**Problem Set 4**

Weifang Zheng 50248714

1. **Problem 1**

**Demonstrate that a neural network to maximize the log likelihood of observing the training data is one that has softmax output nodes and minimizes the criterion function of the negative log probability of training data set:**

is the label of image .

is the prediction of according to function f, y = f() =

(m is the index of labels (0,1,2,3,4,5,6,7,8,9), n is the index of input images)

We can define the loss function Loss =

Total loss equals to the number of images which are classified into the right classes.

To maximum the Loss, we need to determine the W to make each image can be assigned to the right classes.

softmax function:

the output of output layer is a [1,n] vector.

The index of max element in the vector will be chosen as the prediction.

To maximum the log likelihood can be transferred to minimum the negative log likelihood.

=

p() is evidence which are 1/n

**Demonstrate that a neural network to maximize the posterior likelihood of observing the training data given a Gaussian prior of the weight distribution is one that minimizes the criterion function with L2 regularization**

, , …

is the Gaussian noise with mean 0 and variance

Gaussian likelihood:

Regularize parameter W by imposing the Gaussian prior

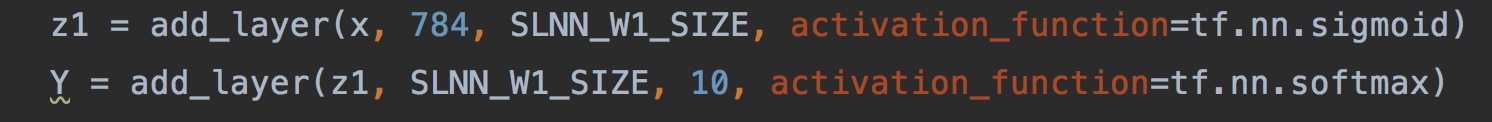
Where is a strictly positive scalar. Hence, combining the likelihood and the prior we have

Logarithm and dropping some constants:+const

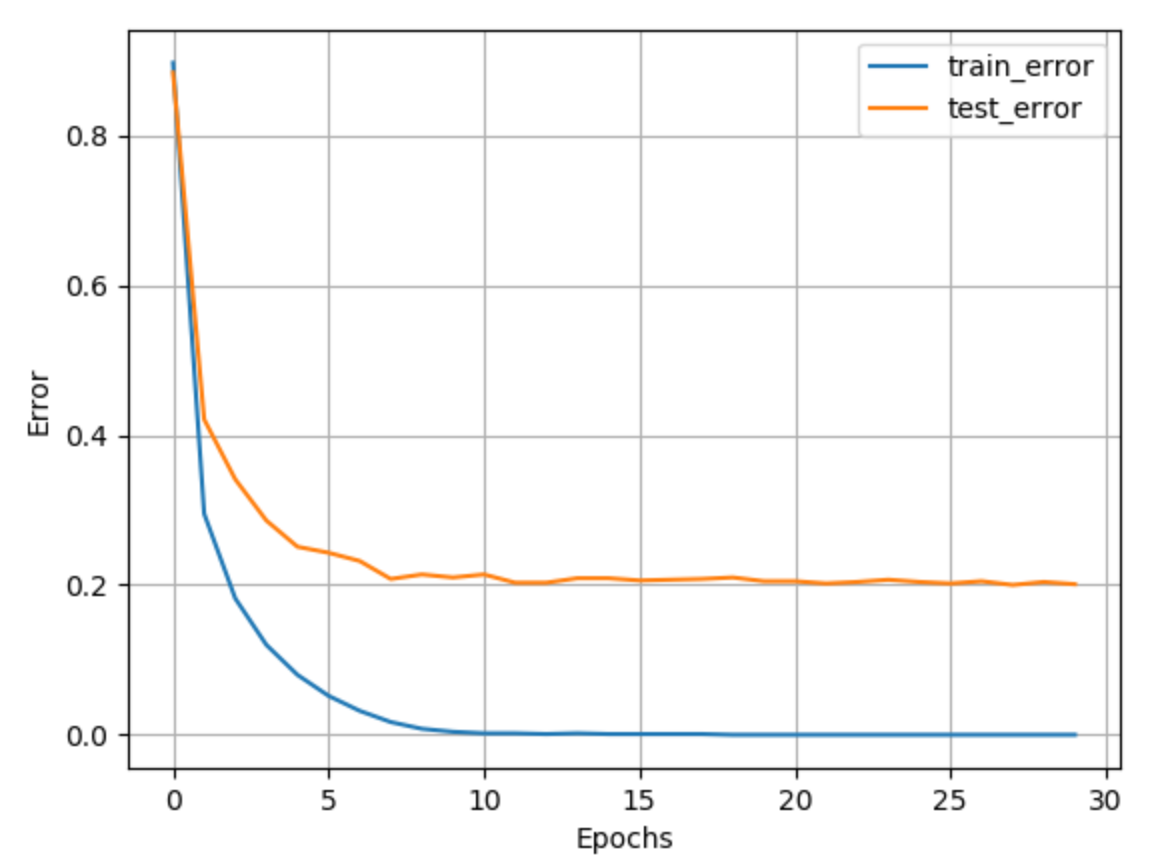
If we maximize the above expression with respect to β we get the so called maximum a-posteriori estimate for β. In this expression it becomes apparent why the Gaussian prior can be interpreted as a L2 regularisation term.

