CSL 467: Homework #3

Due on Friday, September 13, 2014

Narayanan C Krishnan

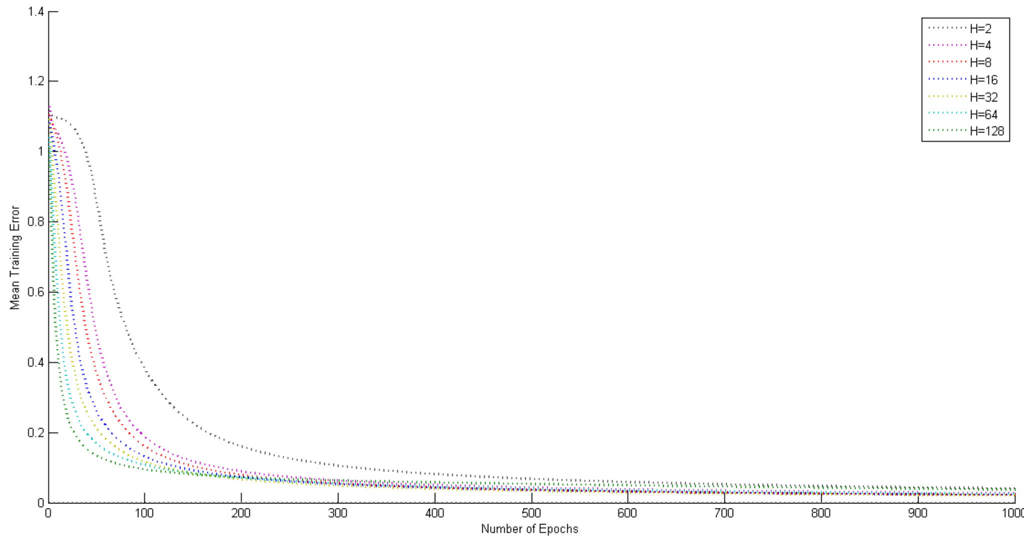
Harsimran Singh

Contents

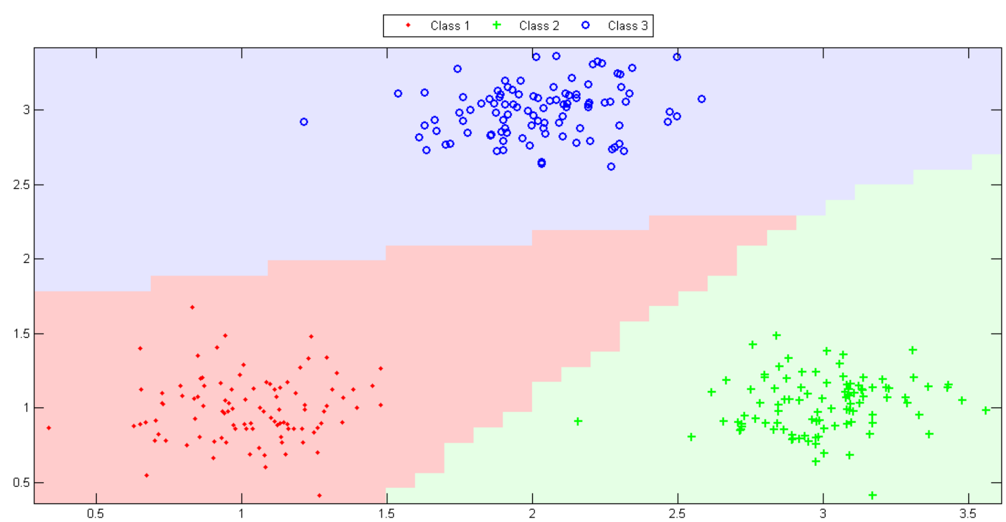
Problem 1

(d)

