Deep Learning (Homework #2)

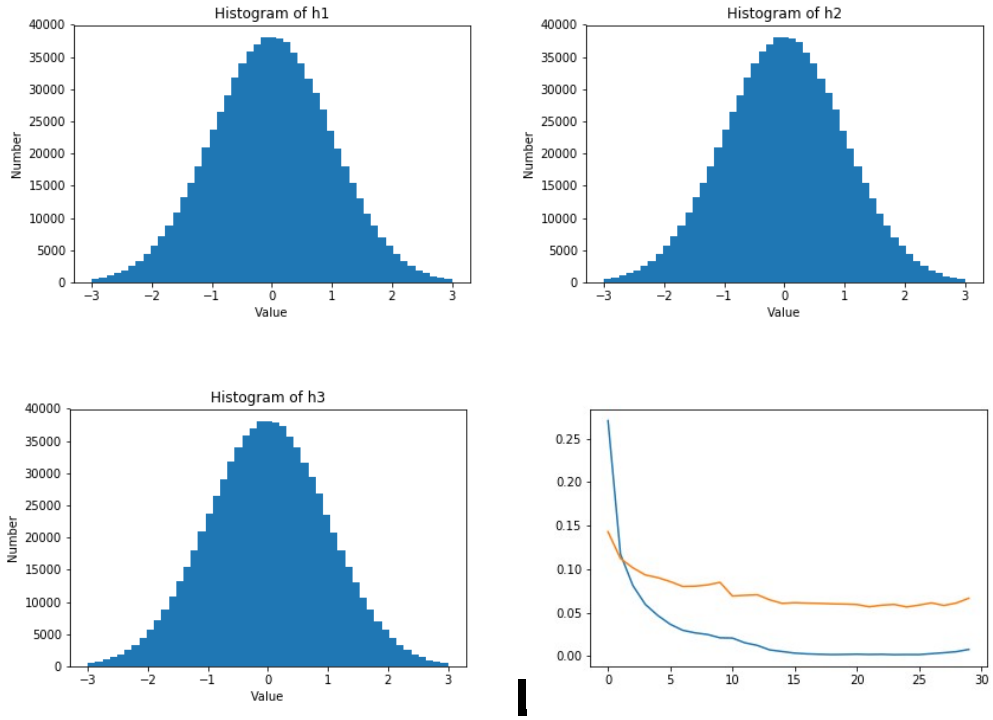
Due date: 4/14

1. Regularization

Overfitting is an important issue in deep learning. This is especially true in modern networks, which often have very large number of weights and biases. To train such a large model effectively, we would like to evaluate the techniques for reducing the effects of overfitting. Increasing the amount of training data or reducing the size of neural networks are two ways of reducing overfitting. Fortunately, there are different methods, which can reduce the overfitting by using L1 regularization, L2 regularization, Dropout. In this problem, you need to construct a Deep Neural Network (DNN) with three mentioned regularization techniques on MNIST dataset.

MNIST is a large dataset of hand-written digits, which contains a training set of 60,000 examples, and a test set of 10,000 examples. Each digit is displayed in an image of 28×28 as follows:

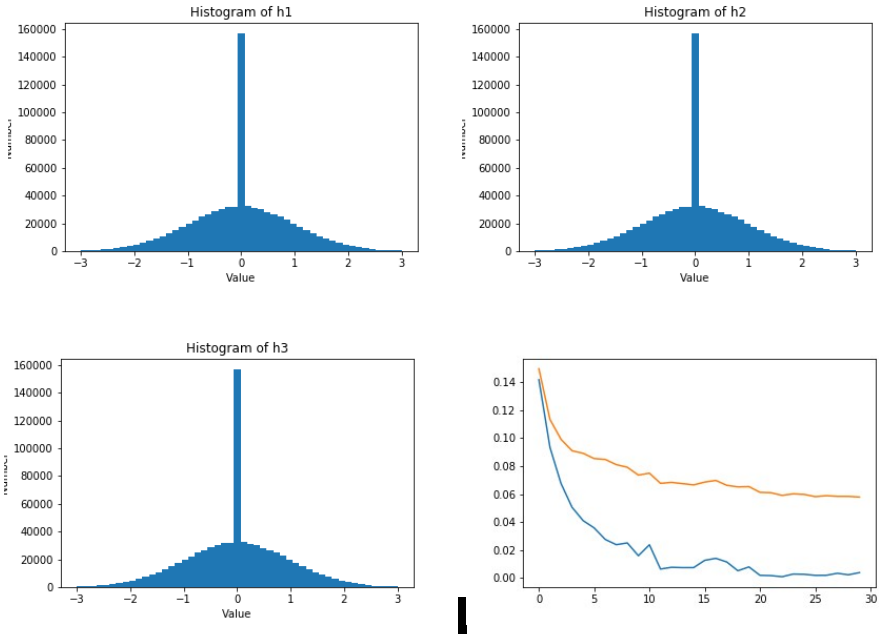
(a) Train a simple DNN and conduct some analyses in terms of accuracy under different settings (number of layers, neurons, etc.) and plot distribution of weights and biases in each layer and the learning curve of training and test sets.



