COMP3314/CSIS0314 : Assignment 2

Due on Monday, March 23, 2015

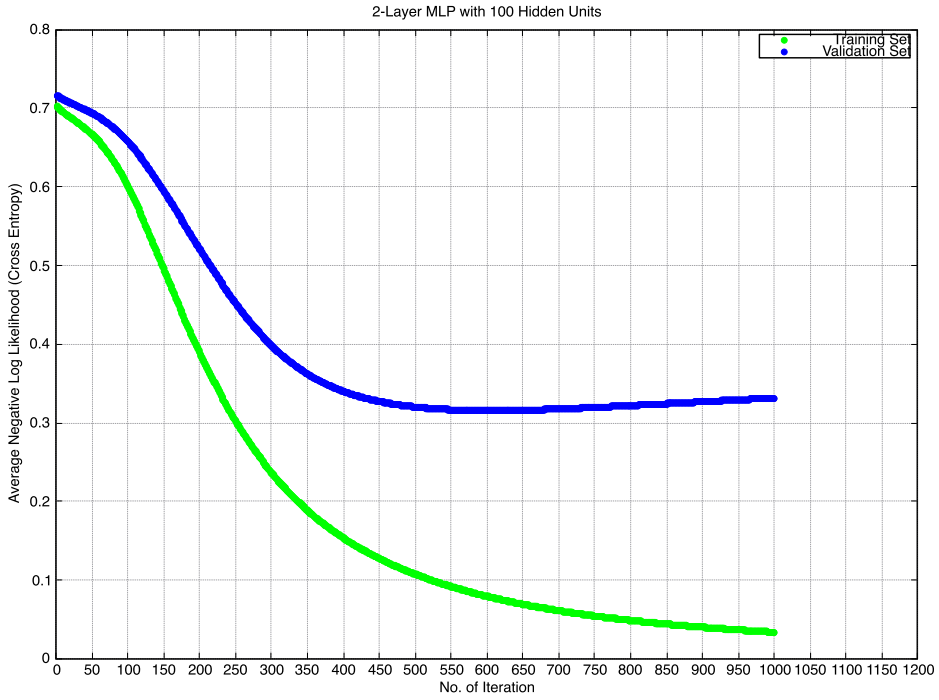
Instructor: Jack Wang

Qian Xin, 3035134147

Neural Network

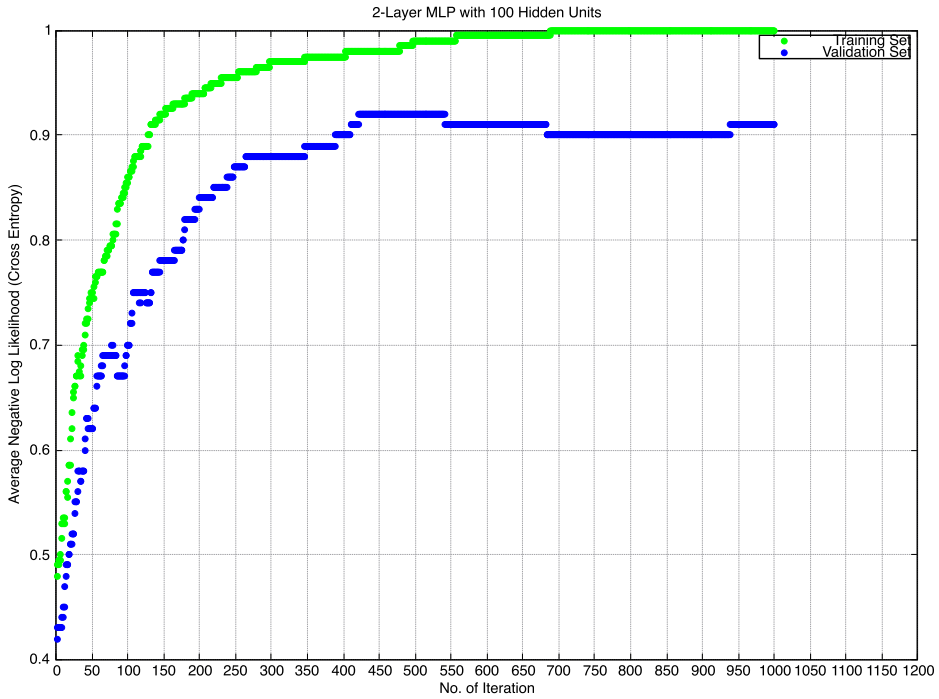
Part 1. Training Error vs. Validation Error

Using the default initialization values, learning rate lambda = 1e-4 and 1000 times gradient descent iterations, the negative log likelihood both starts from 0.70/0.71. On the training set it finally reaches 0.033764, which indicates a very good minimization. The negative log likelihood on the validation set finally reaches 0.331516, which indicates an acceptable minimization.



Part 2. Training Accuracy vs. Validation Accuracy

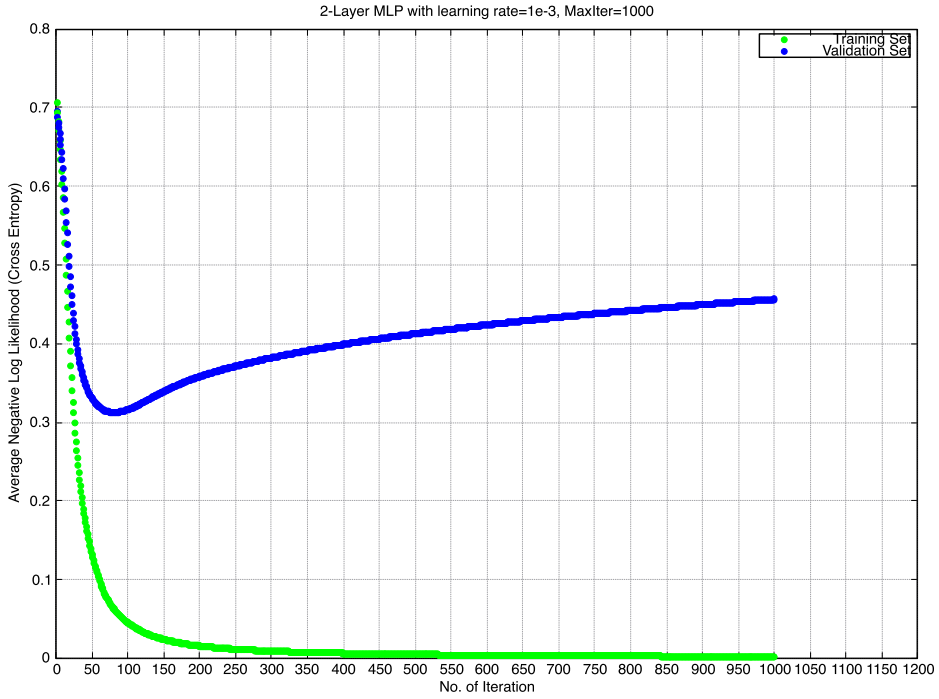
Plotted the classification accuracy as below, the accuracy on the training set reaches 1.00 at around 680 times of iteration, which indicates a good fitwhile the accuracy on the validation set stays around 0.91 since 940 times of iteration.