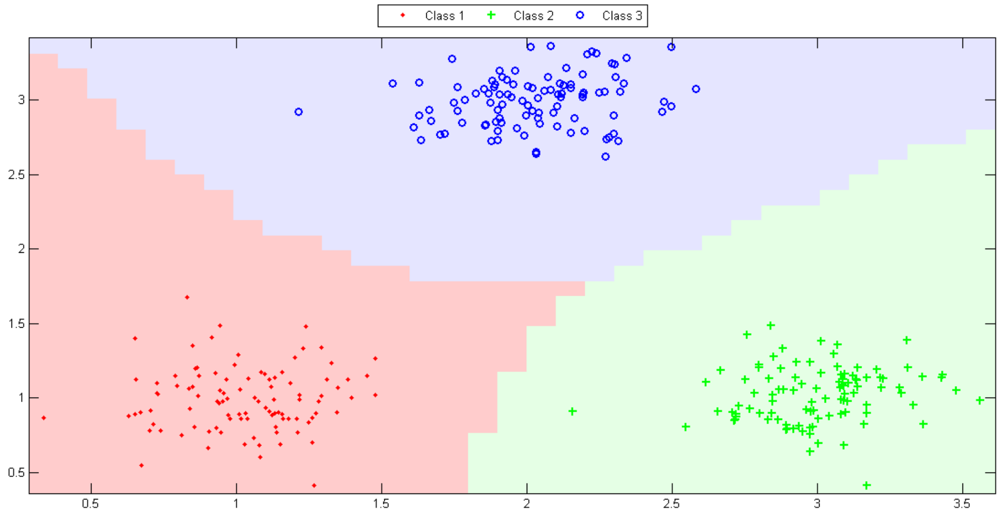
Mean Training Error is plotted vs Number of epochs with varying values of number of hidden layers from 2 to 128 in powers of 2. The Number of epochs are 1000 and learning rate is 0.01. As, we can see that as we increase the number of hidden layers, trainerror decreases faster (faster convergence) but as number of hidden layers increases, the rate of convergence does not increase as much.



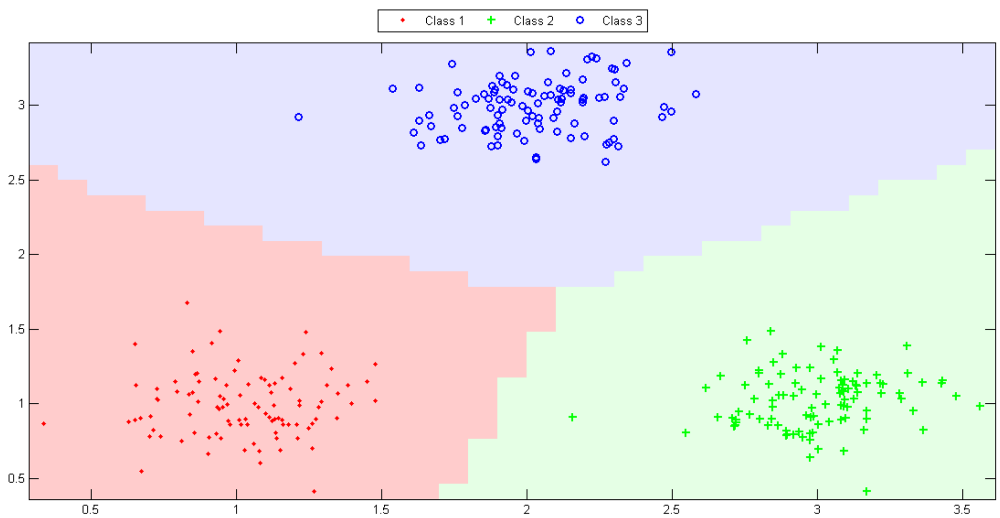
Following is the plot of approximate decision boundary for number of epochs = 1000, ***number of hidden layer units = 2*** and learning rate = 0.01.



Following is the plot of approximate decision boundary for number of epochs = 1000, ***number of hidden layer units = 4*** and learning rate = 0.01.