reference: <https://stats.stackexchange.com/questions/163388/l2-regularization-is-equivalent-to-gaussian-prior>

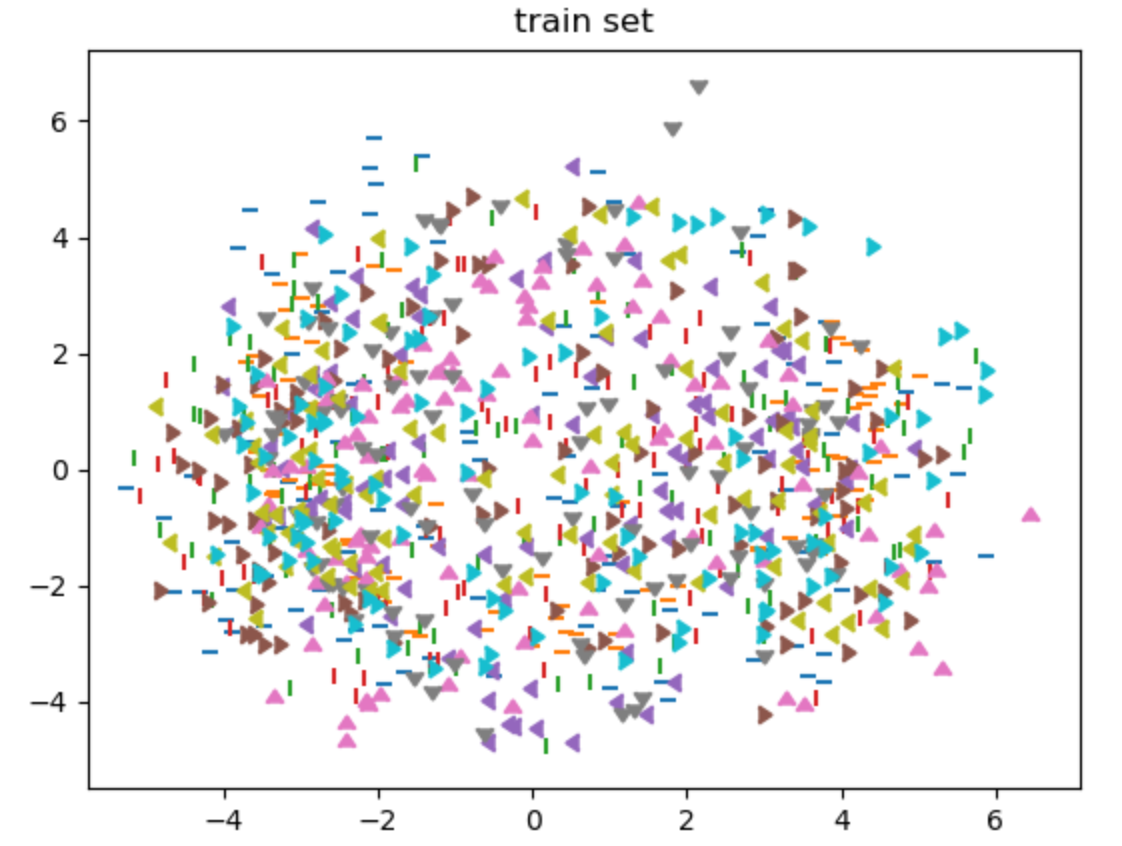
1. **Problem 2**
2. Single hidden layer NN