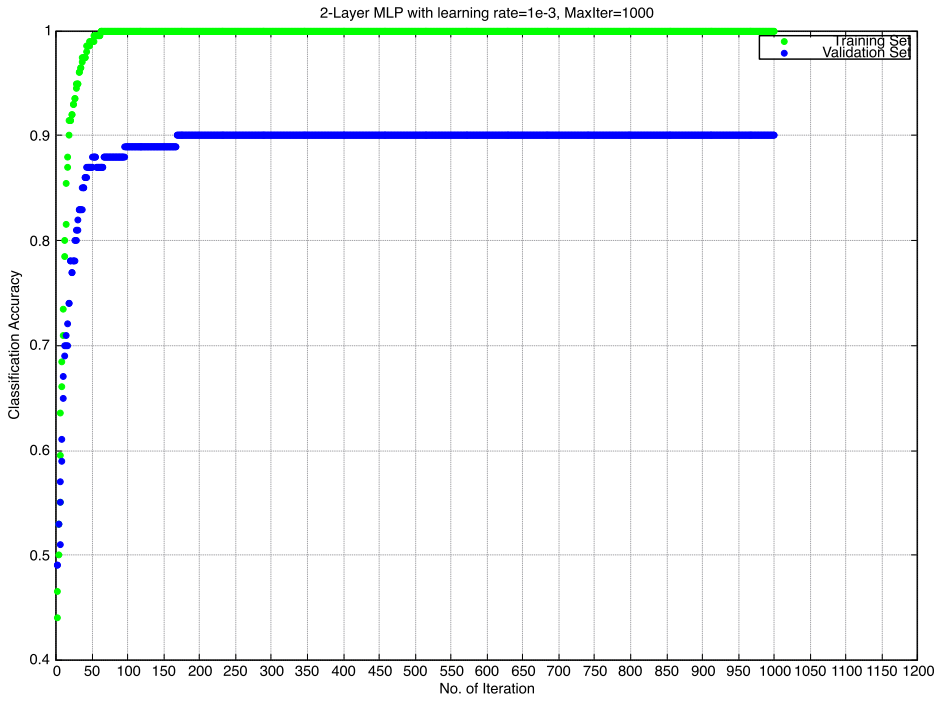


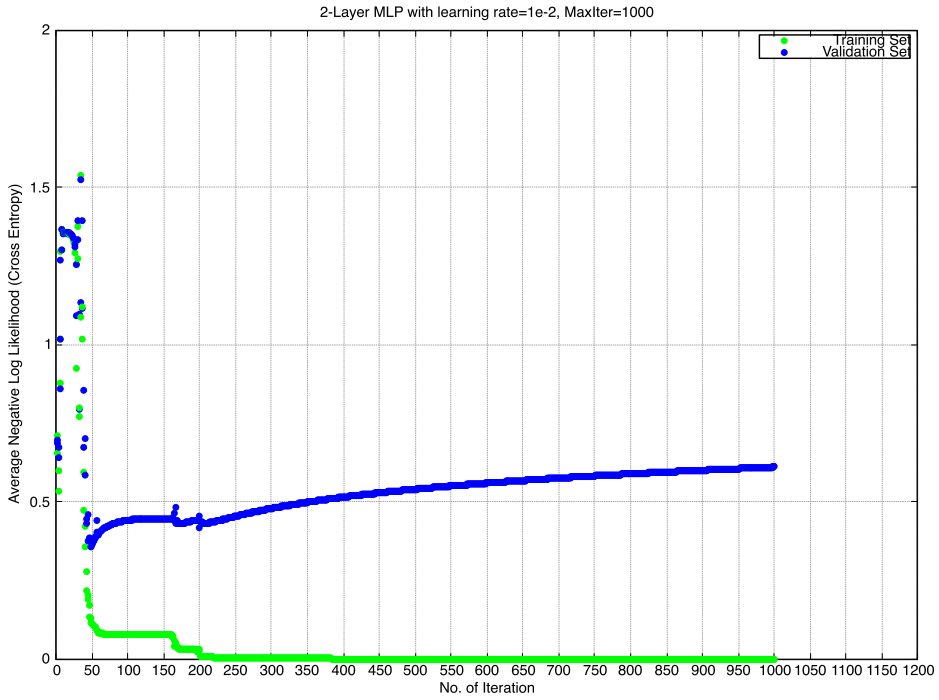
Part 3. Learning Rate

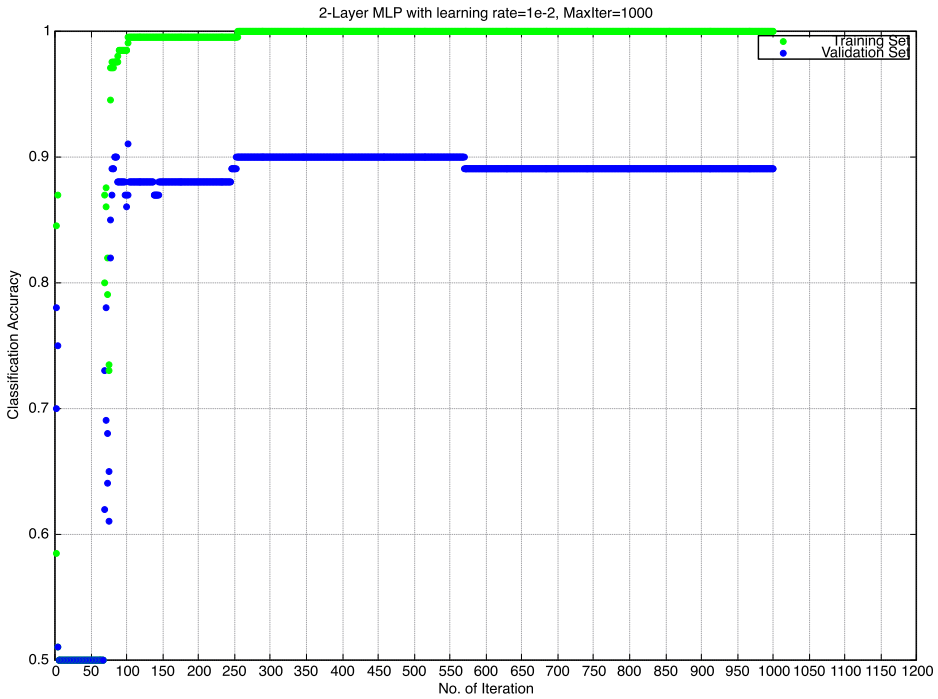
The learning rate adjusts and controls the speed and quality on convergence. If the learning rate is set too large (0.01), then the neural network converges fast but also sacrifices quality. From the below plot, we may observe that (1) the negative log likelihood slightly goes up instead of keeping flat when number of iteration grows, which is a sign of overfitting (see in 1e-3,1e-2 plot); (2) volatility at the beginning of the iteration instead of stable descent (see in 1e-2 plot); (3) slight lower in accuracy compared to small learning rate.

