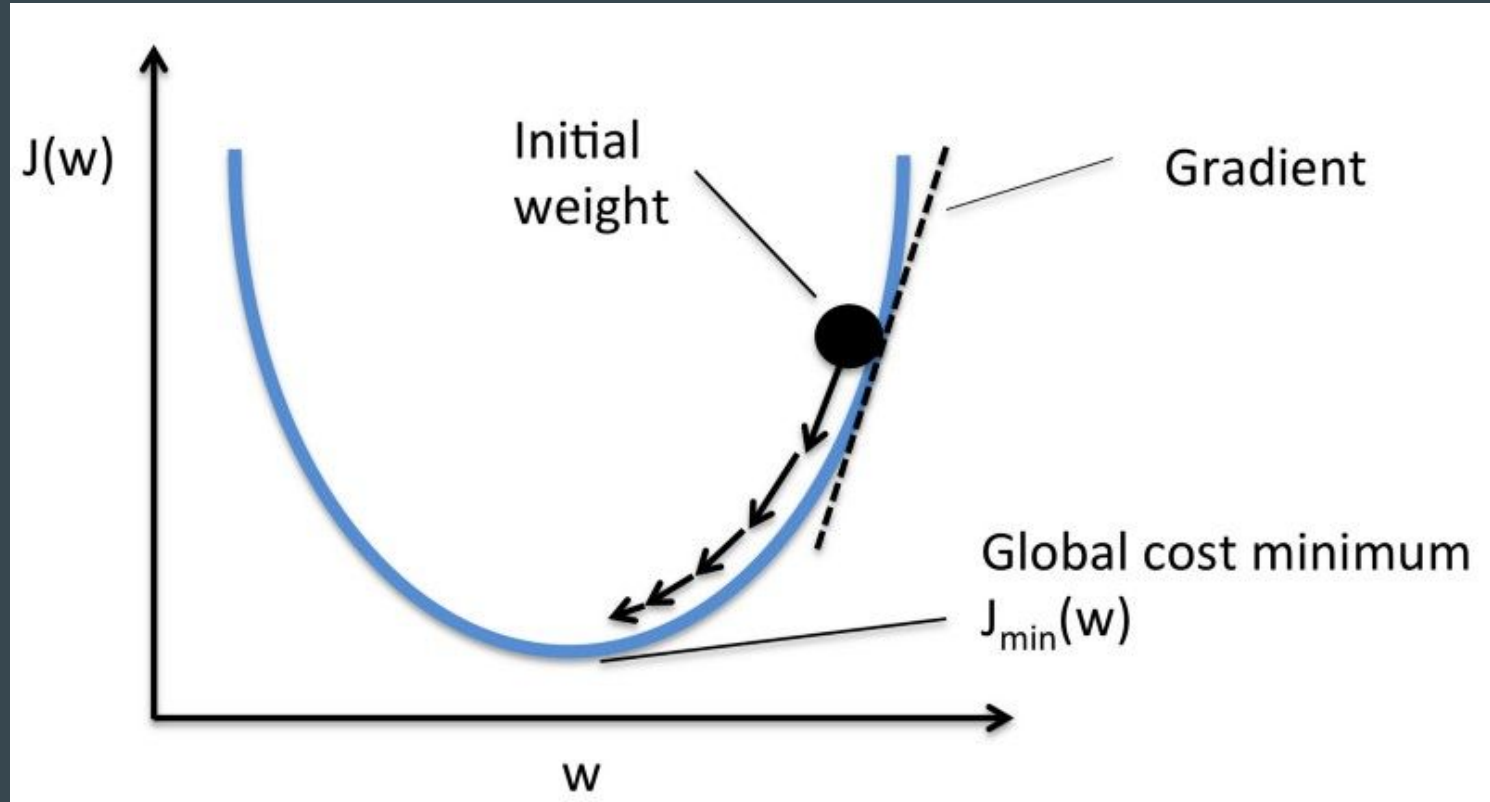


Brendan Mulcahy
@Brendan Mulcahy

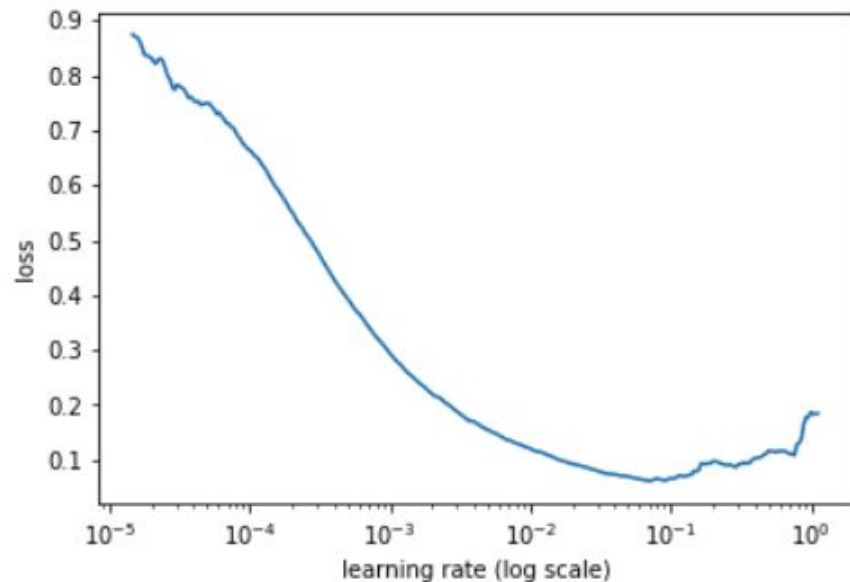
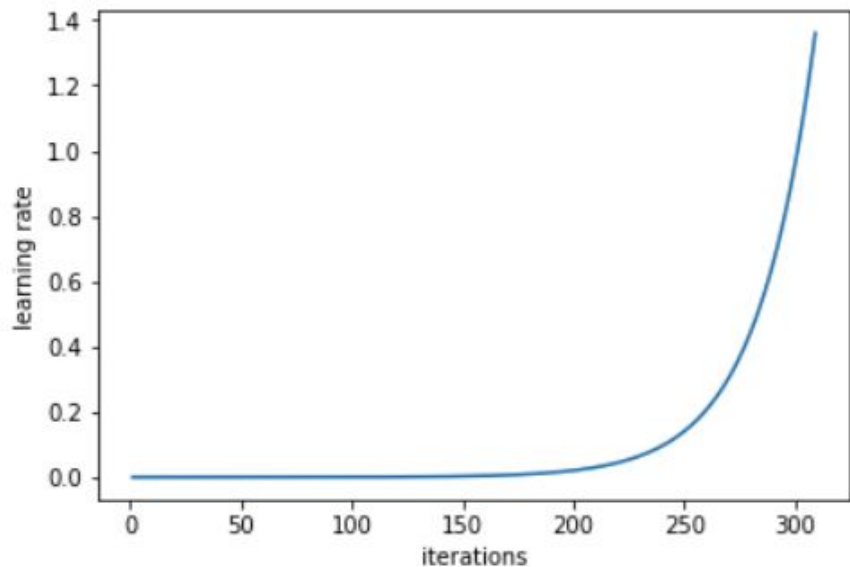
Agenda

1. Lesson 2 key concepts
2. Discussion/questions, e.g.:
 - a. Lesson 2 video content
 - b. Lesson 2 jupyter notebook
 - c. Related thoughts