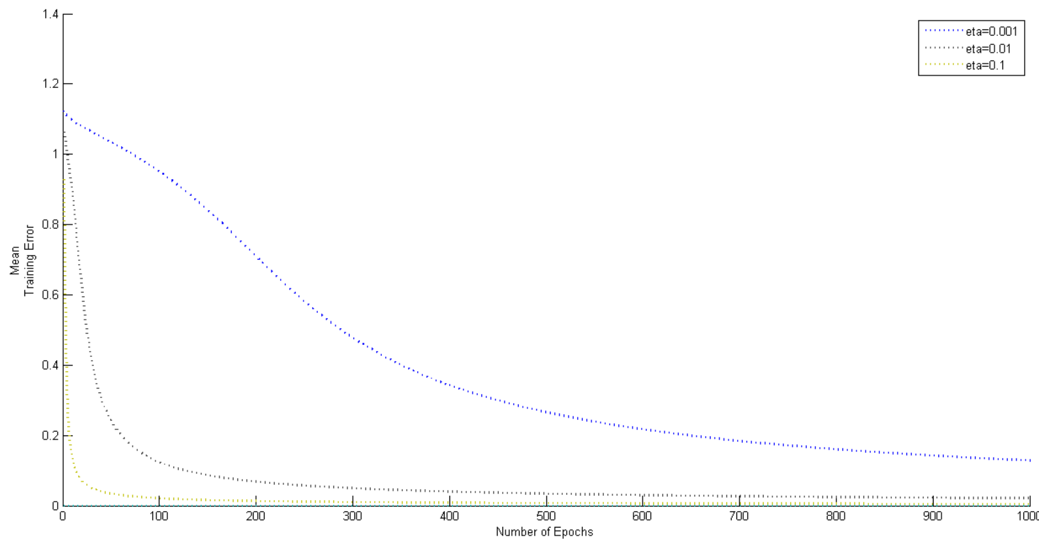
Following is the plot of approximate decision boundary for number of epochs = 1000, ***number of hidden layer units = 64*** and learning rate = 0.01.



The observation taken from above three plots is that as number of hidden layer units increases decision boundary becomes more complex. We can see that when number of hidden layer units are 2 decision boundaries are almost straight.

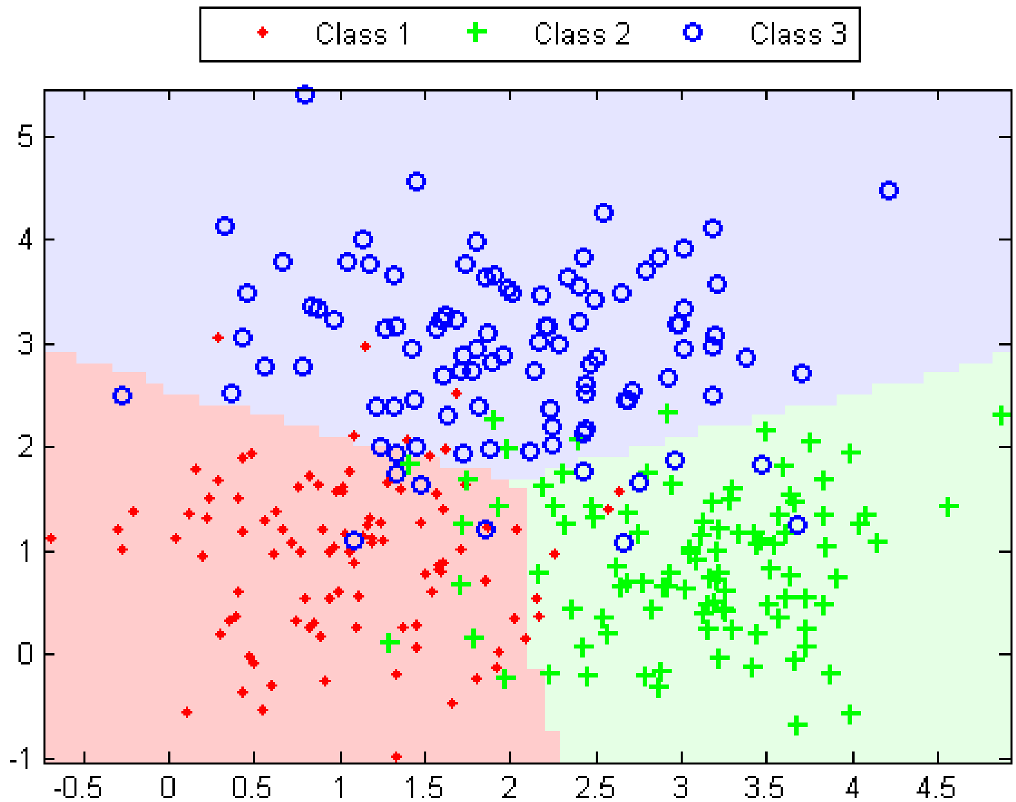
(e)

Mean Training Error is plotted vs Number of epochs with varying values of learning rate (0.001, 0.01, 0.1). The Number of epochs are 1000 and number of hidden layer units are 16. As learning rate increases, the rate of convergence increases.
At ***eta=0.001***, mean training error is 0.2 even after 1000 epochs.
At ***eta=0.01***, mean training error is reduced to 0.2 after 100 epochs approximately.
At ***eta=0.1***, mean training error is almost zero after 100 epochs.