It is more appropriate to set the learning rate to a small value but not too slow value. As long as the learning rate is mall enough to ensure convergence, its value determines only the speed at which the training converges. In this case, I would choose the best value as 1e-3.



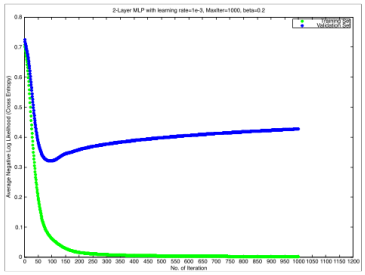
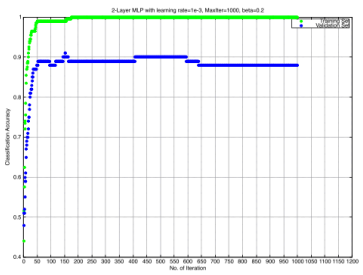
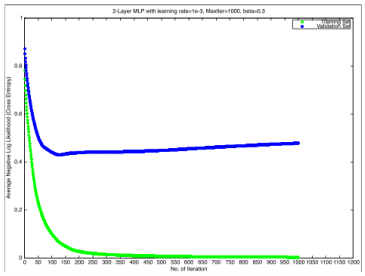
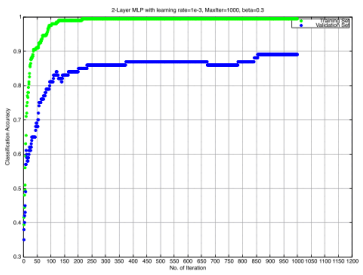
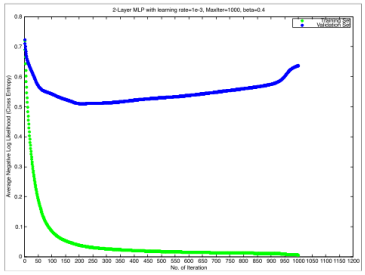
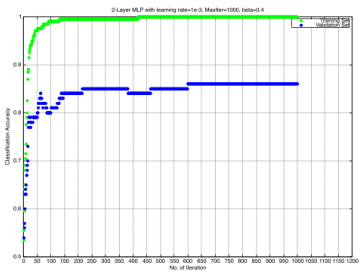
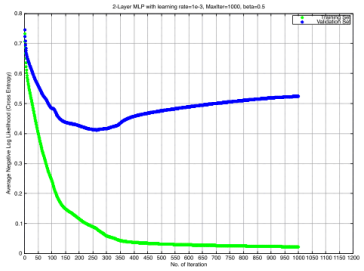
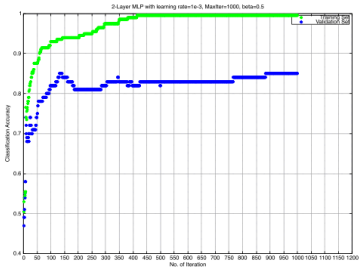

2-Layer MLP with learning rate=1e-3, MaxIter=1000
Training Set
Validation Set
Classification Accuracy
No. of Iteration
1
0.9
0.8
0.7
0.6
0.5
0.4
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200

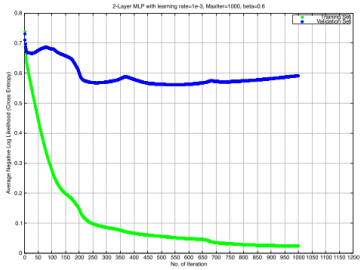

2-Layer MLP with learning rate=1e-2, MaxIter=1000
Training Set
Validation Set
Average Negative Log Likelihood (Cross Entropy)
No. of Iteration
2
1.5
1
0.5
0
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200

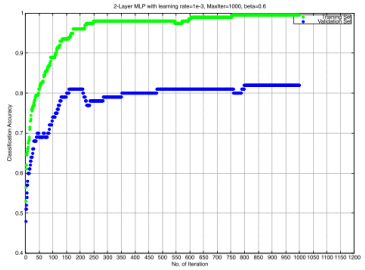

2-Layer MLP with learning rate=1e-2, MaxIter=1000
Training Set
Validation Set
Classification Accuracy
No. of Iteration
1
0.9
0.8
0.7
0.6
0.5
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200

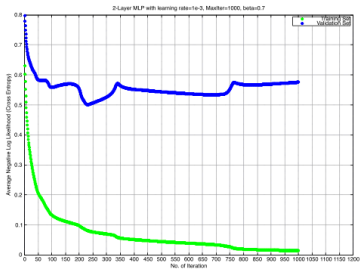
Part 4. Initialization

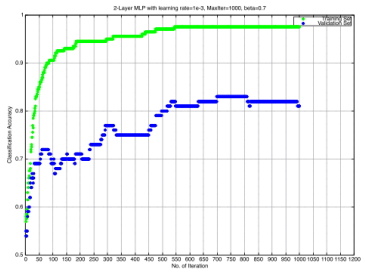
Below are 18 plots of the changes w.r.t. beta parameter from 0.2 to 1 (beta=0.1 omitted), followed with analysis.

