CSL407 Machine Learning

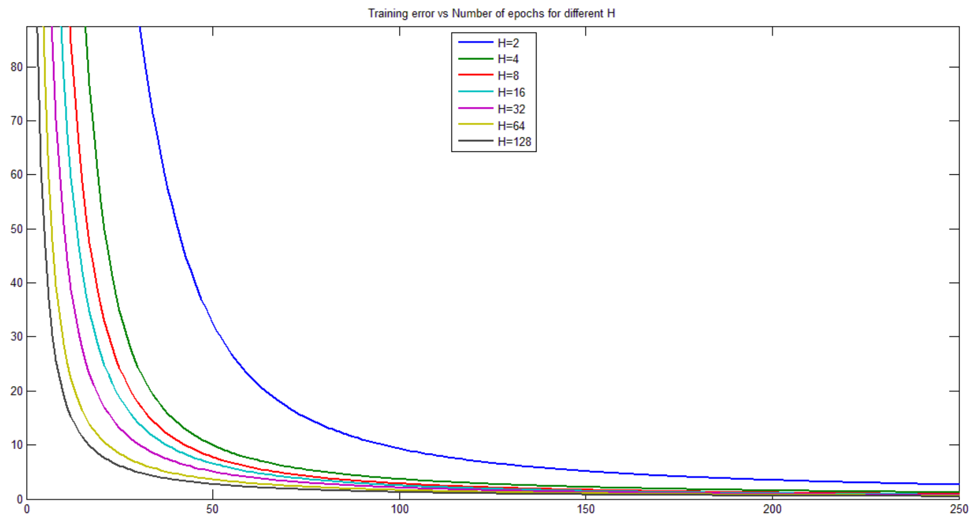
Homework3

Jaskaran Singh(2011cs1012)

Ans1

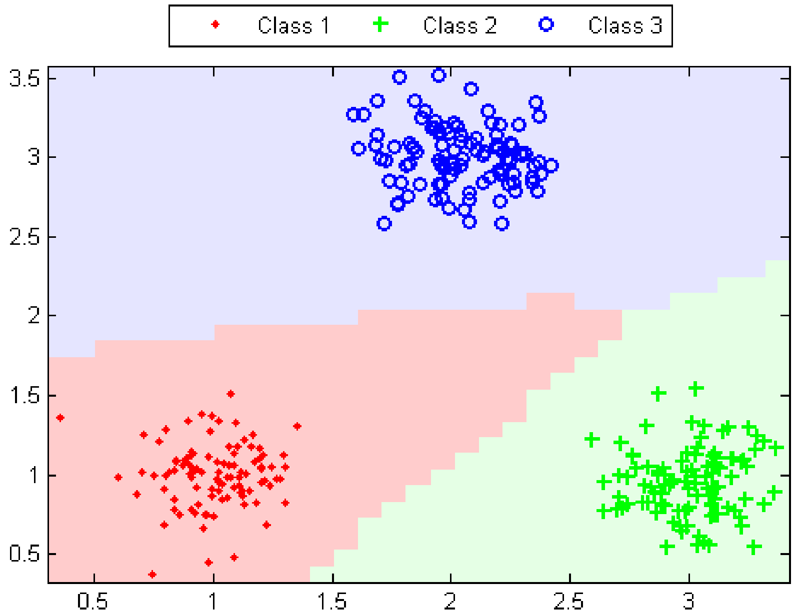
d)