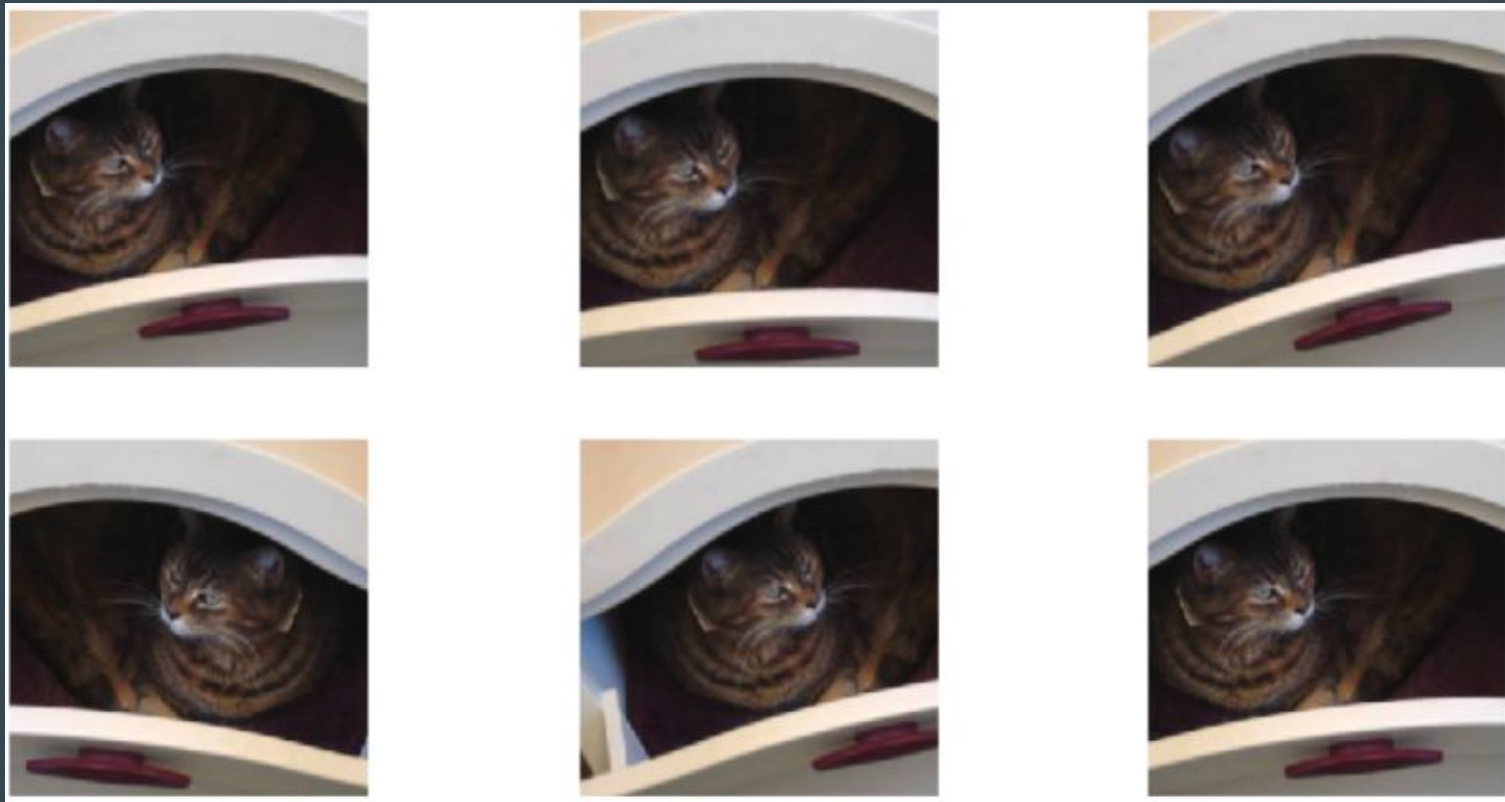
Learning rate



Learning rate finder