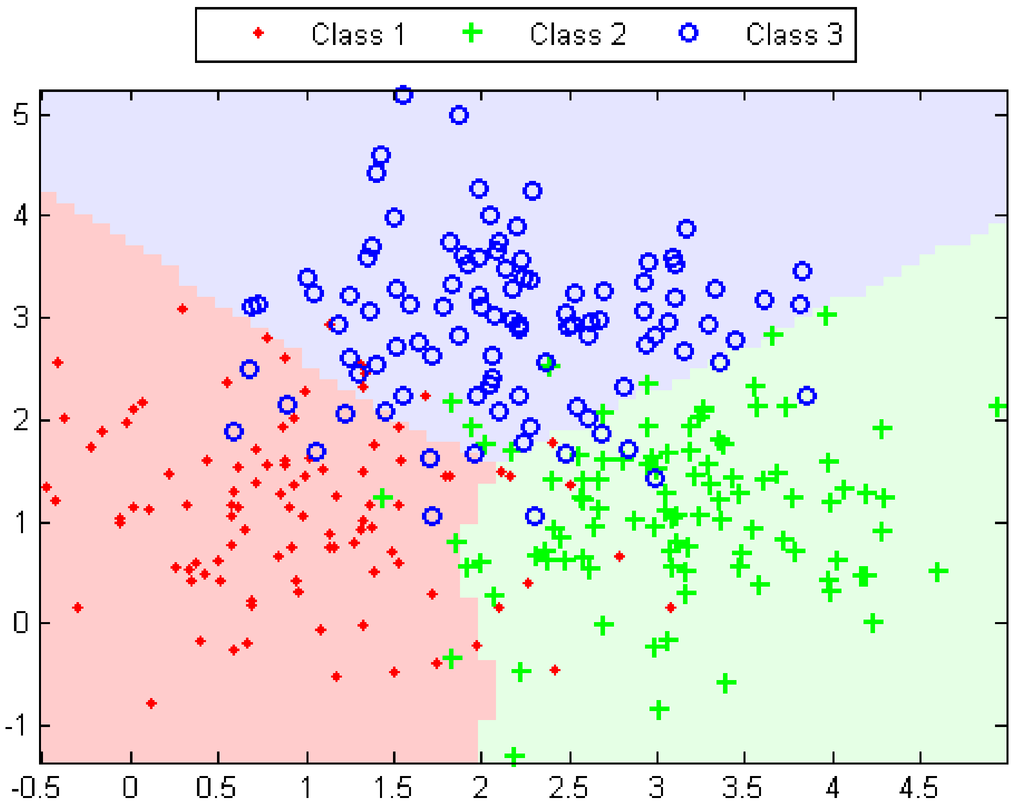


(f)