Training error vs Number of Epochs for different number of hidden layer units

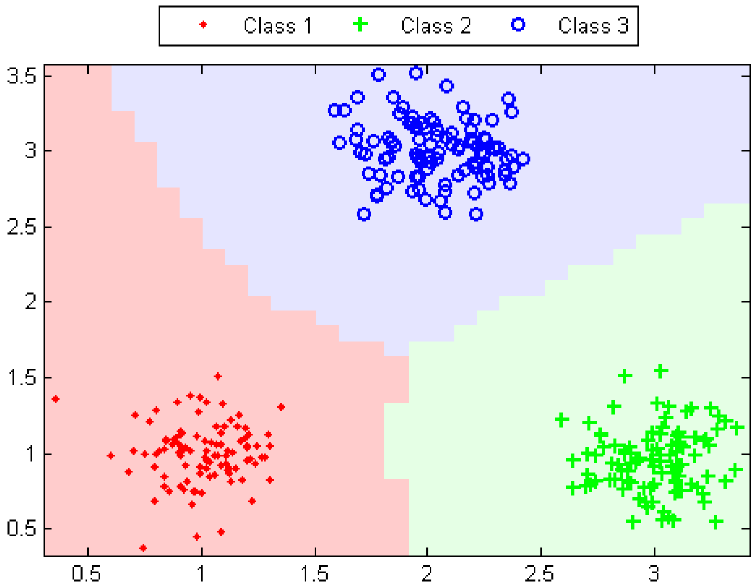


Conclusion: More the number of hidden layer units, lower the training error for a fixed number of epochs.

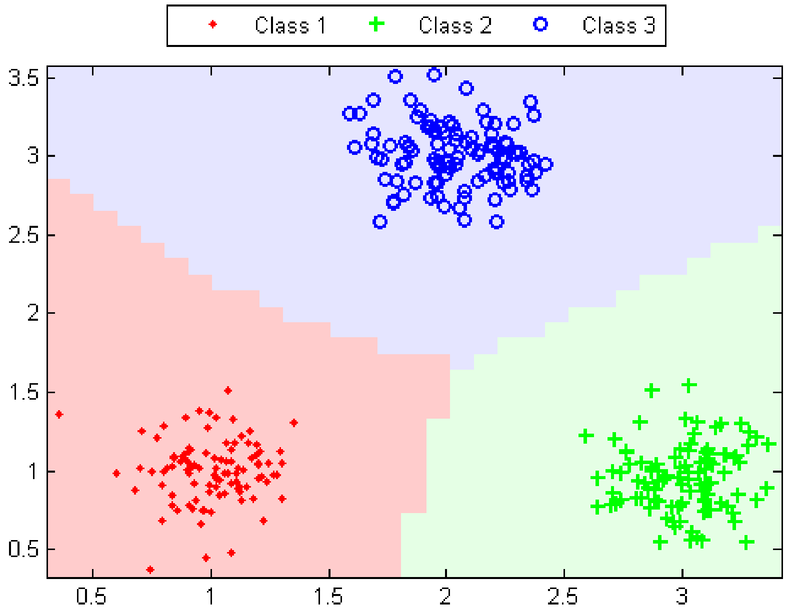
Decision boundary for H=2



Decision boundary for H=4



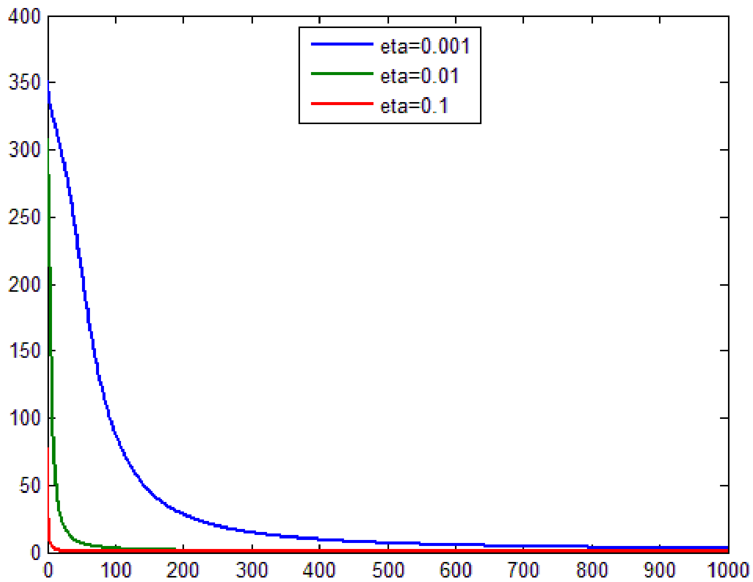
Decision boundary for H=64



Conclusion: More the number of hidden layer units , more the overfitting.Thus ,tighter the decision boundary around a particular training class. For H=64, the decision boundary of the red class is much tighter as compared to the boundary for H=2.