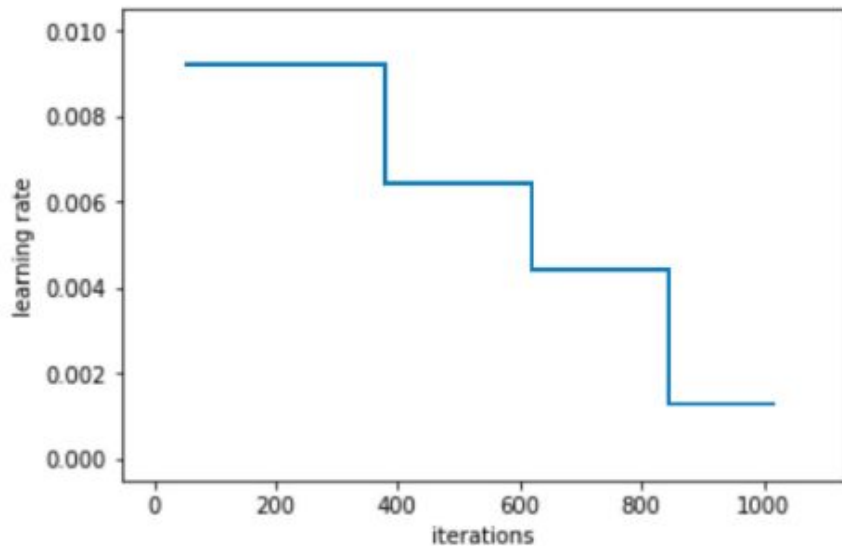


Data augmentation

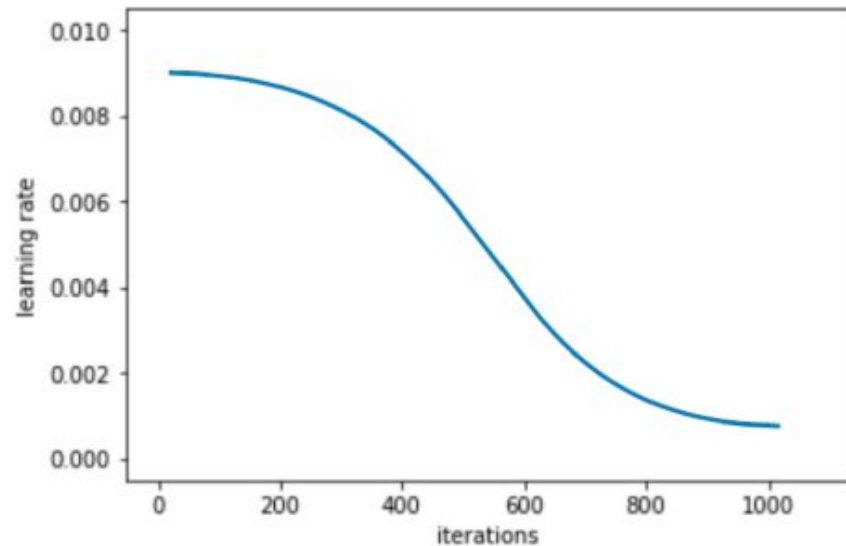