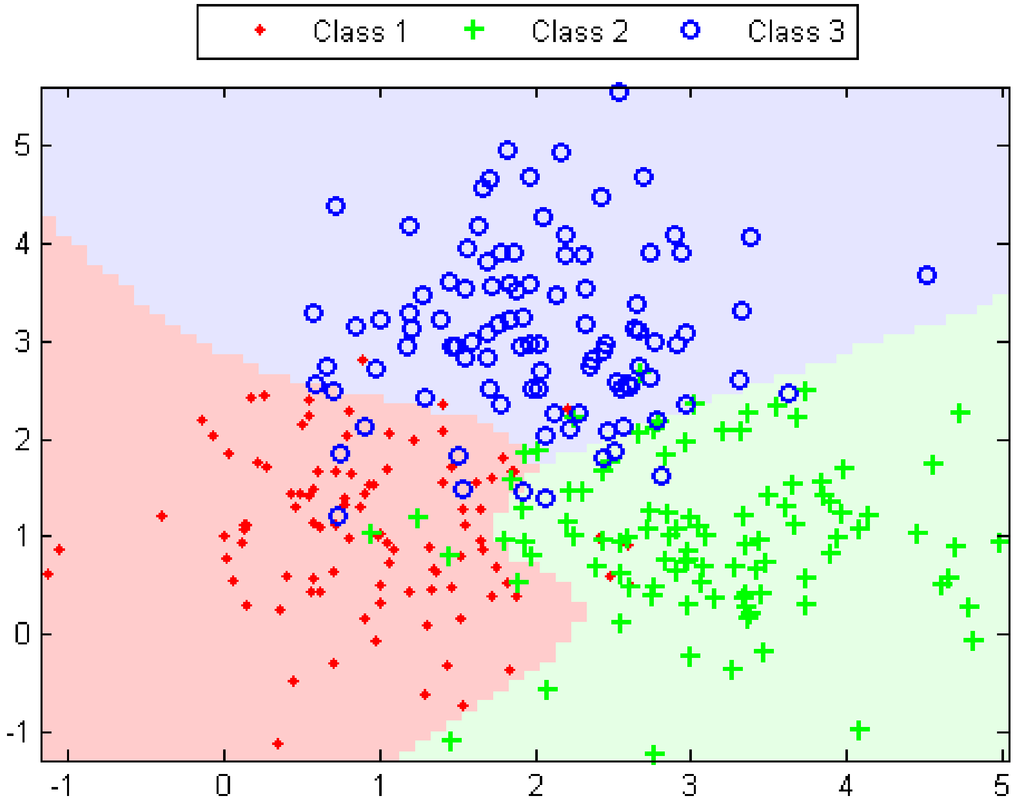
Following is the plot of approximate decision boundary for ***number of epochs = 1000***, number of hidden layer units = 2 and learning rate = 0.01. Training error was 0.3491.



Following is the plot of approximate decision boundary for ***number of epochs = 5000***, number of hidden layer units = 2 and learning rate = 0.01. Training error was 0.3455.



Following is the plot of approximate decision boundary for ***number of epochs = 10000***, number of hidden layer units = 2 and learning rate = 0.01. Training error was 0.3399.



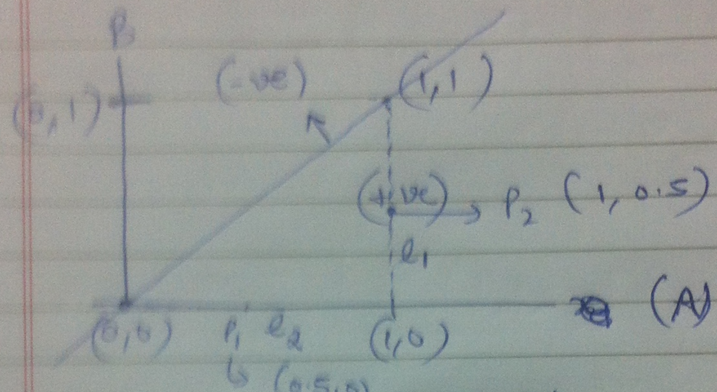
As number of epochs are increased the decision boundaries are becoming more complex, and they are fitting the training data more accurately.

Problem 2

Problem 3

(2)

A	B	$A \wedge \neg B$
0	0	0
0	1	0
1	0	1
1	1	0



Any line ~~passing through~~ intersecting l_1 & l_2 will classify the data.

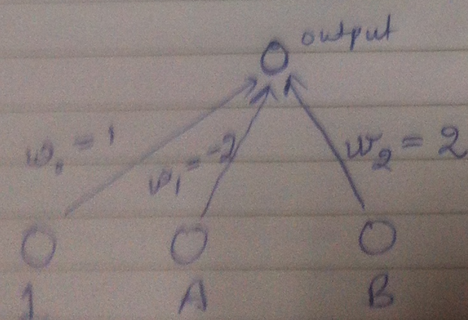
Take line passing through P_1 & P_2

$$2B - 2A + 1 = 0$$

Compare with

$$w_0 + w_1 A + w_2 B$$

Perceptron will look like follows



classmate

Date

Page

If $w_0 + w_1 A + w_2 B > 0$

then input is negatively classified

& if $w_0 + w_1 A + w_2 B < 0$

then input is positively classified.