e)

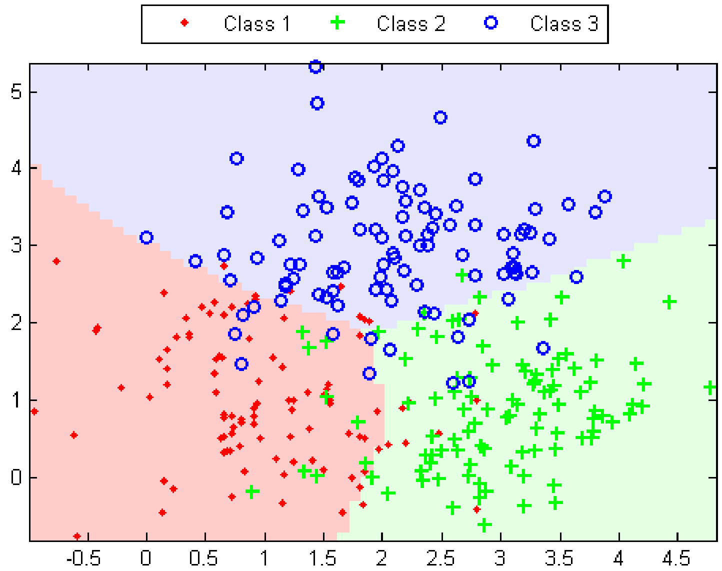
Training error Vs Number of Epochs for different values of learning rates



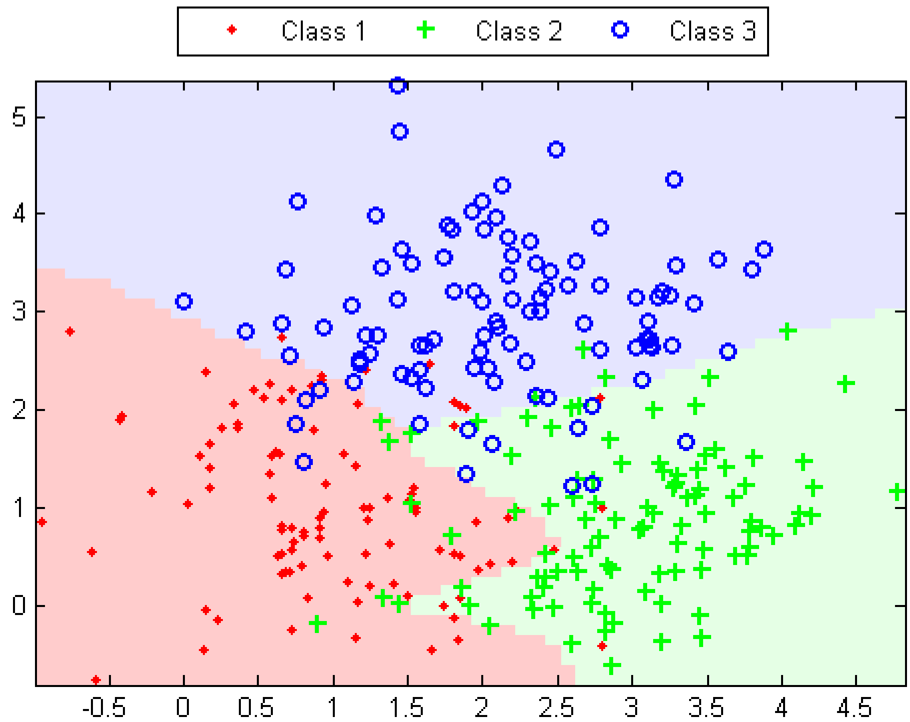
Conclusion: Greater the learning rate,faster the convergence. With eta=0.1 , error becomes almost 0 at the start itself, whereas with eta=0.001 , error becomes 0 after 800 epochs.