Cosine annealing

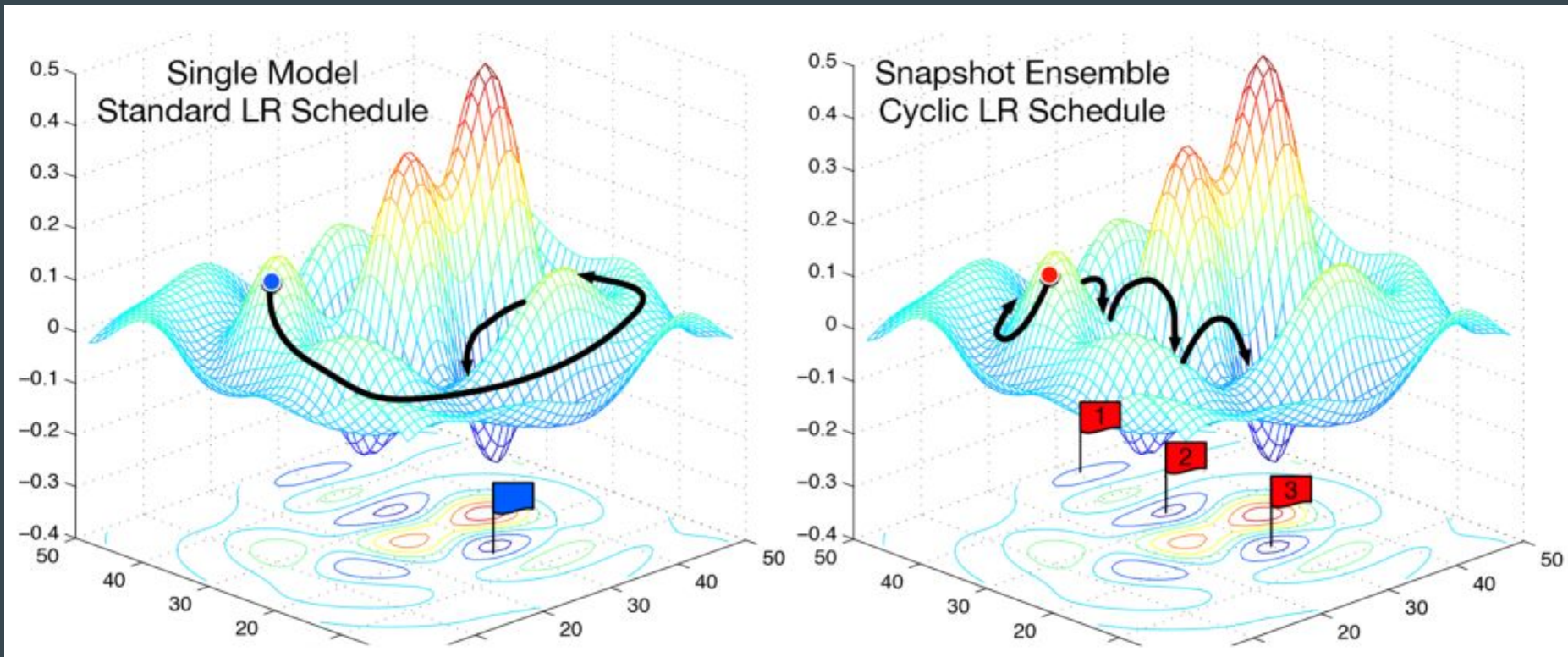
Stepwise Annealing



