

Report your conclusions regarding the relation between  $B$  and  $\gamma$  and other findings you have learned from this exercise.

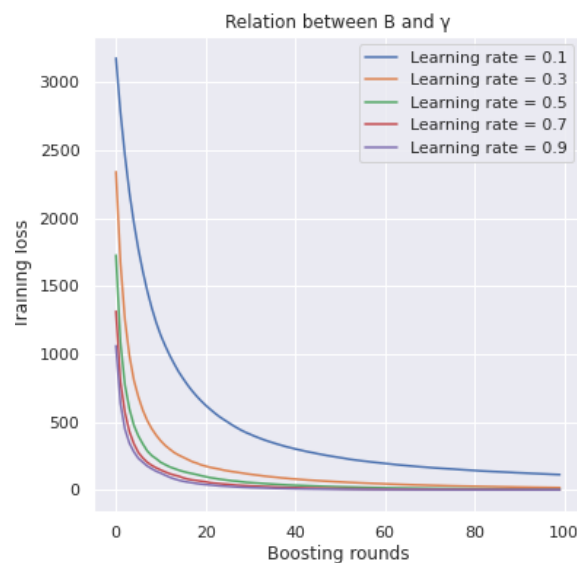


圖 1 Training Loss 與 Boosting Round 的關係圖( $B=100$ )

從圖 1 可以清楚看到，根據不同大小的 learning rate，在最開始所得到的 training loss 就有明顯差異，設定 learning rate 愈大則一開始 loss 就愈低，與模型的收斂速度成正比，而在上圖中可以看到在 boosting rounds 到 100 時，除了 learning rate 為 0.1 的曲線，其他線的 loss 都會非常接近，這就可以發現若模型一開始所設定的 learning rate 太小，在模型還沒得到最佳解之前，就會先停止迭代更新，為了要證實這件事，我另外又把 boosting rounds 設成 200，在圖 2 就可以看到 learning rate 為 0.1 的曲線又比圖 1 來的更接近最佳值。

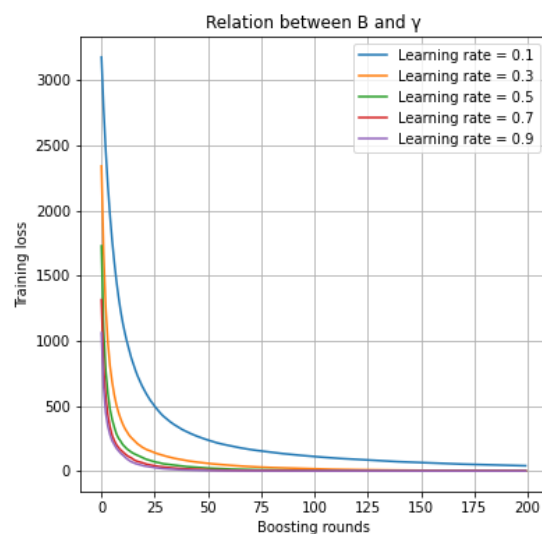


圖 2 Training Loss 與 Boosting Round 的關係圖( $B=200$ )