f)

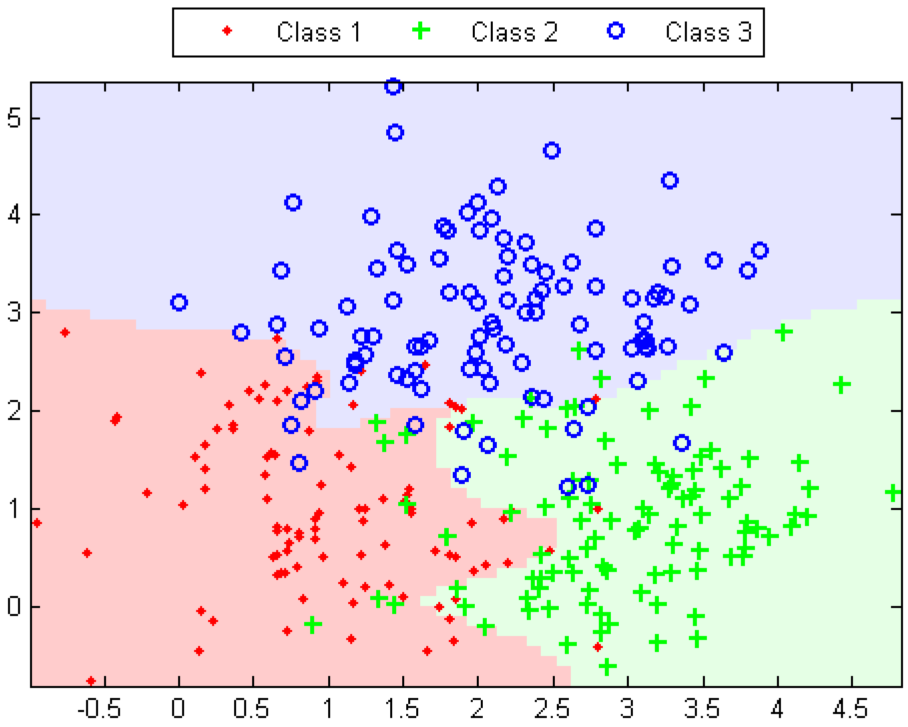
Training error at 1000 epochs=108.3275



Training error at 5000 epochs=91.6284



Training error at 10000 epochs=86.3546



Conclusion: As the number of epochs increases the decision boundaries overfits the training examples which shows that the complexity of the hypothesis learnt increases.

Ans2

Epoch 1

Forward propagation for training example 1:

c_1=sigmoid(0.1*1+0.1*1+0.1*0=0.2)=0.5498

y_1=d_1=sigmoid(0.1*1+0.1*.5498=.15498)=0.5387

Forward propagation for training example 2:

c_2=sigmoid(0.1*1+0.1*0+0.1*1=0.2)=0.5498

y_2=d_2=sigmoid(0.1*1+0.1*0.5498=.15498)=0.5387

Error=$\frac{1}{2*(2)} * \sum_{i=1}^{2}(t_i - y_i)^2$

vd0_new=vd0_old+delta_vd0=0.1+(0.3*(1-0.5387)*0.5387*(1-0.5387)*1+0.3*(0-0.5387)* 0.5387*(1-0.5387)*1)/2=0.0971

vdc_new= vdc_old+delta_vdc=0.1+(0.3*(1-0.5387)*0.5387*(1-0.5387)*0.5498+0.3*(0-0.5387)* 0.5387*(1-0.5387)* 0.5498)/2=0.0984

wc0_new= wc0_old+delta_wc0=

0.1+0.3*((1-0.5387)*0.5387*(1-0.5387)*0.1*0.5498*(1-0.5498)*1 + (0-0.5387)*0.5387*(1-0.5387) *0.1*0.5498*(1-0.5498)*1)/2=0.0999

wca_new=0.1+0.3*((1-0.5387)*0.5387*(1-0.5387)*0.1*0.5498*(1-0.5498)*1+0)/2=0.1004

wcb_new=0.1+0.3*(0+(0-0.5387)*0.5387*(1-0.5387)*0.1*0.5498*(1-0.5498)*1)/2=0.0995

