The German University in Cairo

Department Computer Science and Engineering

CSEN 1067 Machine Learning Algorithms

Assignment 2 Report

Team Name:
Berliners

Team Members:
31-1020 Ahmed Hassan Abdou Gomaa Ouda
31-15420 Mohamed Atef Hussein Ahmed
31-626 Patrick Mounir Naeem Attia

October 27, 2017

1 Different Configurations

1.1 Number of Hidden nodes

It is clear from Figure 1 that the the one with the least number of hidden layers has the smallest generalization gap between the training error and validation error, and it is also the one wit the least mean absolute error. This is because with more hidden layers the model tends to over-fit as the validation error increases and the training error decreases apparent, the model starts to go out of its way to fit the points of in the training set.

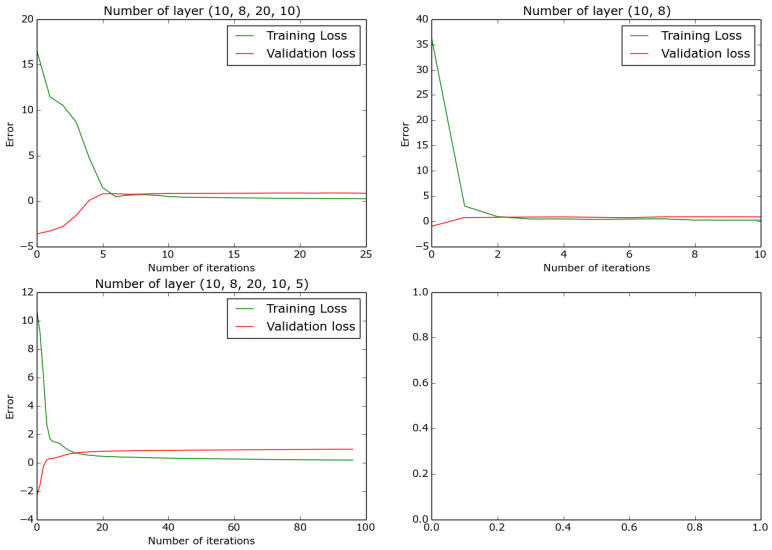

Figure 1: Number of Hidden Nodes Error.

1.2 Learning Rate

It is apparent from Figure 2 that there is a difference between learning rate of 0.001 and 0.00001, it is clear that the smaller learning rate the error descends much smoother, since the step size at each iteration is smaller. The larger learning rate stops much earlier that the smaller one, this is also again to the difference in step size and the learning rate reaches the local minimum much quicker.

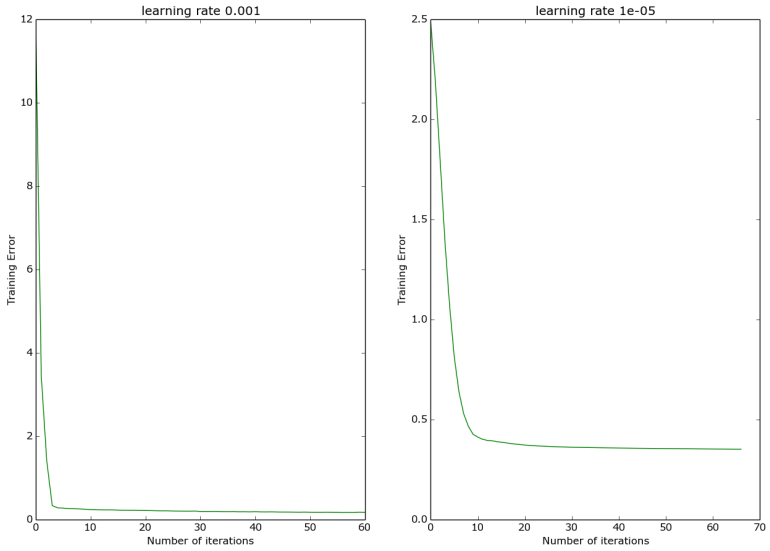

Figure 2: Learning Rate error.

2 Optimal Configuration

As seen in both Figures 1 and 2, that the optimal configuration would be when the learning rate is 0.001 and the Hidden Layers $(10, 8)$ with this we would get the left figure in Figure 3 that there is a gap which may be due some overfitting. In order to prevent this, we use regularization of the wights with $\lambda = 1e^{-5}$, which would make the error of the validation and the training closer as seen in Figure 3.

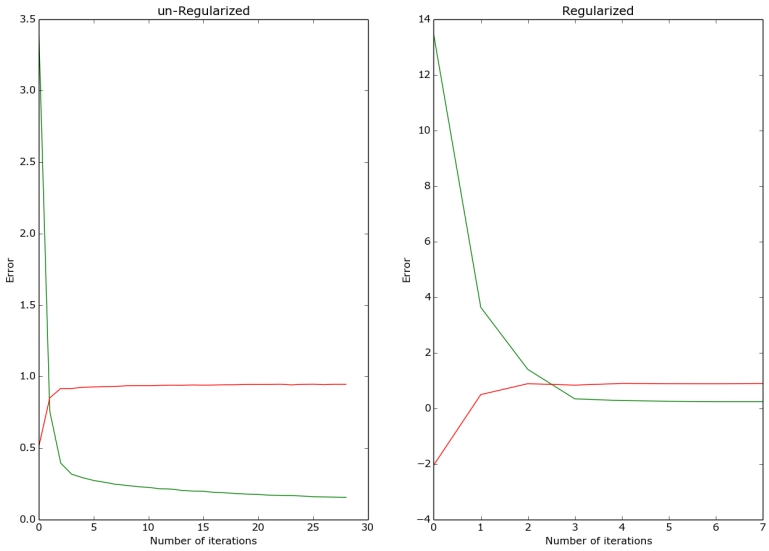

Figure 3: Optimal Configuration.