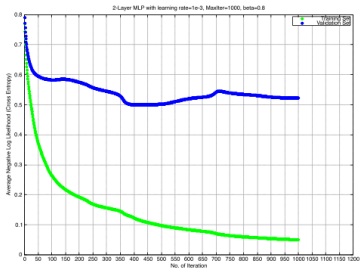









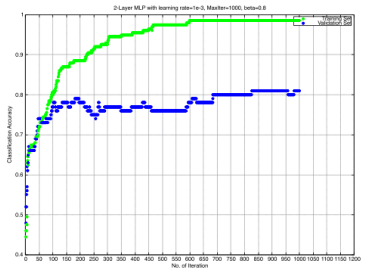

2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.6
Training Set
Validation Set
Average Negative Log Likelihood (Cross Entropy)
No. of Iteration

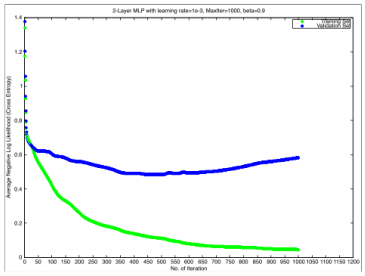

2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.6
Training Set
Validation Set
Classification Accuracy
No. of Iteration

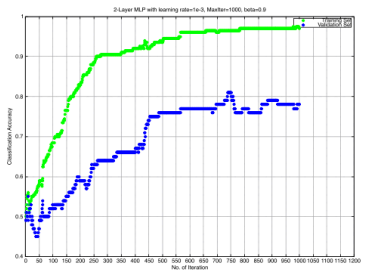

2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.7
Training Set
Validation Set
Average Negative Log Likelihood (Cross Entropy)
No. of Iteration

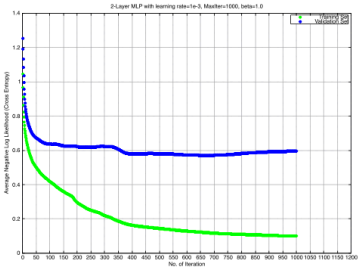

2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.7
Training Set
Validation Set
Classification Accuracy
No. of Iteration

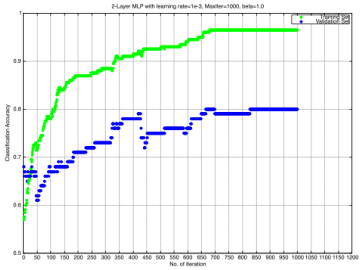

2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.8
Training Set
Validation Set
Average Negative Log Likelihood (Cross Entropy)
No. of Iteration


2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.8
Training Set
Validation Set
Classification Accuracy
No. of Iteration


2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.9
Training Set
Validation Set
Average Negative Log Likelihood (Cross Entropy)
No. of Iteration


2-Layer MLP with learning rate=1e-3, MaxIter=1000, beta=0.9
Training Set
Validation Set
Classification Accuracy
No. of Iteration





Please see analysis in the next page.

With increase in the beta parameter, the negative log likelihood on the training set remains stable. However, the negative log likelihood on the validation set grows more and more volatile and unstable. The minimisation are also performed poorly. The accuracy on the validation set also fluctuate. The final accuracy on both training set and validation set fall (deteriorate).

When we look back into the code, we may notice

```
% Initialize connection weights, beta controls the average magnitude
beta = 0.1;
% Weights and biases for layer 1
W1 = beta*randn(D, M);
```

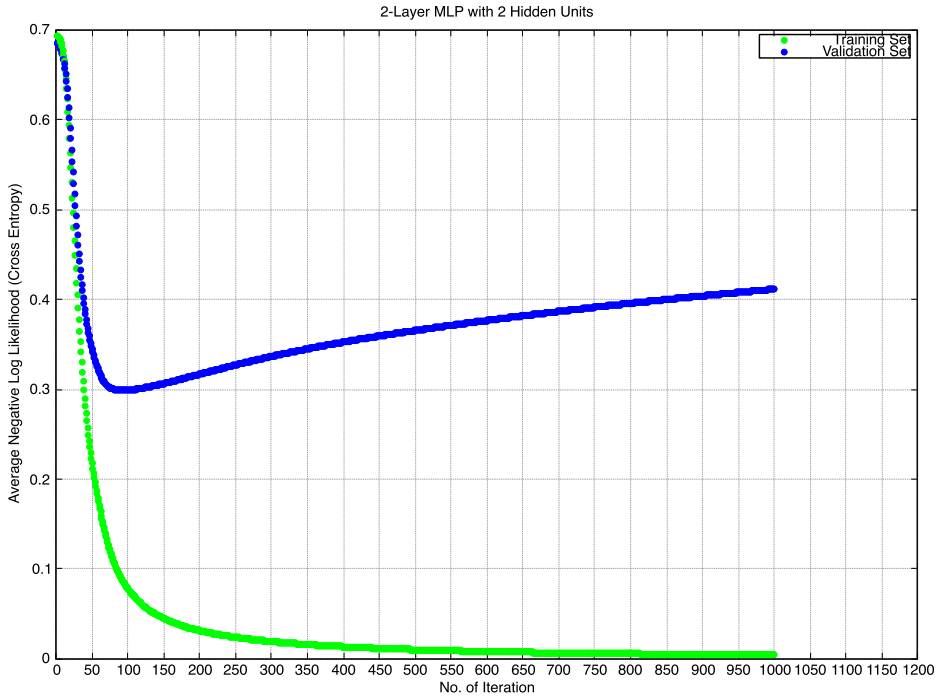
Because beta parameter controls the variance of the connection weights. In this way, it controls the average magnitude of weights. A good idea is to pick values of beta / values of the weights at random following a distribution which helps the optimization process to converge to a meaningful solution. According to my research, it's a good idea to choose initial weights of a neural network from the range (-1/sqrt(d), 1/sqrt(d)) where d is the number of inputs. In our case, d = 28*28 = 784. The best range is (-0.036, 0.036). Here we may choose beta as 0.1.

Part 5. Hidden Units

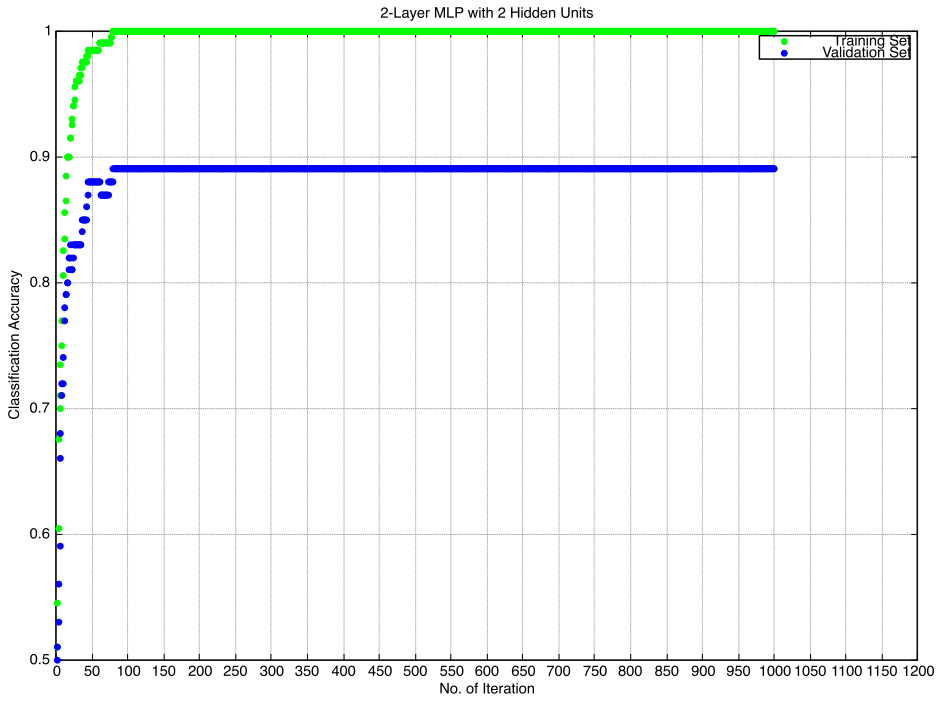
There is not much change made on the curve simply based on hidden units number. However, based on the final accuracy and early stopping method mentioned in the lecture, we may choose large M and stop the learning process when validation error starts increasing.