Epoch2

c_1=sigmoid(0.0999*1+0.1004*1+0.0995*0)=0.5499

y_1=d_1=sigmoid(0.0971*1+0.0984*.5499)=0.5377

c_2= sigmoid(0.0999*1+0.1004*0+0.0995*1)=0.5497

y_2=d_2=sigmoid(0.0971*1+0.0984*.5497)=0.5377

vd0_new=vd0_old+delta_vd0 + alpha*delta_vd0_old=0.0971+(0.3*(1-0.5377)*0.5377*(1-0.5377)*1+0.3*(0-0.5377)* 0.5377*(1-0.5377)*1)/2 + 0.9*(-.0029)=**0.0940**

vdc_new= vdc_old+delta_vdc + alpha*delta_vd0_old =0.0984+(0.3*(1-0.5377)*0.5377*(1-0.5377)*0.5499+0.3*(0-0.5377)* 0.5377*(1-0.5377)* 0.5497)/2 + 0.9*(-.0016)=**0.0954**

wc0_new= wc0_old+delta_wc0+alpha*delta_wc0_old=

0.0999+0.3*((1-0.5377)*0.5377*(1-0.5377)*0.0984*0.5499*(1-0.5499)*1 + (0-0.5377)*0.5377*(1-0.5377) *0.0984*0.5497*(1-0.5497)*1)/2 + 0.9*(-0.0001) = **0.0999**

wca_new=0.1004+0.3*((1-0.5377)*0.5377*(1-0.5377)*0.0984*0.5499*(1-0.5499)*1+0)/2 + 0.9*(0.0004) = **0.1011**

wcb_new=0.0995+0.3*(0+(0-0.5387)*0.5377*(1-0.5377)*0.0984*0.5497*(1-0.5497)*1)/2 + 0.9*(-0.9005)= **-0.7114**

Ans3

R=A^~B

A	B	R
0	0	0
0	1	0
1	0	1
1	1	0

Points belonging to class R=0 are {(0,0) (0,1) (1,1)}

Points belonging to class R=1 are (1,0)

Both the classes are linearly separable.

The line A-B-0.5=0 separates both the classes.

Since points belonging to class 0 when substituted in the equation give –ve value

And point (1,0) gives positive value.

Thus, the weights of the perceptron are:

W_{y0}=-0.5

W_{yA}=1

W_{yB}=-1

Ans4

<u>Derivation</u>

$E(w,v) = -\sum_{i=1}^{K} t_i \log(y_i)$ (Cross Entropy error function)

$y_i = \frac{e^{(zv_i{}^T)}}{\sum_{j=1}^{K} e^{(zv_j{}^T)}}$ (Softmax function)

$$z_h = \tanh\,(\sum_{j=0}^{D} x_j w_{hj})$$

$$\frac{\partial E}{\partial v_{ih}} = \sum_{j=1}^{K} \frac{\partial E}{\partial y_j} * \frac{\partial y_j}{\partial v_{ih}} = \frac{\partial E}{\partial y_i} * \frac{\partial y_i}{\partial v_{ih}} + \sum_{j=1, j\neq i}^{K} \frac{\partial E}{\partial y_j} * \frac{\partial y_j}{\partial v_{ih}}$$

1) $\frac{\partial y_i}{\partial v_{ih}} = y_i(1-y_i)z_h$

2) $\frac{\partial E}{\partial y_i} = -\frac{t_i}{y_i}$

3) $\frac{\partial \mathrm{E}}{\partial \mathrm{y_i}} * \frac{\partial \mathrm{y_i}}{\partial \mathrm{v_{ih}}} = -\mathrm{t}_i(1-y_i)z_h$

