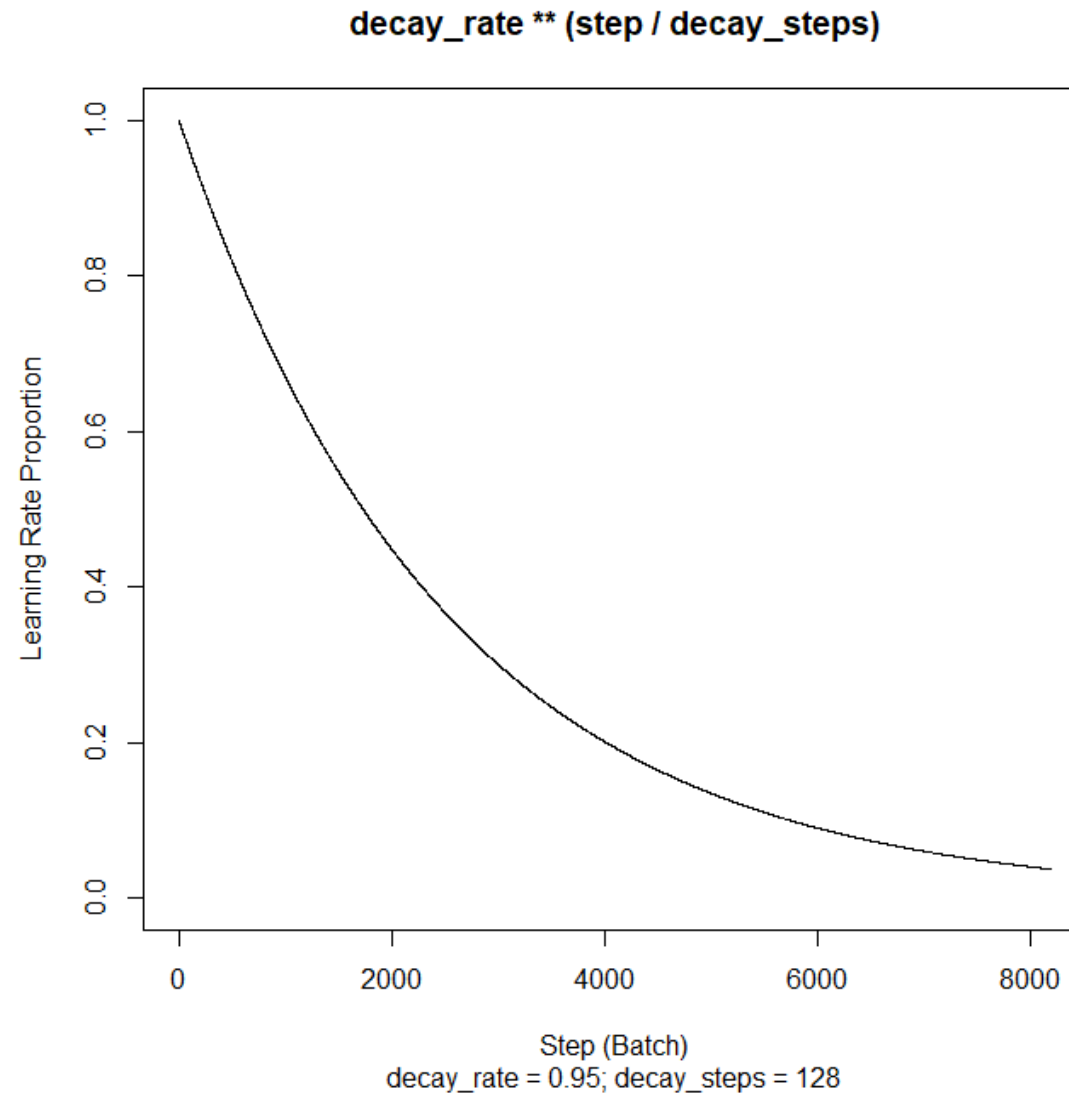
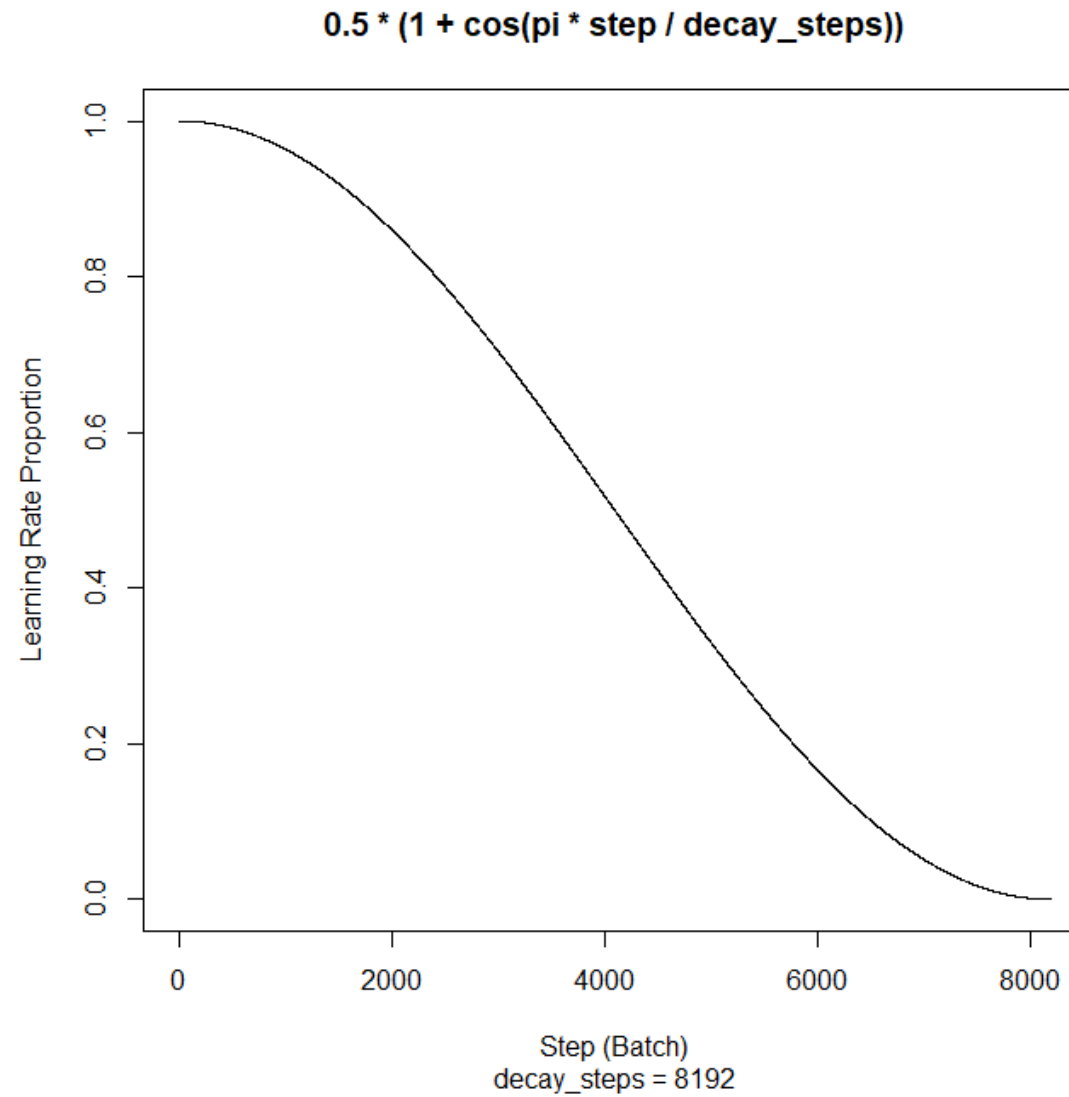


Exponential Decay [for learning rate]



We're gradually reducing the learning rate to 95% of the previous value every 128 steps

Cosine Decay [for learning rate]



Essentially the cosine function for inputs in $[0, \pi]$ shifted to the output range $[0, 1]$; can be combined with “restarts”