(b) Train a DNN with L1 regularization:

$$E = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk} \ln t_{nk} + \alpha||\mathbf{w}||_1$$

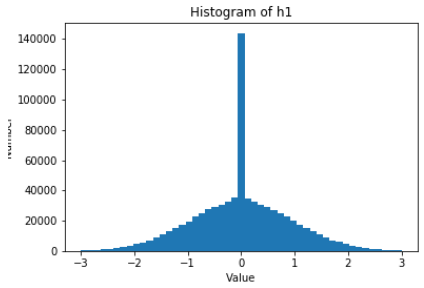
Please analyze the distribution of weights and biases in each layer under different settings (value of α, number of layers, etc.) and plot the learning curve of training and test sets.

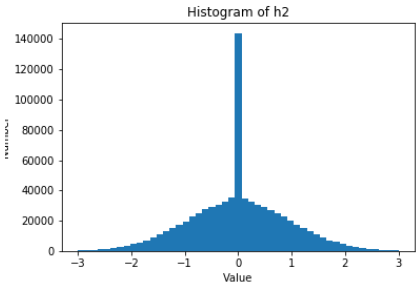


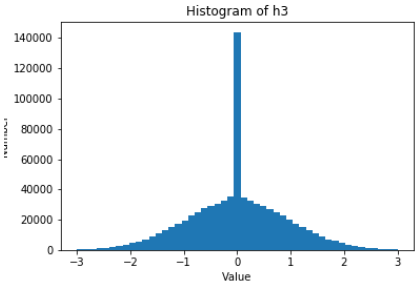
(c) Train a DNN with L2 regularization:

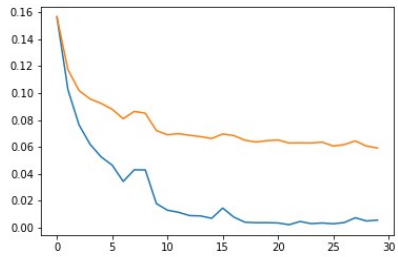
$$E = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} y_{nk} \ln t_{nk} + \alpha||\mathbf{w}||_2^2$$

Please analyze the distribution of weights and biases in each layer under different settings (value of α, number of layers, etc.) and plot the learning curve of training and test sets.

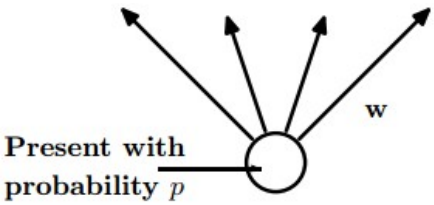


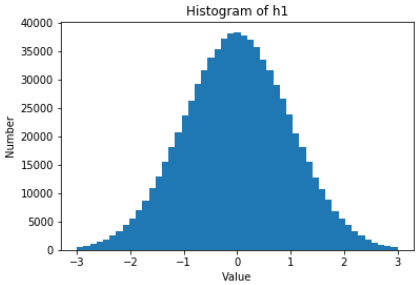


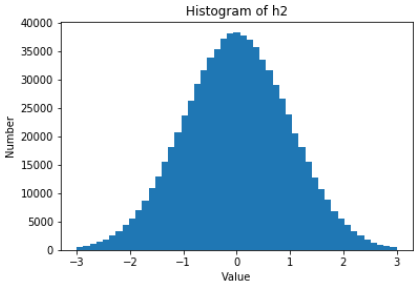


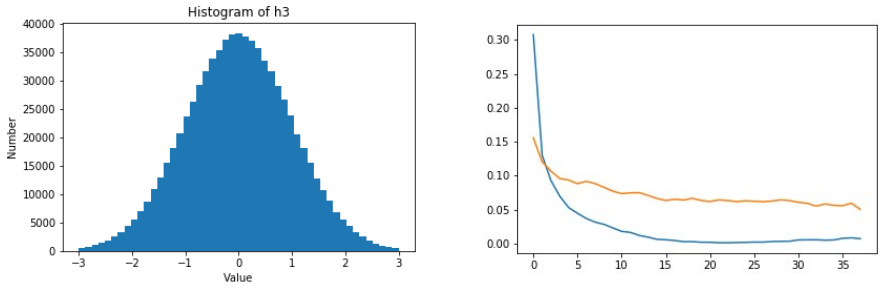


(d) Train a DNN with Dropout. Please analyze the distribution of weights and biases in each layer under different settings (value of p, number of layers, and etc.) and plot the learning curve of training and test set.





