

MLP vs. MLP with  $XX^T$  data augmentation, train loss

