1-Cycle Training and Learning Rates Finder

1-cycle Training:

Leslie Smith designed a schedule for learning rates separated into two phases. One where the learning rate grows from minimum to maximum(warmup). And one where it decreases from maximum to minimum(annealing).

1-cycle training allows us to use a much higher learning rate than other types of trainings. This has two benefits.

1. By training with higher learning rates, we train faster. A phenomenon called Super Convergence.
2. By training with a higher learning rate, we overfit less because we skip over the sharp local minima to end up in a smoother (and more generalizable) part of the loss. Love this schedule for Maine inspired and two from one where learning.

Learning Rate Finder:

-Finding the right learning rates.

Leslie Smith created the Learning Rate Finder in 2015. His idea was to start with a very, very small learning rate, something so small that we would never expect it to be too big to handle. We use that for one mini batch, find what the losses are afterwards, and then increase the learning rate by a certain percentage (e.g., doubling it each time). Then we do another mini batch, track the loss and double the learning rate again. We keep doing this until the loss gets worse instead of better. Now select at this point, a learning rate a bit lower than this shift.

Advice to pick either of these:

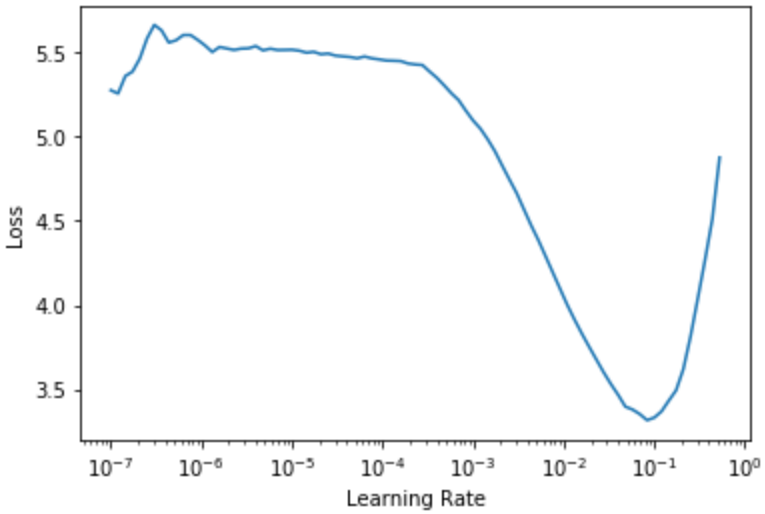
- One order of magnitude less than where the minimum loss was achieved (I.e., the minimum divided by 10).
- The last point where the loss was decreasing clearly.
-

Example with fastai:

The learning rate finder computes those points on the curve to help you. Both these rules usually give around the same value.

```
learn = cnn_learner(dls,resnet34,metrics=error_rate)

lr_min, lr_steep = learn.lr_find()
```



print(f"Minimum/10: {lr_min:.2e}, steepest point: {lr_steep:.2e}")

Minimum/10: 8.32e-03, steepest point: 6.31e-03