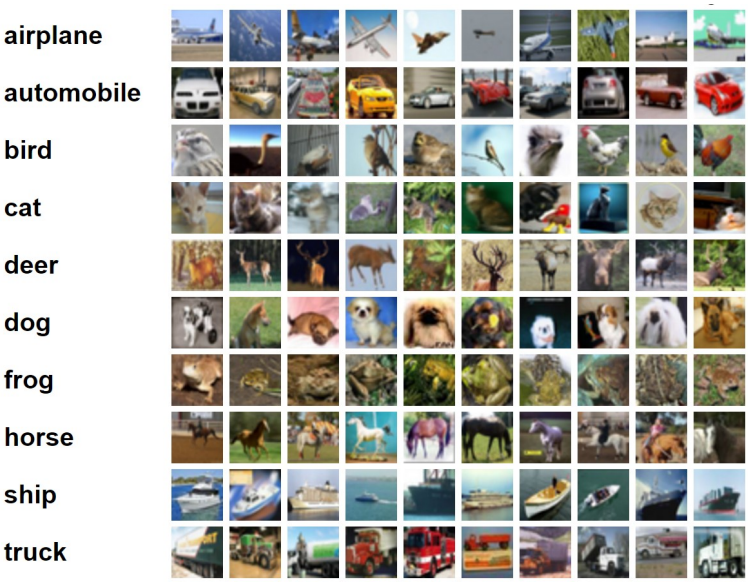




(e) Discuss the difference between L1 regularization, L2 regularization and Dropout in terms of your experiments from (a) to (d).

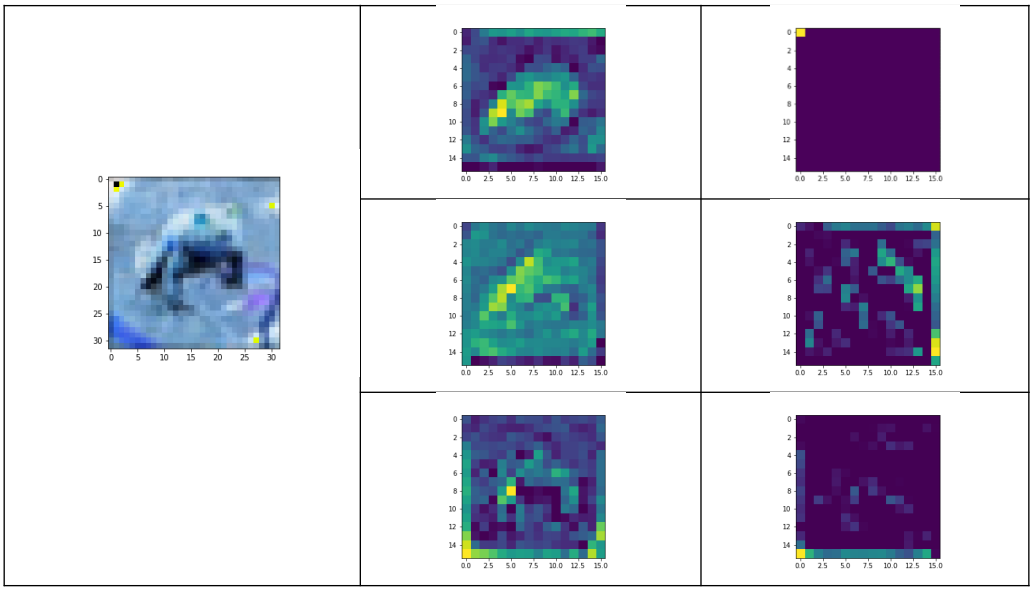
2. Convolutional Neural Network

In this problem, you will construct a Convolutional Neural Network (CNN) for image recognition using CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 32×32 color images in 10 classes, with 6,000 images per class. There are 50,000 training images and 10,000 test images. Here are the classes in the dataset, as well as 10 random images from each:



(a) Implement a CNN for image recognition using CIFAR-10. Analyze the effect of different settings including stride size and filter size. You should show the learning curve of training set and the final test error rate.

(b) Show some of feature maps in hidden layers as follows:



(c) Discuss about your feature maps.

MNIST dataset: http://yann.lecun.com/exdb/mnist/ or you can refer to tensorflow library

CIFAR-10 dataset: https://www.cs.toronto.edu/~kriz/cifar.html