****

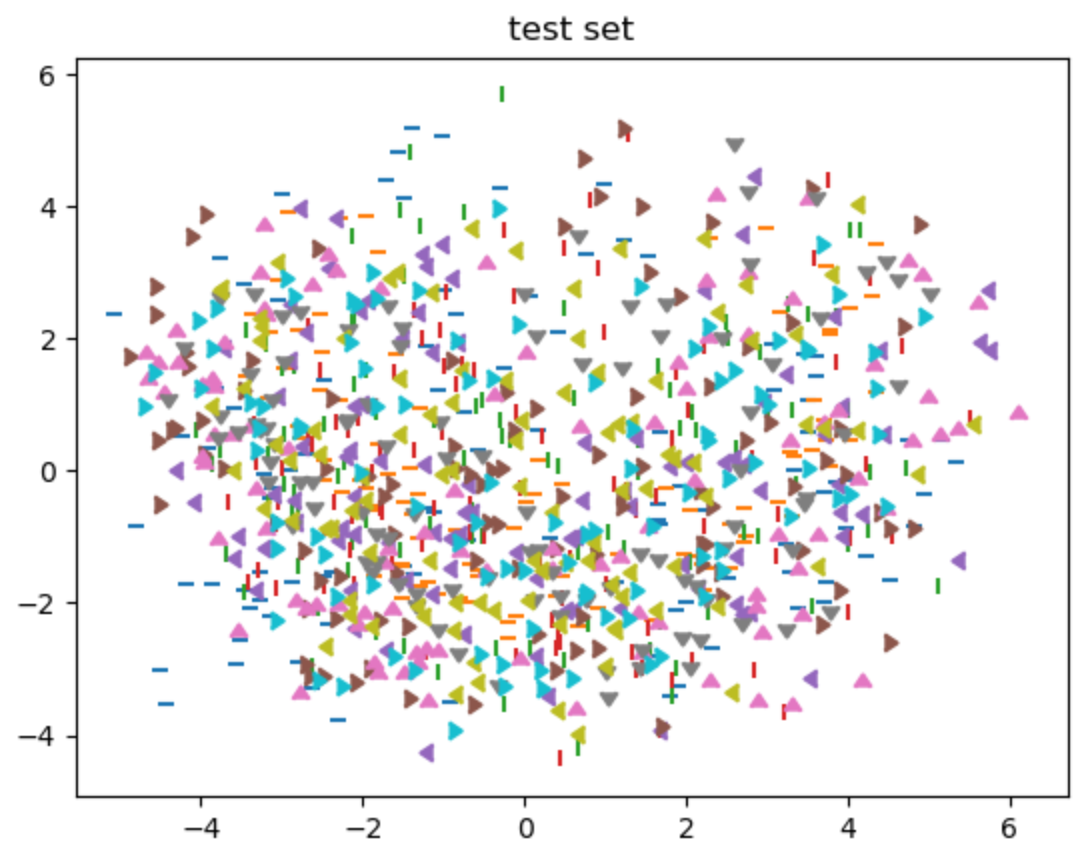
Plot:



  criterion function on training data set（ PCA）：



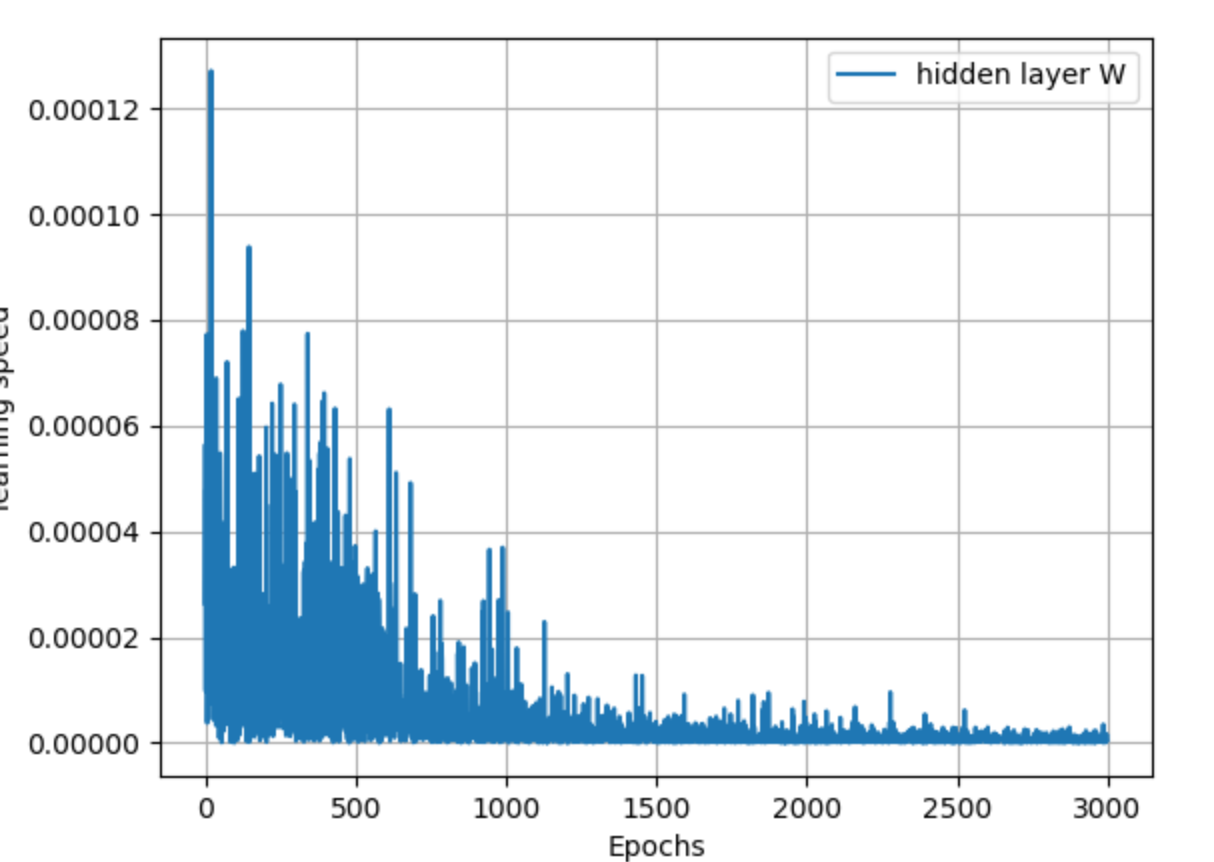
criterion function on testing data（ PCA）：



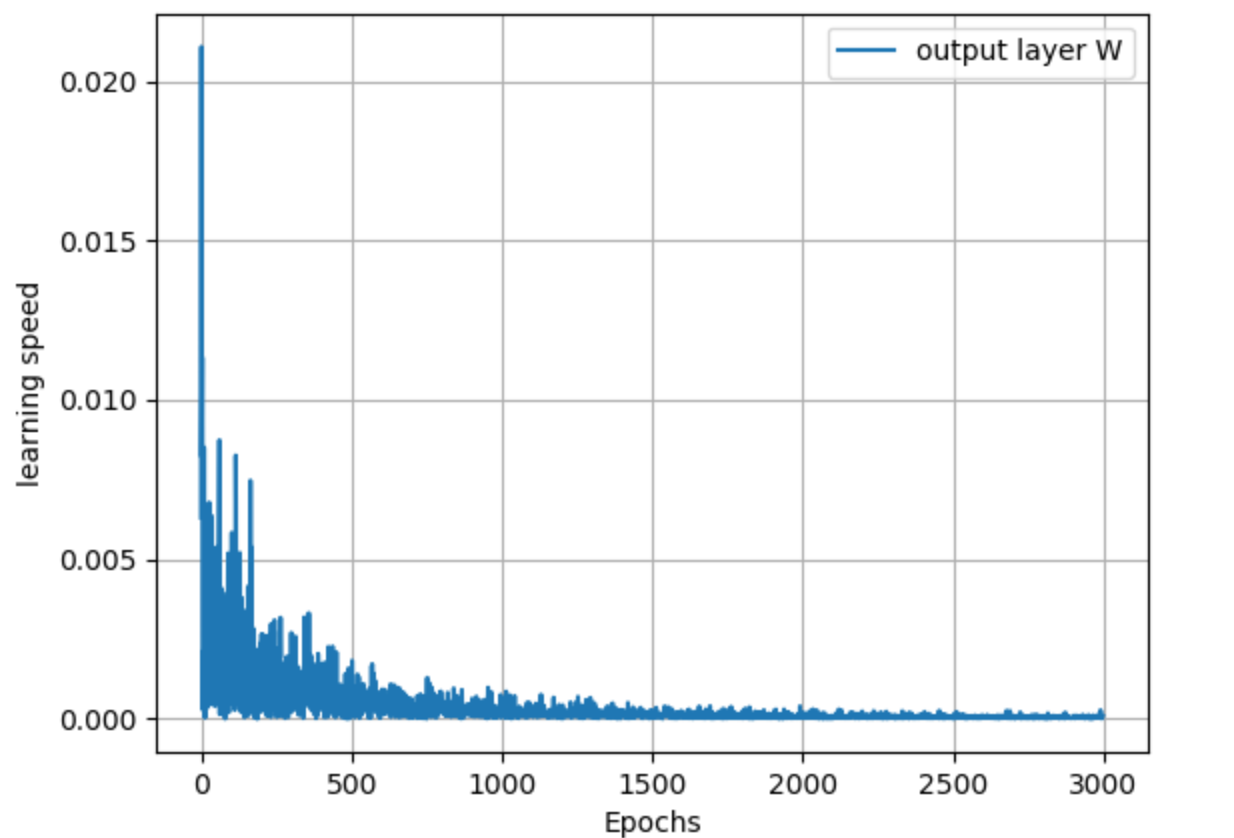
Learning speed:

Use L1 regularizer to estimate the learning speed.

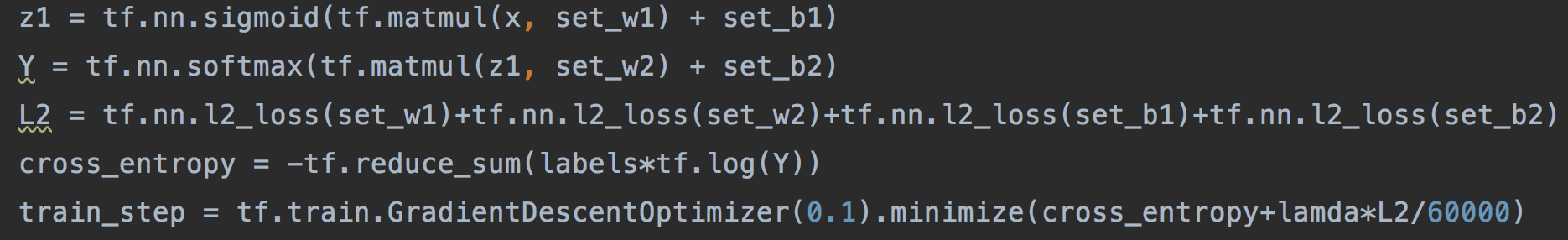
learning speed of W in hidden layer

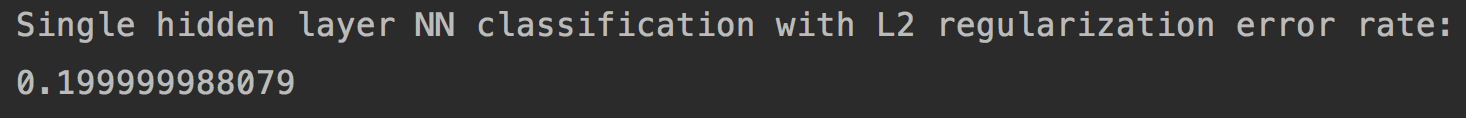


learning speed of W in output layer

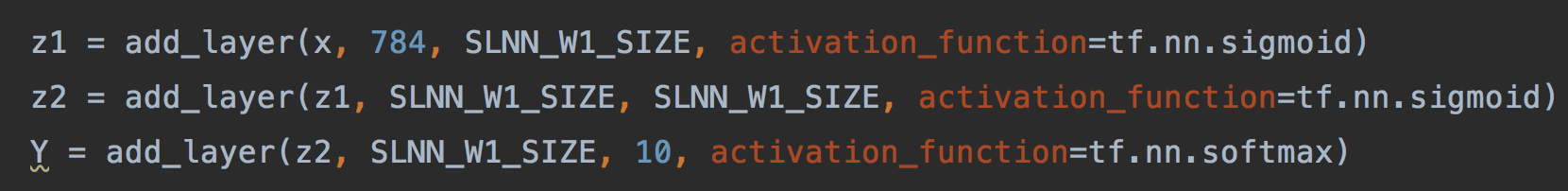


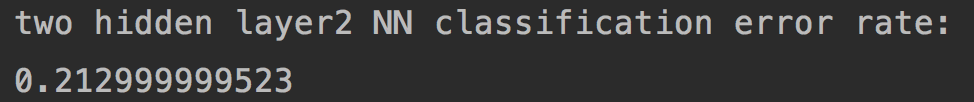
1. Single hidden layer NN with L2 regularizer



