Project 2

Lizhang Miao

Abstract

In this report, we reproduce experiments in some papers to explore the capacity and non-overfitting puzzle of neural network. Different architectures of neural networks and CIFAR-10 are used in our experiments.

1 Experiment1

To test the capacity of MLP, we use standard CIFAR-10 and modified version dataset. Following setups are used in experiment1 to get more complicated datasets.

- True: the original CIFAR-10;
- Random label: true labels are replaced by random numbers;
- Shuffle pixels: a random permutation of the pixels is chosen and then the same permutation is applied to all the images in both training and test set.
- Random pixels: random permutation is applied to each image independently.
- Gaussian: a Gaussian distribution (with matching mean and variance to the original image dataset) is used to generate random pixels for each image.

In figure 1, we see regardless how data becomes complicate, MLP can fit the data well that training error achieve 0. While in figure 2, test accuracy gradually grows and stays the same after network getting 0 training error.

In figure 3, we explore more about the situation when labels are partially corrupted. According to the test accuracy, the dataset is indeed becomes more and more wierd, although the training loss or training error can achieves 0.

2 Experiment2

In experiment2, we train MLP on CIFAR-10 classification problem, and increase the parameters gradually. After 200 epochs or 0 training loss, we output the training error/loss and test error/loss, respectively. We observe that test error and loss perform inconsistently while training do as shown in figure 4.

3 Analysis

In experiment 1, we can easily see the expression power of neural networks. With the number of parameters increasing, the power improves, in another word, it gets better training accuracy. My explanation for the non-overfitting is following. Neural network is trying to find a feature representation such that the gap between different classes are enlarged. The more parameters it has, the sharper the representation is. For samples which can be classified ”intrinsicly”, more parameters help to distinct them better. For confusing samples, more parameters cannot lead to correct misclassifying, but also increase the distances. That may lead to loss increasing and keep the classification the same.

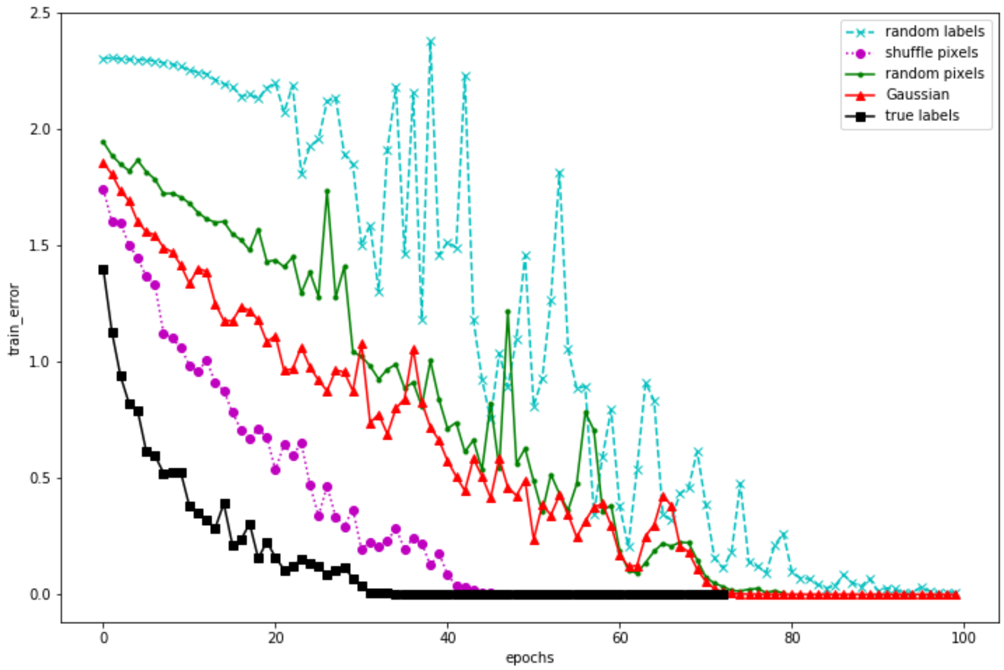

Figure 1: Training error decays with epochs under various dataset settings.

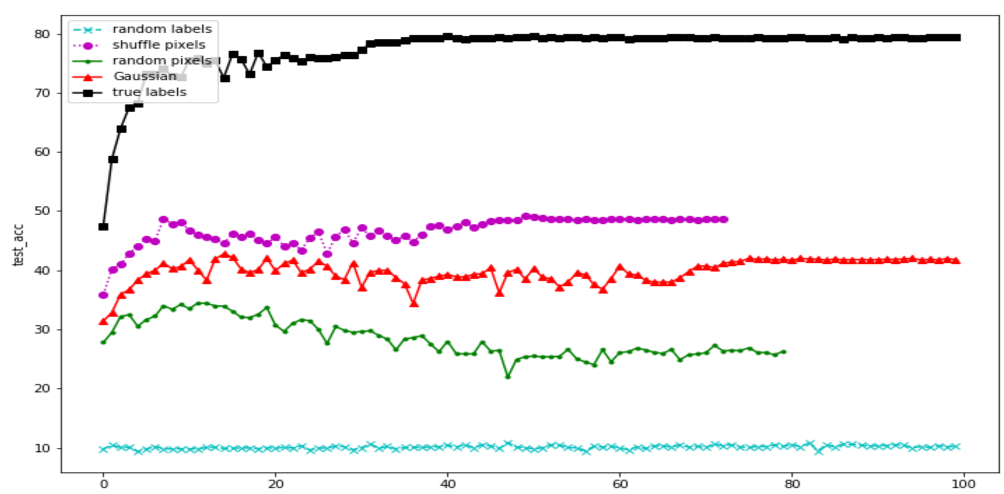

Figure 2: Test accuracy grows with epochs until network converges.

