



In this homework, I tried vanilla SGD with three different learning rates (0.1, 0.01, 0.001), and RMSprop with default hyperparameters provided by PyTorch and three different learning rates (0.1, 0.01, 0.001).

In the task of training a ResNet for binary classification, RMSprop with $lr = 0.001$ and default hyperparameters is found to be the best among all combinations.

In general, RMSprop is found to be better than vanilla SGD such that it converges faster and on average reaches higher accuracy on validation set given the same number of epochs. Smaller learning rate also improves the training regardless of optimizer in this homework.