Problem 4

(4) (a) Error, $E = -\sum_{n=1}^{N} \sum_{i=1}^{K} t_{ni} \log(ydash_{ni})$

$$\Delta v_{hj} = -\frac{\partial E}{\partial V_{hj}}$$

$$= \frac{\partial}{\partial V_{hj}} \left[\sum_{n=1}^{N} \sum_{i=1}^{K} t_{ni} \log(ydash_{ni}) \right]$$

for $i = h$ ($\because$ for soft max func. differential is different for $i=h$ & $i \neq h$)

$$\frac{\partial \left(t_{nh} \log(ydash_{nh}) \right)}{\partial V_{hj}}$$

$$= \frac{t_{nh}}{ydash_{nh}} \cdot \frac{\partial}{\partial V_{hj}} \left(\frac{\exp\left(\sum_{j=1}^{H} z_{nj} v_{hj}\right)}{\sum_{k=1}^{K} \exp\left(\sum_{j=1}^{H} z_{nj} v_{kj}\right)} \right)$$

$$= t_{nh}(1 - ydash_{nh})\, z_{nj} \quad -(1)$$

for $i \neq h$

$$\frac{\partial \left(t_{ni} \log(ydash_{ni}) \right)}{\partial V_{hj}} = \partial \left(\frac{e^{\sum_{j=1}^{H} (z_{nj} v_{ij})}}{\sum_{k=1}^{K} e^{\sum_{j=1}^{H} z_{nj} v_{kj}}} \right) \times \frac{t_{ni}}{ydash_{ni}}$$

$$= -\frac{e^{\sum_{j=1}^{H} z_{nj} v_{hj}}}{\sum_{k=1}^{K} e^{\sum_{j=1}^{H} (z_{nj} v_{kj})}} \times z_{nj} \times ydash_{ni}$$

$$= -\frac{t_n \, ydash_{ni}}{ydash_{ni}} \times ydash_{ni} \times z_{nj}$$

$$= -ydash_{nh} \cdot z_{nj} \quad -(2)$$

from ① & ②

$$\frac{\partial E}{\partial v_{hj}} = \sum_{n=1}^{H}\left(t_{nh}\, z_{zj} - \sum_{i=1}^{K} t_{ni}\, ydash_{nh}\, z_{nj}\right)$$

$$\therefore \Delta v_{hj} = \sum_{n=1}^{N} z_{nj}\left(t_{nh} - ydash_{nh}\right)$$

$$\Rightarrow v = v + \eta * \Delta v_{hj}$$

$$v = v + \eta \sum_{n=1}^{N}\left(t_{nh} - ydash_{nh}\right) z_{nj}$$

Similarly,

$$\Delta w_{hj} = -\frac{\partial}{\partial w_j}\left[-\sum_{n=1}^{N}\sum_{k=1}^{K} t_{ni}\log\left(ydash_{ni}\right)\right]$$

$$= \frac{\partial\, t_{ni}\log\left(ydash_{ni}\right)}{\partial w_{hj}}$$

$$= \frac{t_{ni}}{ydash_{ni}}\frac{\partial}{\partial z_{nh}}\left(\frac{e^{\sum_{j=1}^{H}(z_{nj} v_{ij})}}{\sum_{k=1}^{K} e^{\sum_{j=1}^{H} z_{nj} v_{kj}}}\right)\frac{\partial z_{nh}}{\partial w_{hj}}$$

$$\frac{t_{ni}}{ydash_{ni}}\left(v_{ih}\, ydash_{ni} - \frac{ydash_{ni}\left(\sum_{k=1}^{K}\left[\exp\sum_{j=1}^{H}(z_{nj} v_{kj})\right]\right)}{\sum_{k=1}^{K}\exp\sum_{j=1}^{H}\left(z_{nj} v_{kj}\right)}\right) * \frac{\partial z_{nh}}{\partial w_{hj}}$$