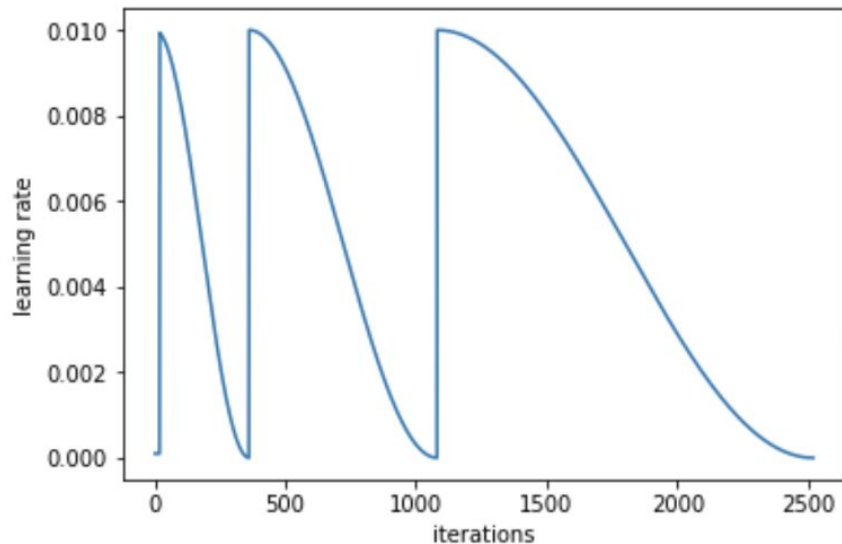
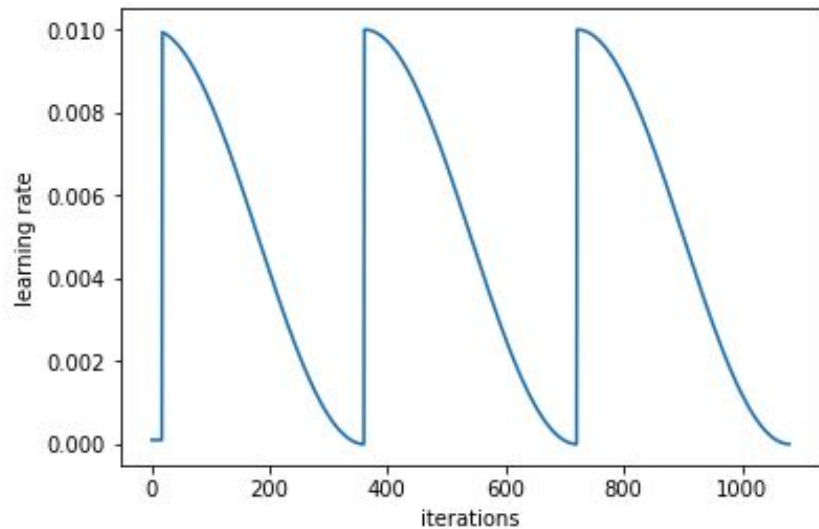
Cosine Annealing