Observing from validation set accuracy, I would choose hidden units as 100.

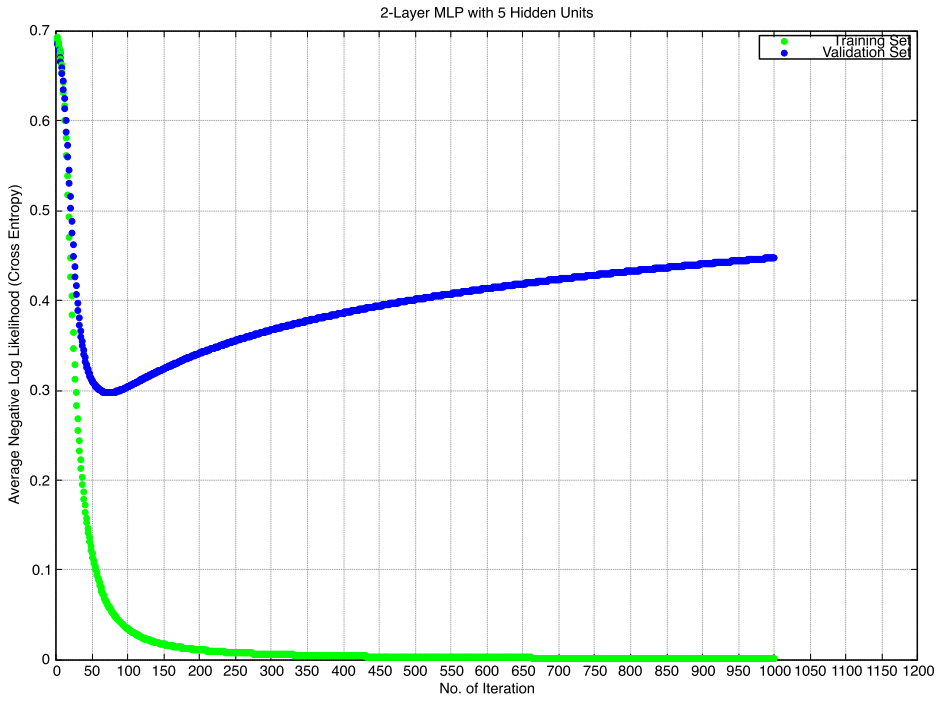
Hidden units = 2, negative log likelihood on validation set = 0.498041



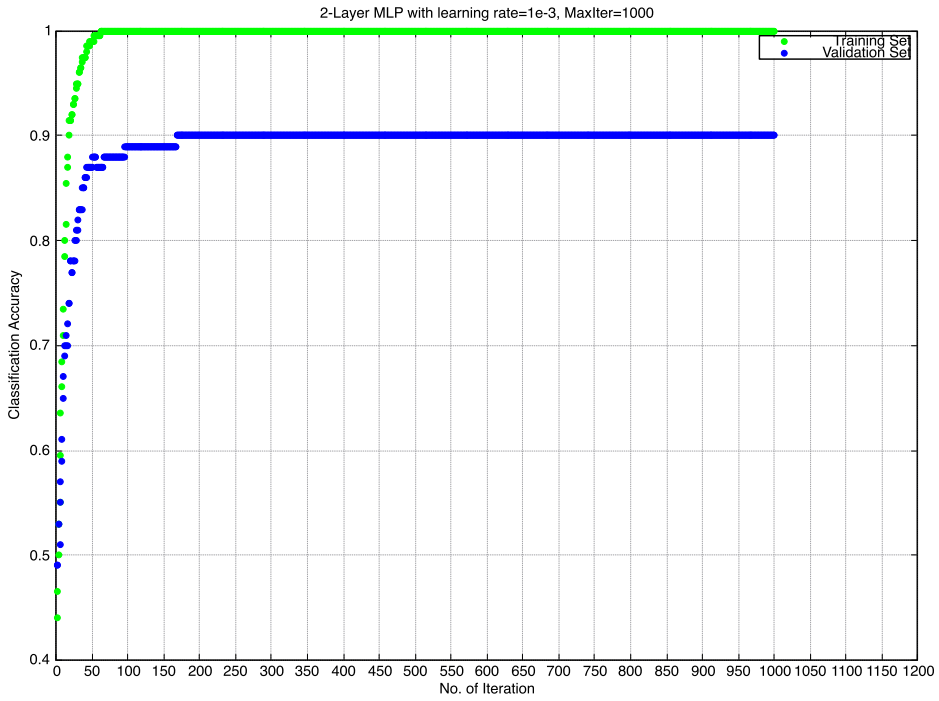
Hidden units = 2, accuracy on validation set = 0.89