4) $\frac{\partial y_j}{\partial v_{ih}} = -y_j y_i z_h$

5) $\frac{\partial \mathrm{E}}{\partial \mathrm{y_j}} * \frac{\partial \mathrm{y_j}}{\partial \mathrm{v_{ih}}} = \mathrm{t}_j y_i z_h$

$$\frac{\partial E}{\partial v_{ih}} = -\mathrm{t}_i(1-y_i)z_h + \sum_{j=1,j\neq i}^{K} \mathrm{t}_j y_i z_h = -\mathrm{t}_i(1-y_i)z_h + y_i z_h(1-t_i)$$

$$= -z_h(t_i - y_i)\ \{using\ equations\ 1{,}2{,}3{,}4{,}5\ and\ \sum_{i=1}^{K} t_i = 1\}$$

(Weight update equation for v_{ih})

$$\boldsymbol{v_{ih} = v_{ih} - eta * \frac{\partial E}{\partial v_{ih}} = v_{ih} + eta * (t_i - y_i)z_h}$$

$$\frac{\partial E}{\partial w_{hj}} = \sum_{i=1}^{K} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial z_h}\frac{\partial z_h}{\partial w_{hj}}$$

6) $\frac{\partial y_i}{\partial z_h} = y_i v_{ih} - y_i(\sum_{j=1}^{K} y_j v_{jh})$

7) $\frac{\partial z_h}{\partial w_{hj}} = (1 - z_h^2)x_j$

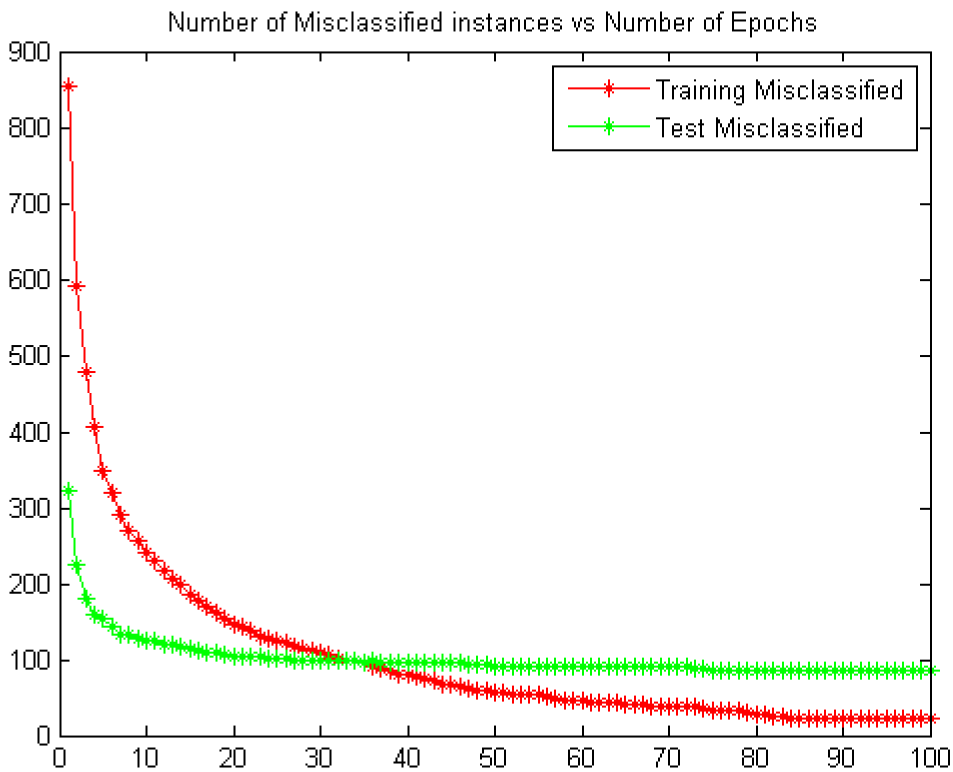
$$\frac{\partial E}{\partial w_{hj}} = \sum_{i=1}^{K} \frac{\partial E}{\partial y_i}\frac{\partial y_i}{\partial z_h}\frac{\partial z_h}{\partial w_{hj}} = \frac{\partial E}{\partial w_{hj}} = \sum_{i=1}^{K} \frac{-t_i}{y_i} * (y_i v_{ih} - y_i(\sum_{j=1}^{K} y_j v_{jh})) * (1 - z_h^2)x_j$$