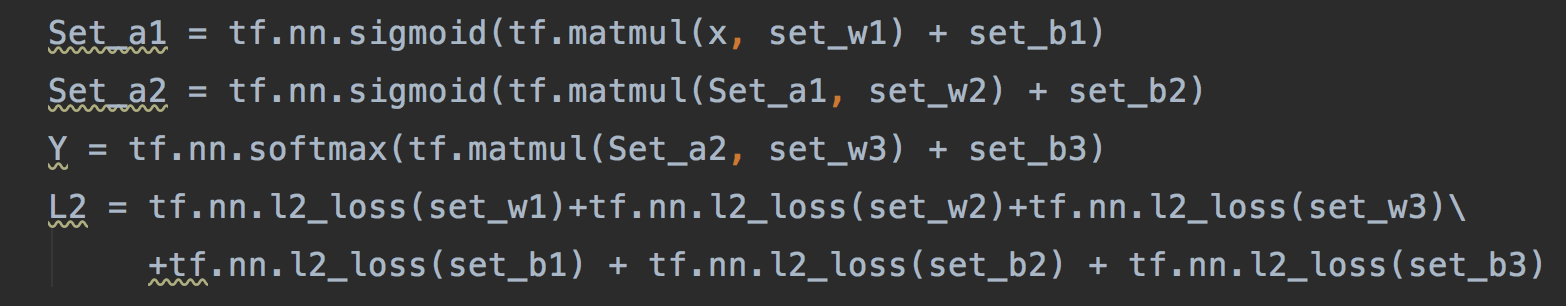


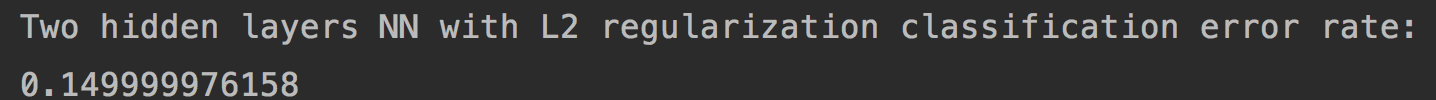
Two hidden layers NN



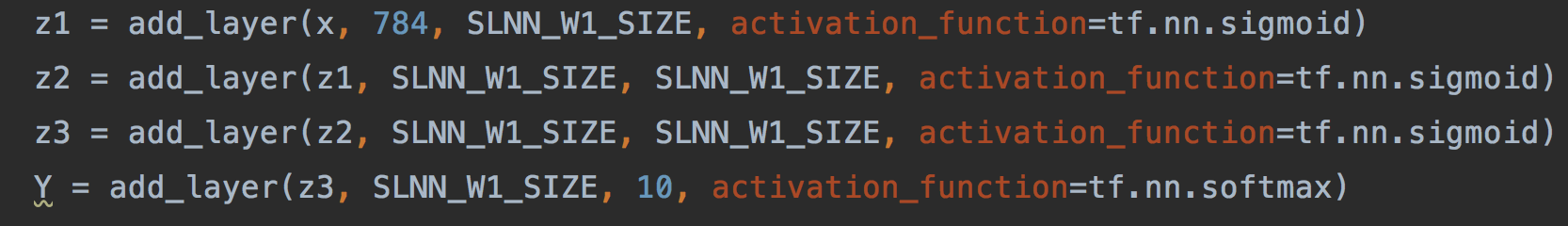


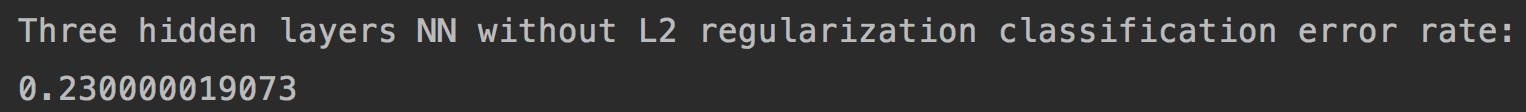
Two hidden layers NN with L2 regularizer



