Learning Rates: What does Differential Learning Rates mean?

It means the layers of your DNN each have different learning rates during training (i.e. first layers have smaller learning rate, last layers have larger learning rate).

Learning Rates: What are cyclical learning rates?

It's essentially doing a logarithmic search for the optimal learning rate.

Learning Rates: What's the first step to finding an optimal learning rate via cyclical learning rates?

Do a training trial run using a low learning rate, but increasing it exponentially with each batch.

Learning Rates: What's the second step to finding an optimal learning rate via cyclical learning rates?

Plot the loss vs the learning rate, and then pick a learning rate right before the loss stops descending.

Learning Rates: What is Cosine Annealing?

Decreasing the learning rate over iterations as the loss function approaches a minimum in a cosine pattern (slow decrease, fast decrease, slow decrease).

Learning Rates: What is Stochastic Gradient Descent with Restarts (SGD-R)?

During training, every once in a while, increasing the learning rate suddenly so that, in

the event that it's in an undesirable local minimum, it will "hop" out.

Learning Rates: What is a cycle in SGD-R?

A cycle is the duration between sudden increases in the learning rate.