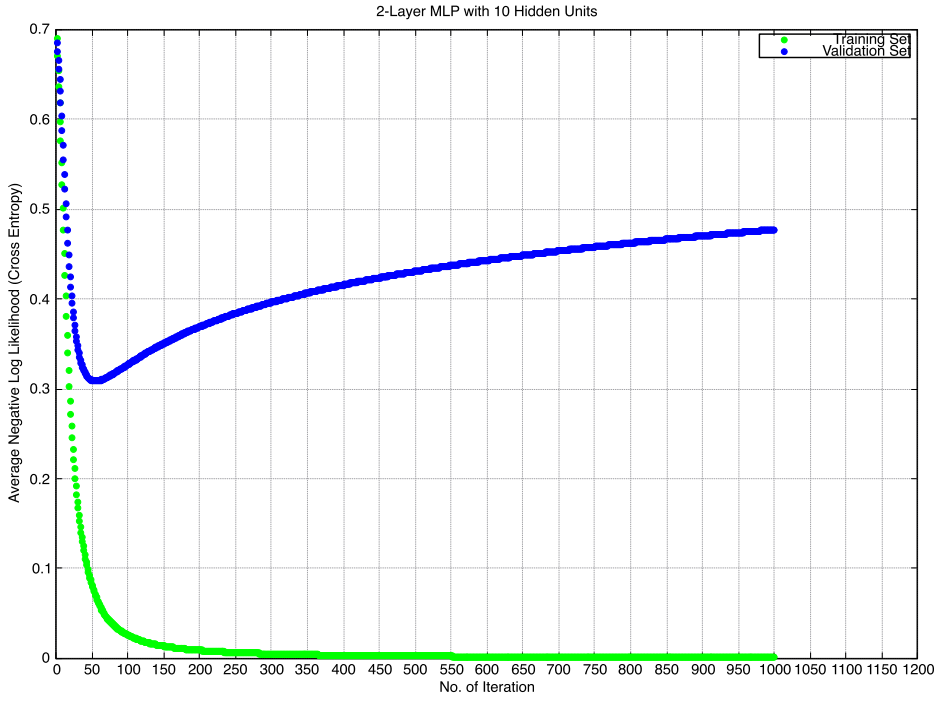
Hidden units = 5 (control group), negative log likelihood = 0.388016



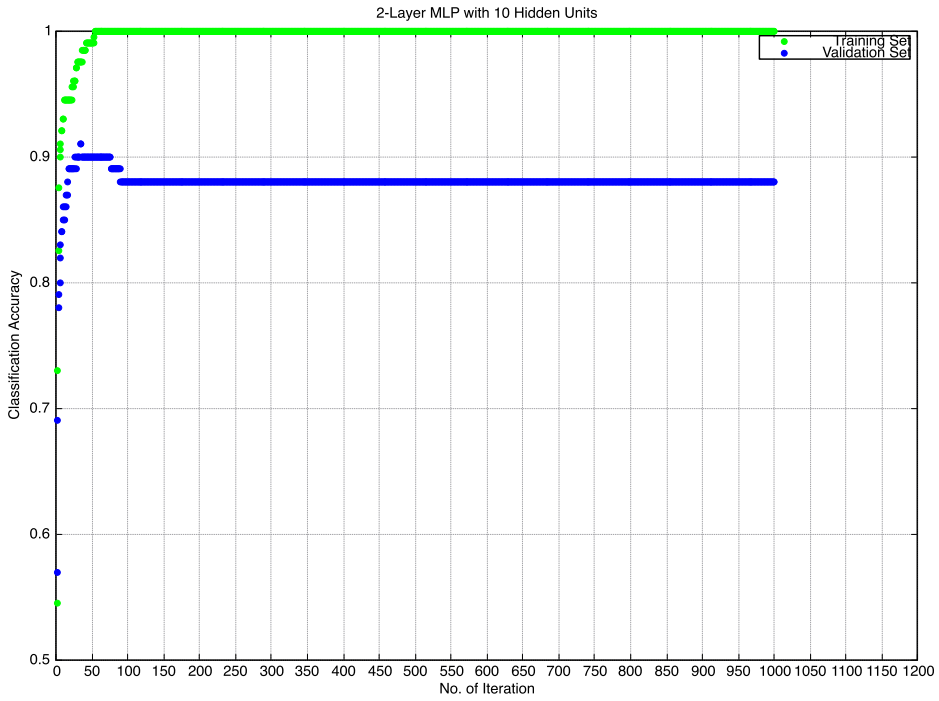
Hidden units = 5 (control group), accuracy on validation set = 0.90



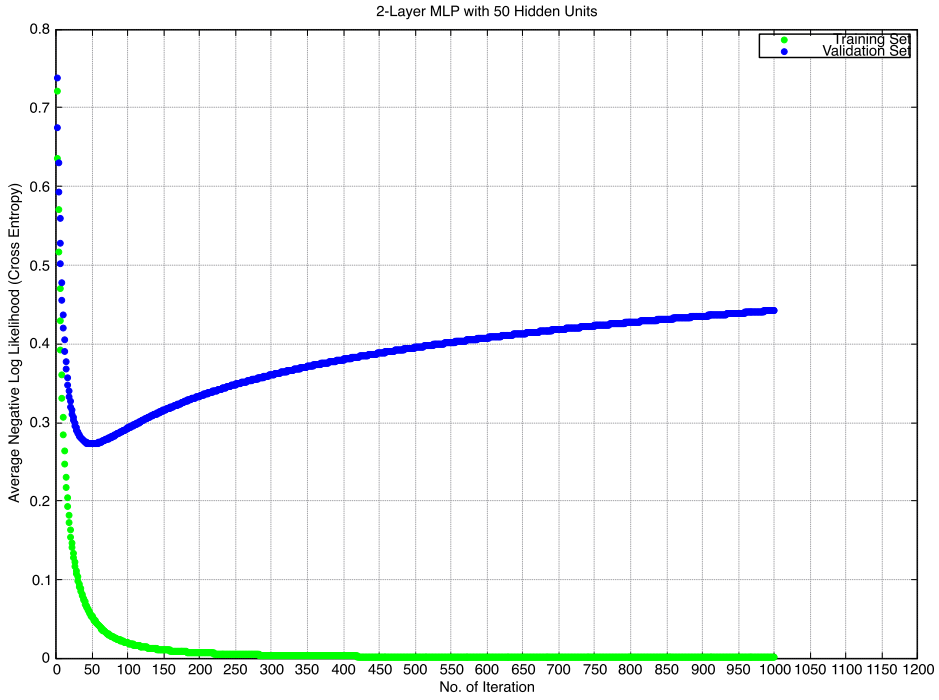
Hidden units = 10, negative log likelihood = 0.47708