$$= (1 - z_h^2)x_j(-\sum_{i=1}^{K} t_i v_{ih} + \sum_{j=1}^{K} y_j v_{jh} * \sum_{i=1}^{K} t_i) = (1 - z_h^2)x_j(\sum_{i=1}^{K}(-t_i v_{ih} + y_i v_{ih}))$$

$$= -(1 - z_h^2)x_j(\sum_{i=1}^{K}(t_i - y_i)v_{ih})\ \ \{using\ equations\ 2{,}6{,}7\ and\ \sum_{i=1}^{K} t_i = 1\}$$

(Weight update equation for w_{hj})

$$\boldsymbol{w_{hj} = w_{hj} - eta * \frac{\partial E}{\partial w_{hj}} = w_{hj} + eta * (1 - z_h^2)x_j(\sum_{i=1}^{K}(t_i - y_i)v_{ih})}$$

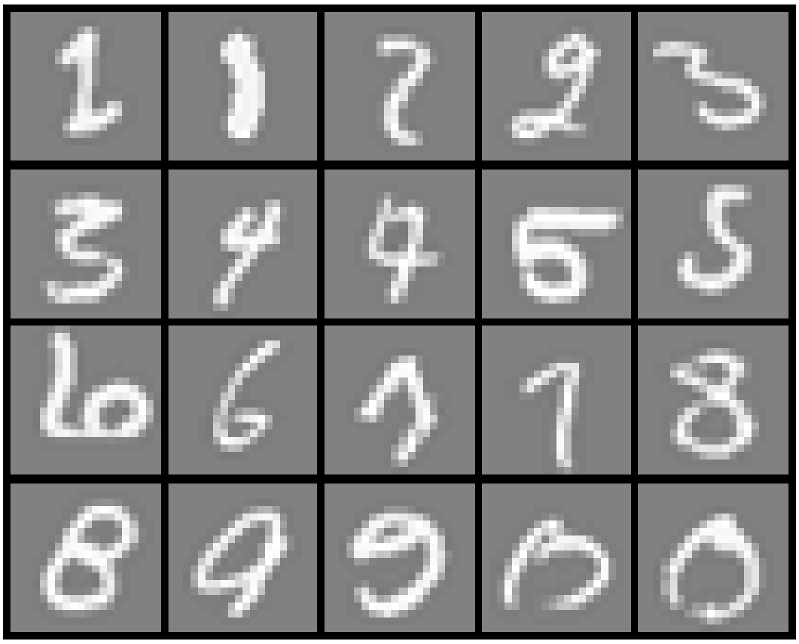


Misclassified examples out of 2990 training examples=22

Misclassified examples out of 1010 validation examples=86

Misclassified examples out of 1000 test examples=71

These are the 20 misclassified examples(row-wise starting from 1st row).



The labels and the prediction for the 20 misclassified examples shown above are :

1. label of example=1
   predicted label of example=2
2. label of example=1
   predicted label of example=7
3. label of example=2
   predicted label of example=8
4. label of example=2
   predicted label of example=3
5. label of example=3
   predicted label of example=6
6. label of example=3
   predicted label of example=5
7. label of example=4
   predicted label of example=9
8. label of example=4
   predicted label of example=7
9. label of example=5
   predicted label of example=8
10. label of example=5
    predicted label of example=2
11. label of example=6
    predicted label of example=2
12. label of example=6

predicted label of example=5

13. label of example=7
    predicted label of example=10
14. label of example=7
    predicted label of example=9
15. label of example=8
    predicted label of example=3
16. label of example=8
    predicted label of example=5
17. label of example=9
    predicted label of example=10
18. label of example=9
    predicted label of example=10
19. label of example=10
    predicted label of example=9
20. label of example=10
    predicted label of example=7