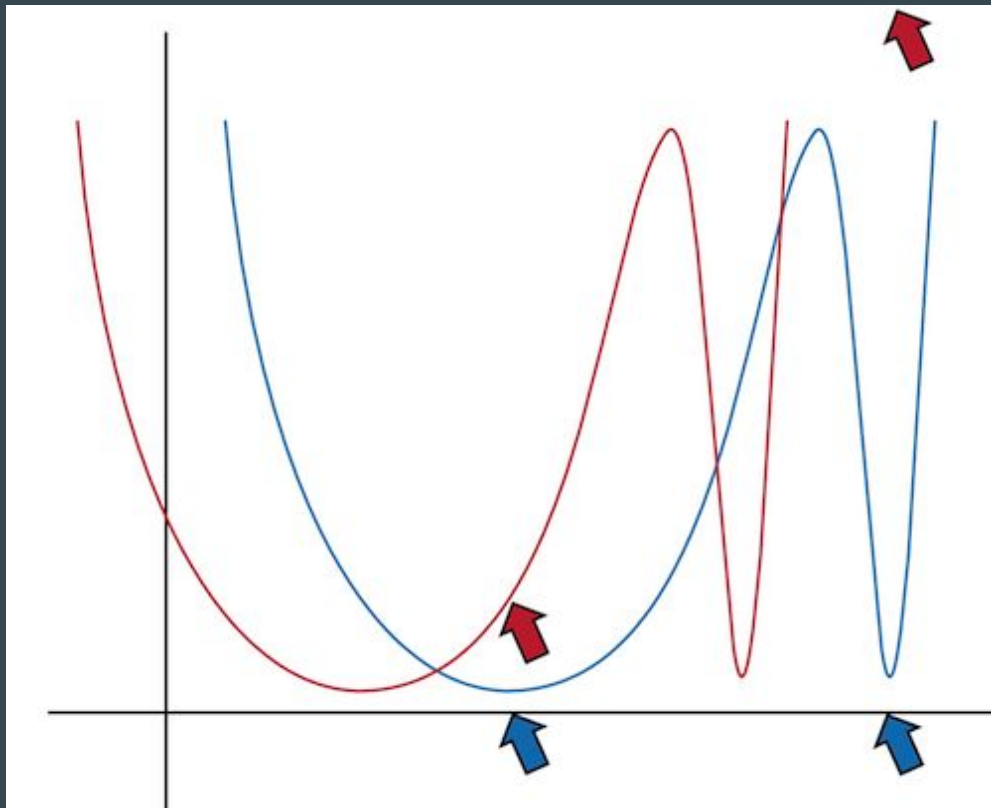
SGD with restarts



Cycle_len and cycle_mult



Spiky minima

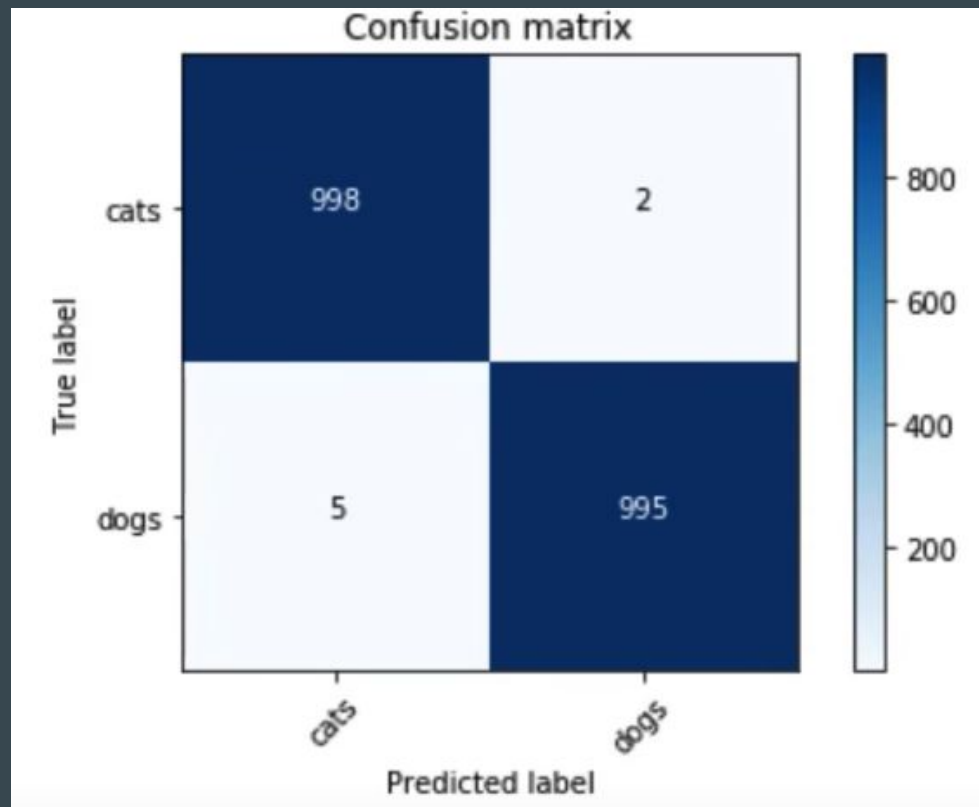


Test time augmentation

“Test Time Augmentation” means is that we are going to take 4 data augmentations at random as well as the un-augmented original (center-cropped). We will then calculate predictions for all these images, take the average, and make that our final prediction. Note that this is only for validation set and/or test set.