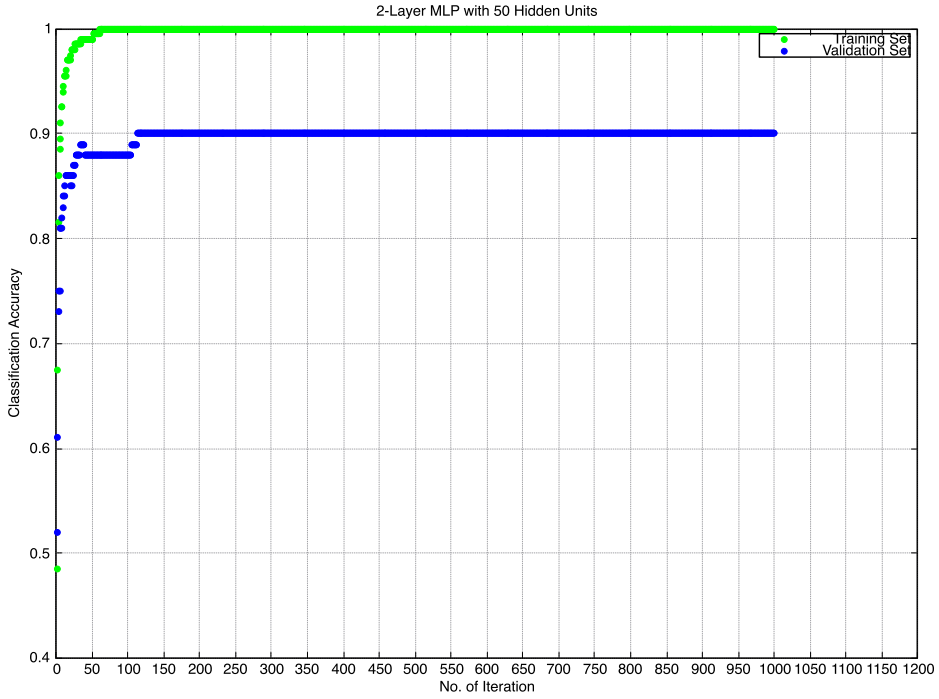
Hidden units = 10, accuracy on validation set = 0.88



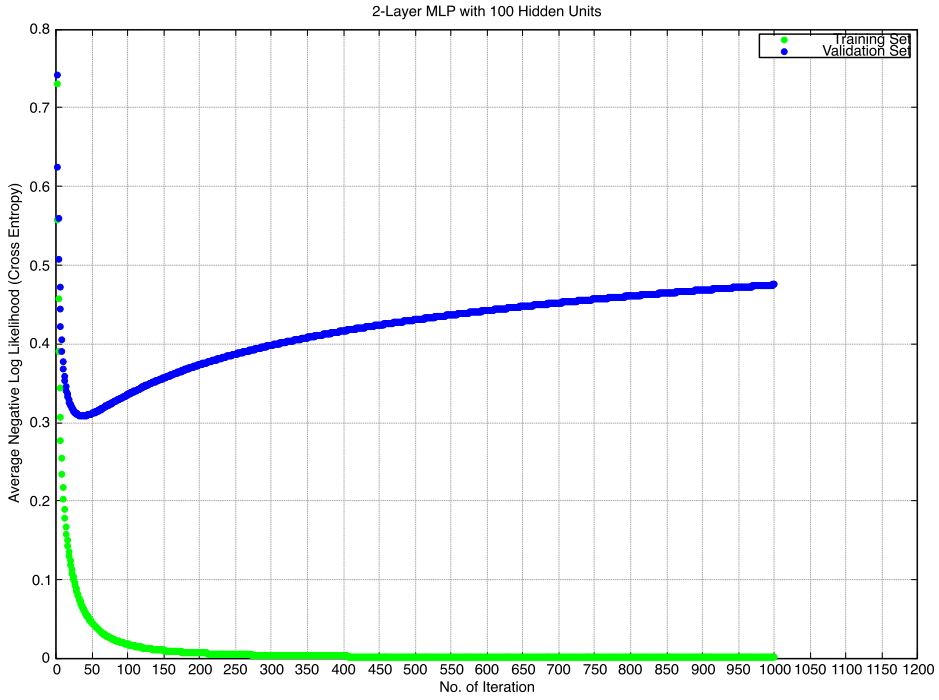
Hidden units = 50, negative log likelihood = 0.44244



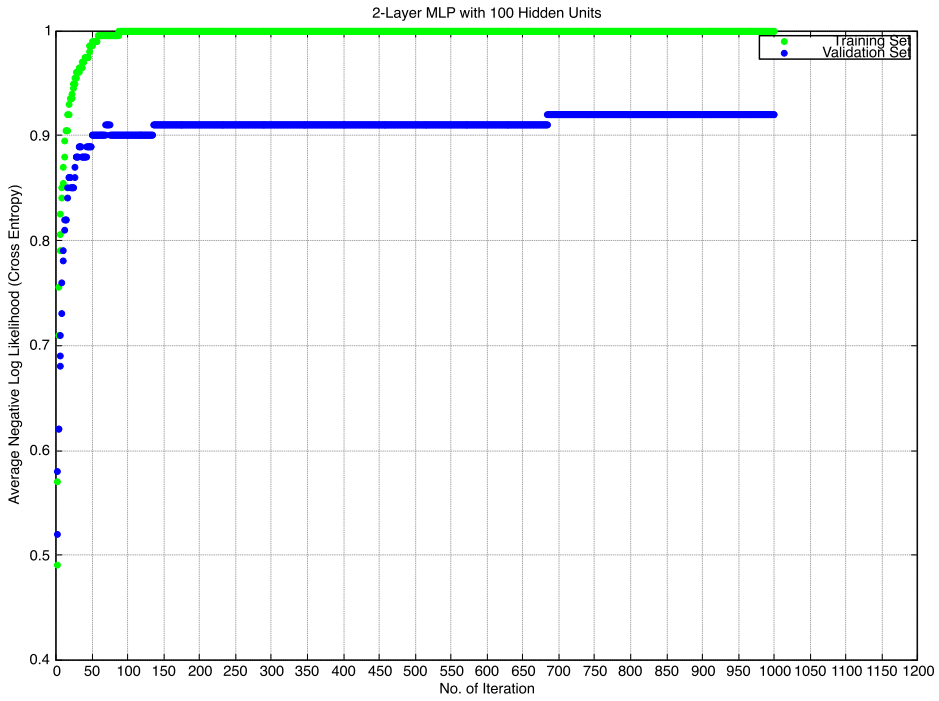
Hidden units = 50, accuracy on validation set = 0.90