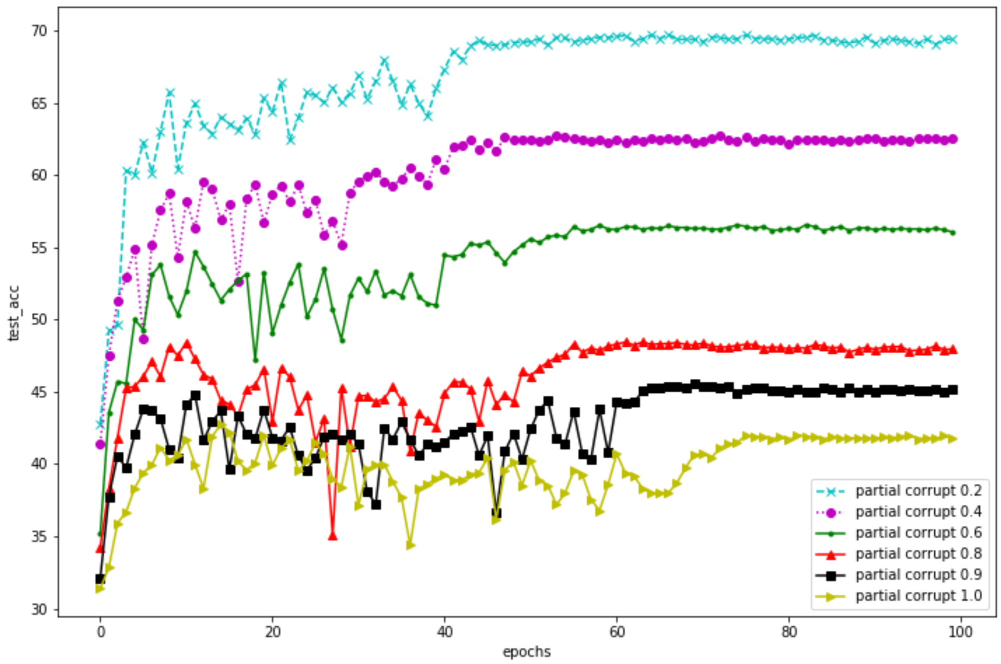

Figure 3: Look into the random labels case, the changes of test accuracy under different percentage of data corruption is shown in this figure.

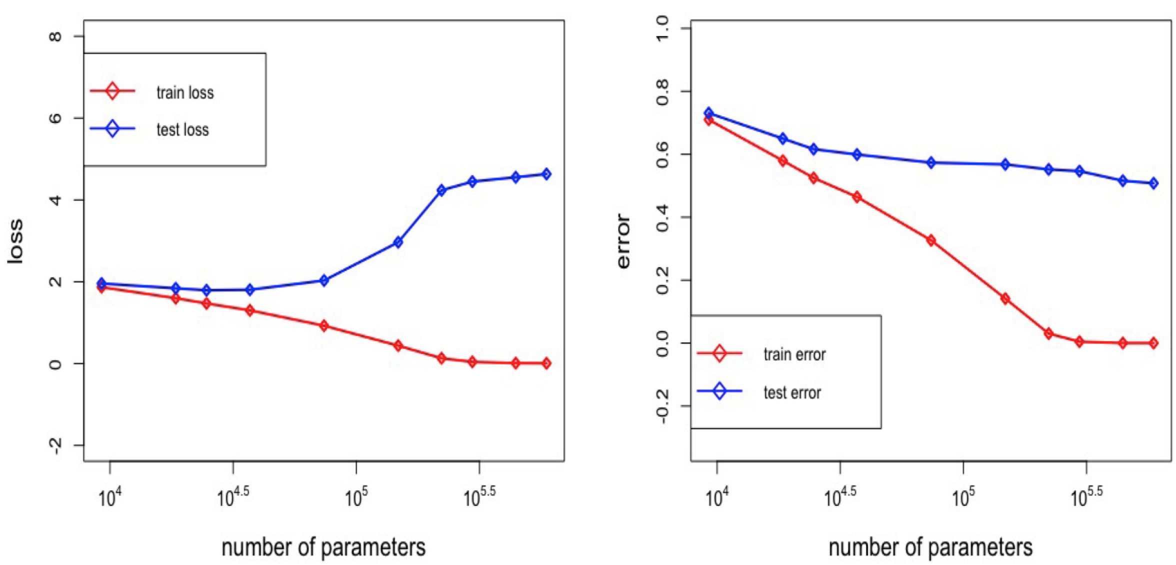

Figure 4: Training error decays with epochs under various dataset settings.