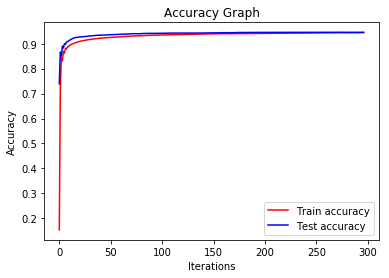
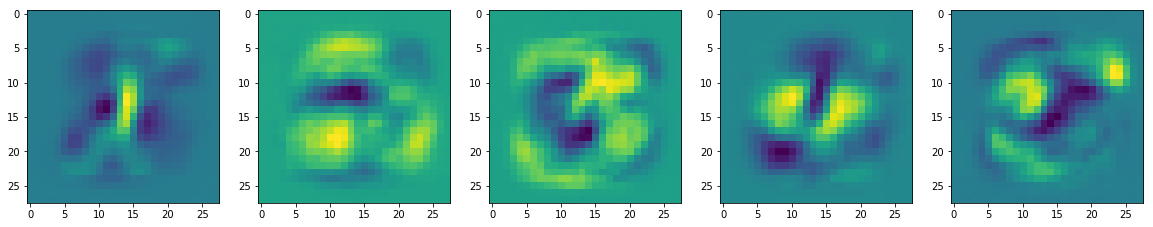
Programming Assignment 1 Report

Note: must initialize weights to small values to make the weight plots work!!

**W** and **b** are the weights and bias in the neural network that is used to predict the digit given an image. The neural network model used in this assignment is a single layer neural net (no hidden layers) with input size of 784 (all pixels of an image) and output size of 5 (one-hot encode corresponding to each digit). The weight is a 784\*5 matrix and the bias is a 5-vector which both initialized with small values. The learning rate of this model is 0.00001. The tolerance is 0.0001 which is the criteria to stop training if the loss difference between two consecutive iteration is less than that value. The gradient descent converges after approximately 300 iterations. The training and testing accuracy graph is shown below.



The learnt weights are displayed below.



The classification performance as a confusion matrix is shown below.

(The code for the confusion matrix is from

<https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html>)