classmate

Date

Page

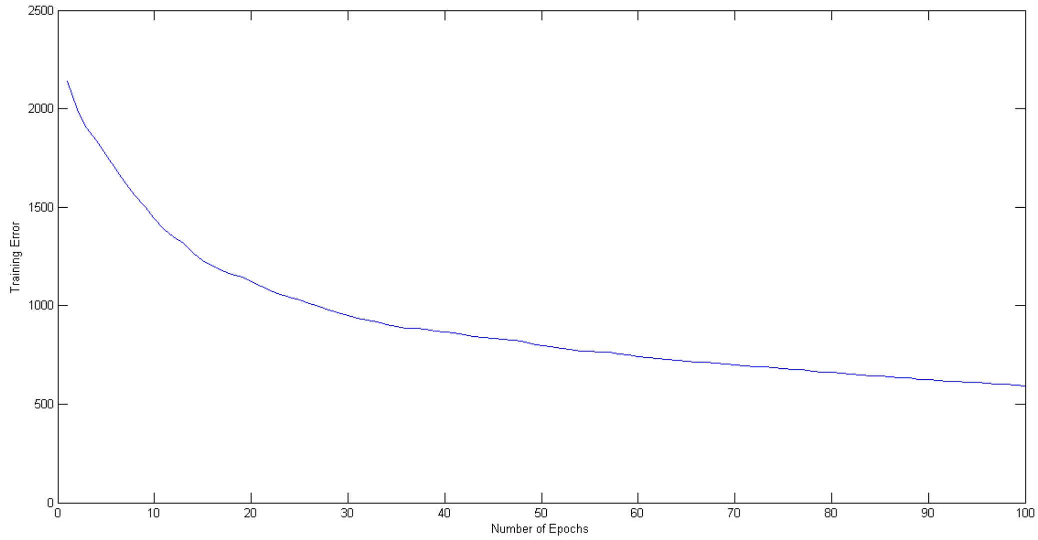
$$= \frac{t_{ni}}{ydash_{ni}} \times \left(v_{ih}\, ydash - ydash_{ni} \left(\sum_{k=1}^{K} v_{kh} * ydash_{nk} \right) \right) * (1 - z_h^2) * x_{nj}$$

$$\frac{\partial E}{\partial w_{hj}} = \sum_{n=1}^{N} \sum_{i=1}^{K} \left(t_{ni}\, v_{ih} - t_{ni} \sum_{k=1}^{K} \left(v_{kh} \cdot ydash \right) \right) * (1 - z_{nh}) * (x_{hj})$$

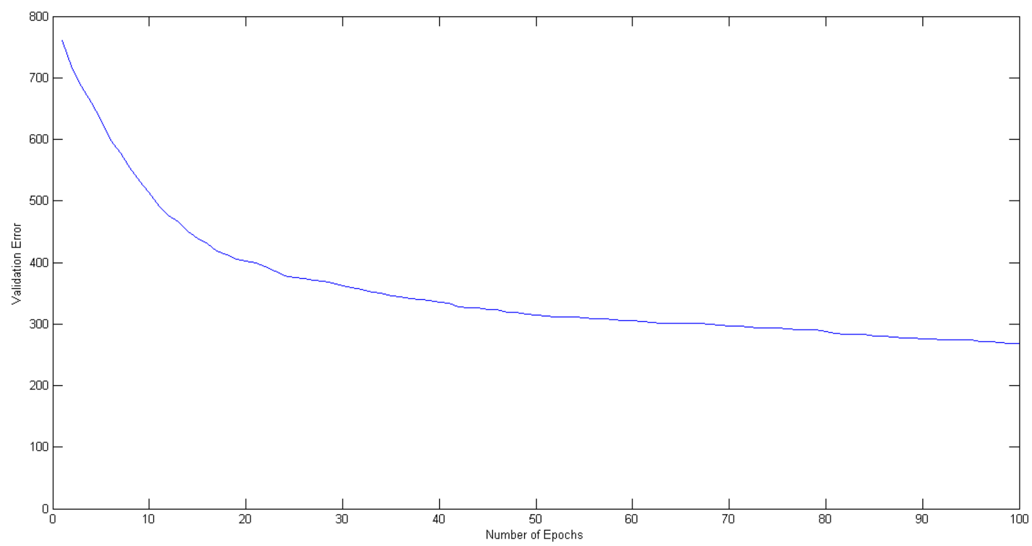
$$\Delta w_{hj} = \sum_{n=1}^{K} \left[\sum_{i=1}^{K} \left((t_{ni} - ydash_{ni}) \cdot v_{ih} \right) * (1 - z_{nh}^2)(x_{nj}) \right]$$

$$w_{hj} = w_{hj} + \eta\, \Delta w_{hj}$$

Plot of Training error v/s number of epochs.



Plot of Validation error v/s number of epochs.



Plot for 2 instances for each of the 10 digits that are misclassied from the test set.