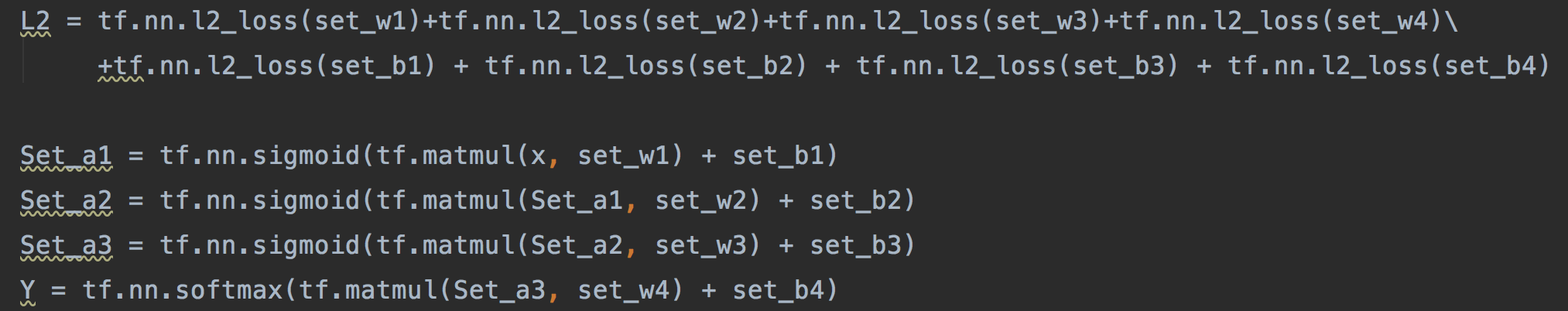


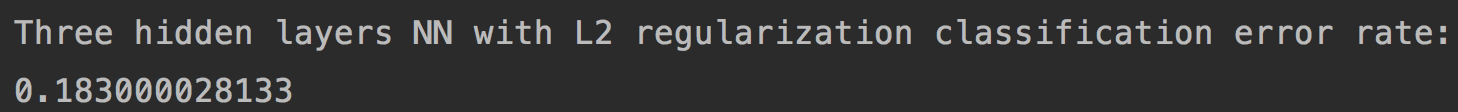
Three hidden layers NN





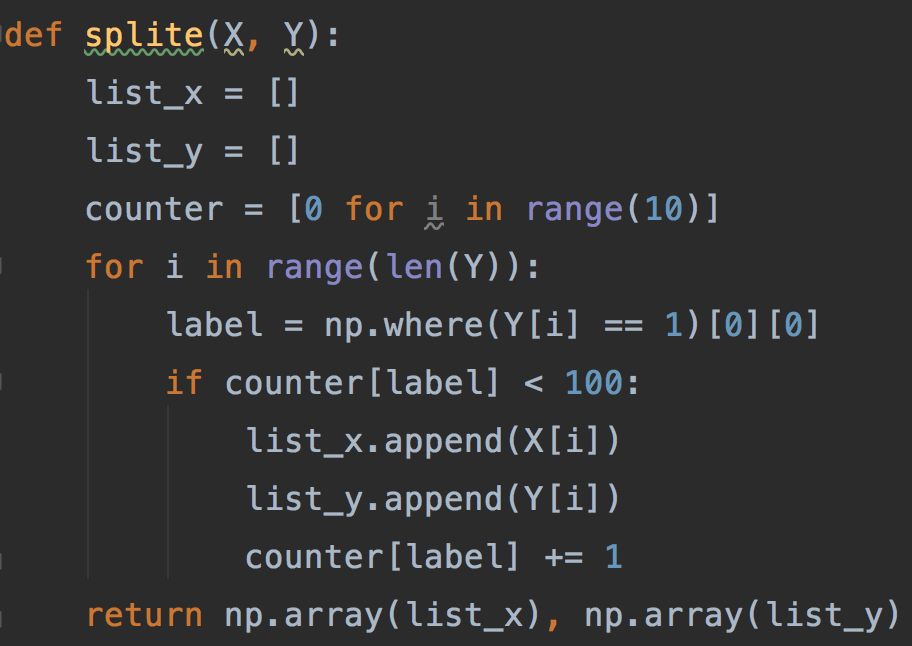
Three hidden layers NN with L2 regularizer





1. CNN

1000 training images (100 images per digit)



Regularize the training of neural network through augment your selection of 1000 images by rotating them for 1-3 degrees clockwise and counter clockwise, and shifting them for 3 pixels in 8 different directions.



