

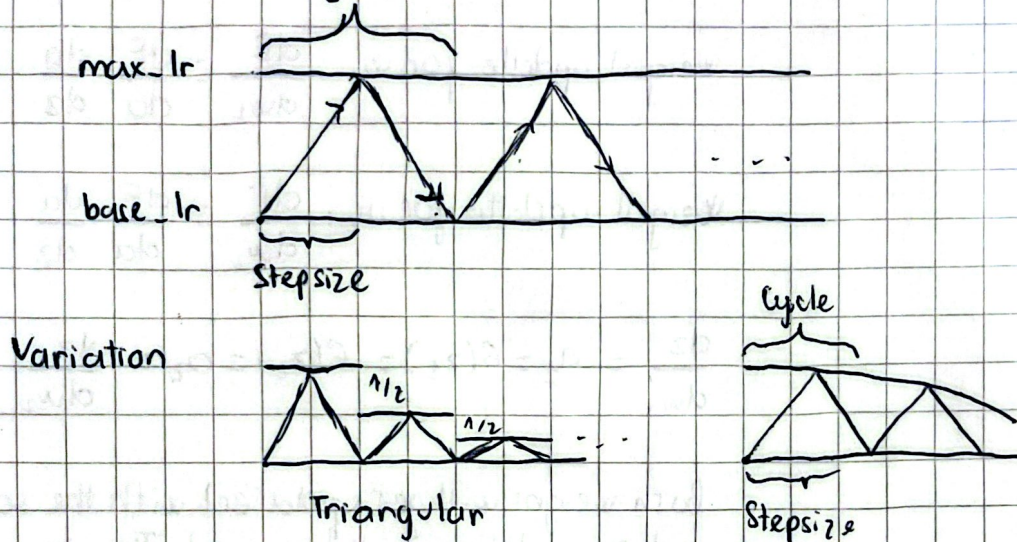
Section 6: Hyper parameter tuning and Learning Rate Scheduling

• Step decay: $lr_{new} = initial_lr \times decay_factor$

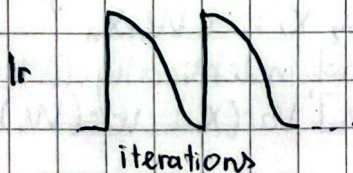
where $\left\{ \begin{array}{l} decay_factor = decay_rate^{fraction} \end{array} \right.$

$\left\{ \begin{array}{l} fraction = (epoch_start_decay_epoch) // decay_every \end{array} \right.$

• Cyclic Learning Rate and Restart



• Cosine Annealing with Warm Restarts



$$\eta_t = \eta_{min} + \frac{1}{2} (\eta_{max} - \eta_{min}) \left(1 + \cos\left(\frac{T_{cur} \pi}{T_i}\right) \right)$$

η_{max} : max lr

T_{cur} : number of epochs since the last restart

T_i : number of epochs between two warm restart

• Batch size vs lr