To do this, all you have to do is `learn.TTA()`

Confusion matrix



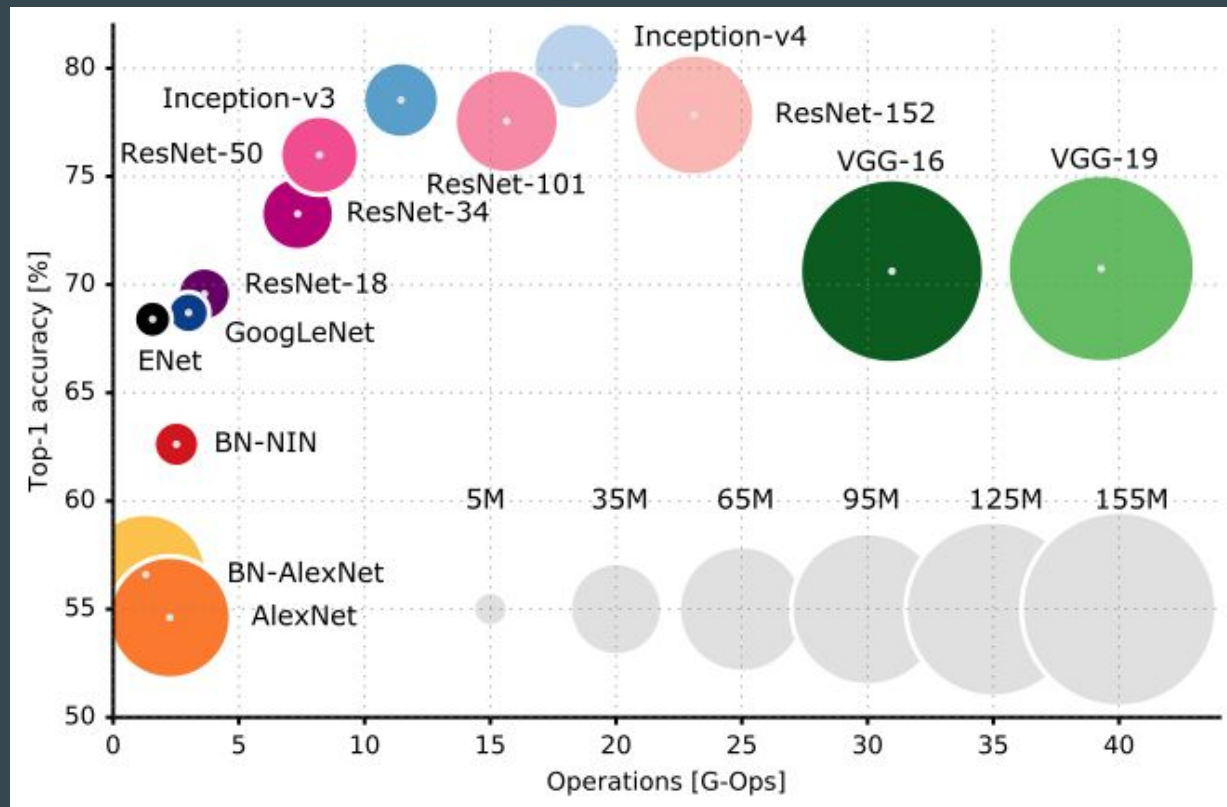
Steps to train a world-class image classifier

1. Enable data augmentation, and `precompute=True`
2. Use `lr_find()` to find highest learning rate where loss is still clearly improving
3. Train last layer from precomputed activations for 1–2 epochs
4. Train last layer with data augmentation (i.e. `precompute=False`) for 2–3 epochs with `cycle_len=1`
5. Unfreeze all layers
6. Set earlier layers to 3x-10x lower learning rate than next higher layer. Rule of thumb: 10x for ImageNet like images, 3x for satellite or medical imaging
7. Use `lr_find()` again (Note: if you call `lr_find` having set differential learning rates, what it prints out is the learning rate of the last layers.)
8. Train full network with `cycle_mult=2` until over-fitting

Neural Network Architectures

Top1 vs. operations, size \propto parameters. Top-1 one-crop accuracy versus amount of operations required for a single forward pass. The size of the blobs is proportional to the number of network parameters; a legend is reported in the bottom right corner, spanning from 5×10^6 to 155×10^6 params. Both these figures share the same y-axis, and the grey dots highlight the centre of the blobs.

<https://arxiv.org/pdf/1605.07678.pdf>



<https://towardsdatascience.com/neural-network-architectures-156e5bad51ba>

Resources

- Wiki lesson 2 <http://forums.fast.ai/t/wiki-lesson-2/9399>
- Lesson 2 notes [Link](#)
- Course page <http://course.fast.ai/>
- Course forums <http://forums.fast.ai/>
- Cleveland Tech Slack <https://cleveland-tech.herokuapp.com/>
 - Join #deep_learning channel
 - Ask questions or share articles
- AI Saturdays <https://nurture.ai/ai-saturdays>
- AI Saturdays guide [Link](#)
- AI Saturdays forums <https://ai6forums.nurture.ai/>
- CAIG Website <https://clevelandaigroup.github.io/>

Thank You to Our Sponsors!



Questions?