Hidden units = 100, negative log likelihood = 0.4751



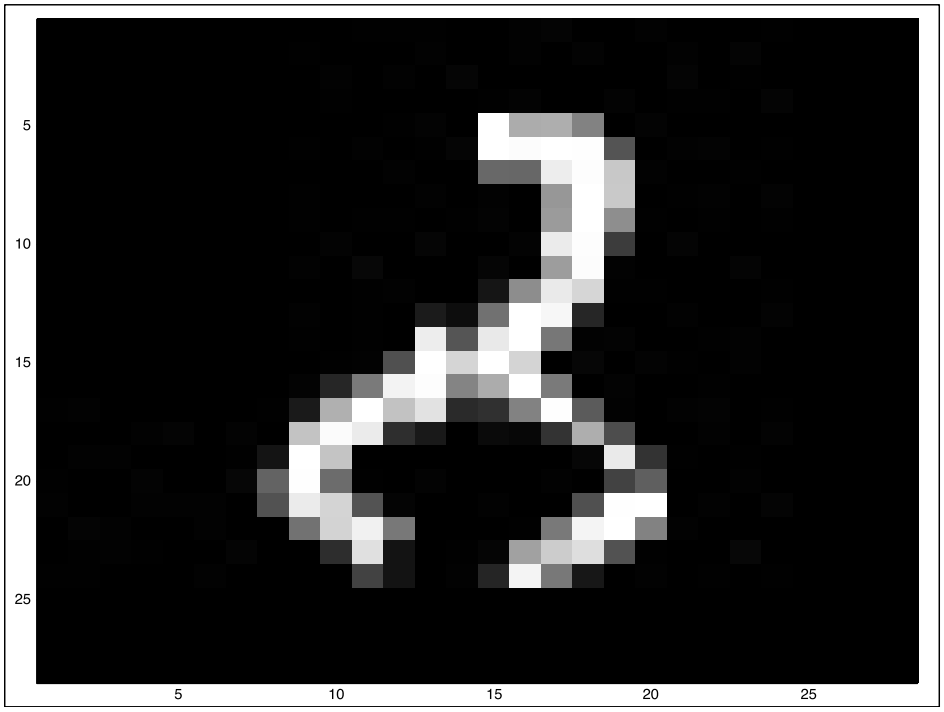
Hidden units = 100, accuracy on validation set = 0.92


2-Layer MLP with 100 Hidden Units
Training Set
Validation Set
Average Negative Log Likelihood (Cross Entropy)
No. of Iteration
1
0.9
0.8
0.7
0.6
0.5
0.4
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000 1050 1100 1150 1200

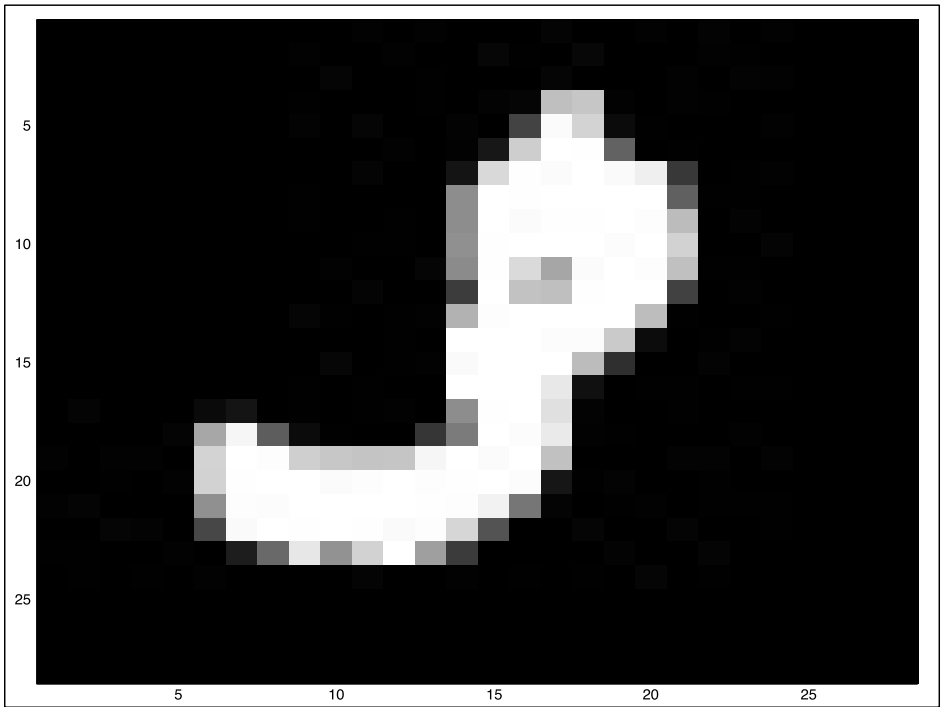
Part 6. Misclassifications

3 misclassified by the network as 8

Example 1: Posterior probability yn = 0.999424

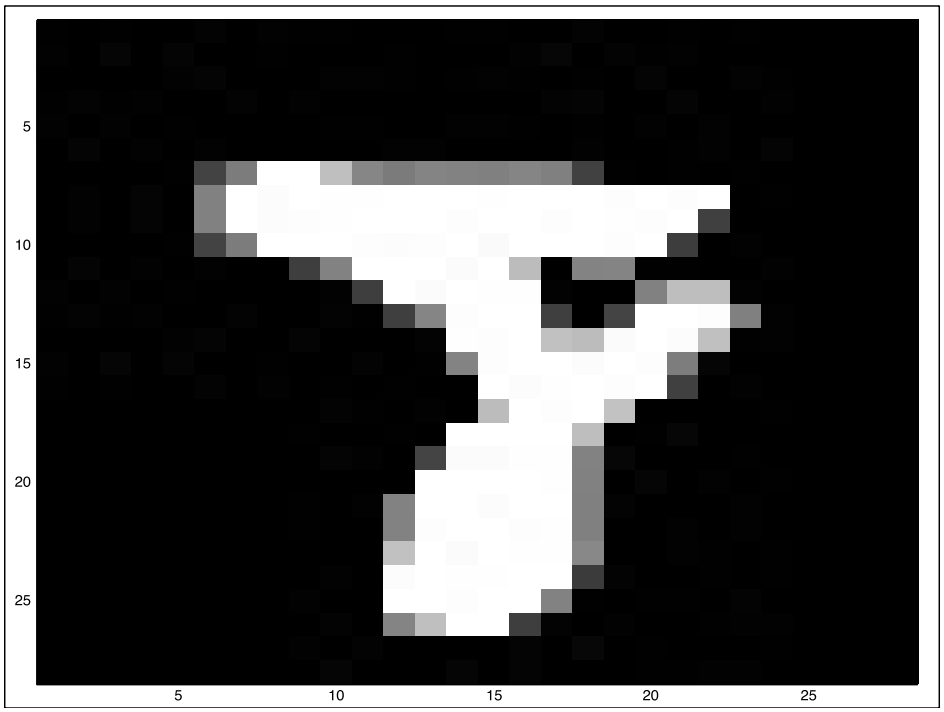


Example 2: Posterior probability yn = 0.994110



8 misclassified by the network as 3

Example 1: Posterior probability yn = 0.033789



Example 2: Posterior probability yn = 0.002483