Homework 3

Bharath Hariharan

September 25, 2012

1   Q1

1. Because the energy function is non-convex, many runs of backpropagation
   don't converge to the right solution. Figure 3 shows a histogram of the
   training error over 75 runs of backpropagation. The training algorithm
   converges to a right solution 20 times out of the 75. Figure 1 shows a
   trained model that classifies the training data perfectly. Figure 2 shows
   the progress of the energy function with the number of iterations for this
   model.

2. Momentum does not seem to help in this case, perhaps because even with-
   out momentum the model converges quickly. Figure 4 shows the conver-
   gence for different values of the momentum: For a momentum value of p,
   the update is taken to be:

$$\Delta w^{(t)} = -(1-p)\eta \nabla_w E + p \Delta w^{(t-1)} \tag{1}$$

   Clearly, varying p doesn't seem to change convergence very much.

2   Q2

A single neuron could not classify the data correctly even over 20 runs of the
training procedure. On the other hand, a two layer network with only two
hidden nodes could classify the data correctly at least some of the times (see
Table 1). However, the probability that a neural network trained from a ran-
dom starting point would manage to classify the data correctly increases as the
number of hidden nodes increases: as such the median and mean number of
errors made goes down.

  Figure 5 shows the learnt "filters" in the hidden layer, for four hidden nodes
for a model that learns to clasify the data correctly. It is hard to make sense of
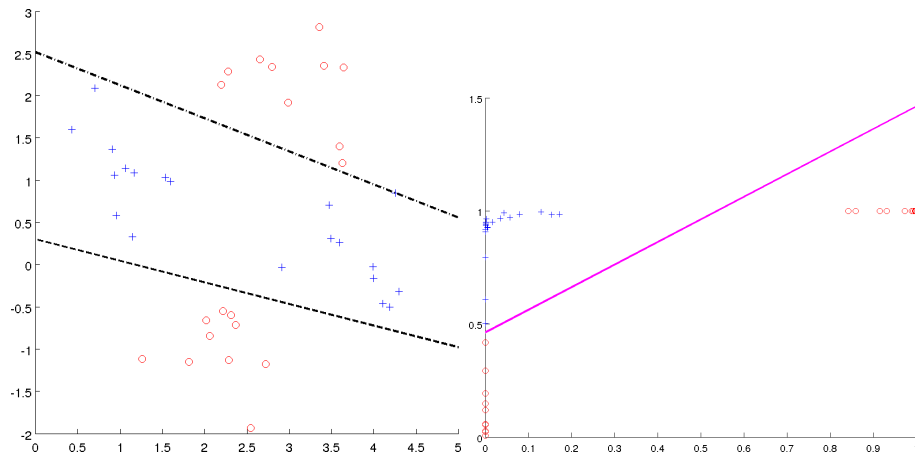them because unlike the training data, they are not binary-valued.

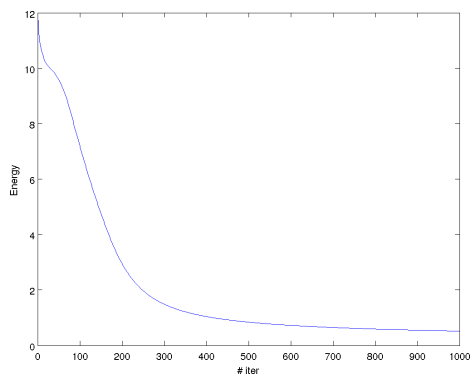Figure 1: (From left to right) The layer 1 hyperplanes and the layer 2 hyperplane for one of the better models

Figure 2: The progress of th energy function vs the number of iterations for the model in Figure 1

	Min error	Median error	Mean error
One layer	1	2	2.45
Two layers, 2 hidden nodes	0	2	1.55
Two layers, 3 hidden nodes	0	0	0.7
Two layers, 4 hidden nodes	0	0	0.2

Table 1: Some statistics on the number of errors made by a single neuron, and by a multilayer network with varying numbers of hidden nodes
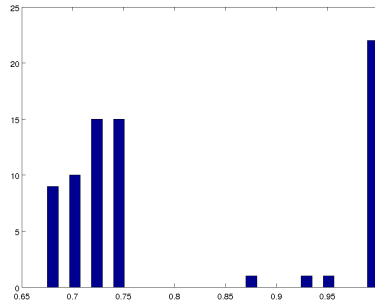
2

Figure 3: A histogram of the accuracies obtained (by thresholding the final output at 0.5) for 75 runs of the training
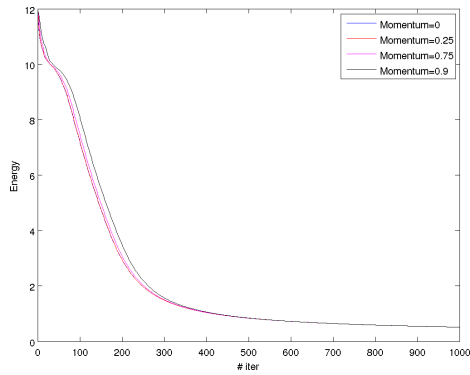

Figure 4: Progress for 4 different values of the momentum, for the model in Figure 1
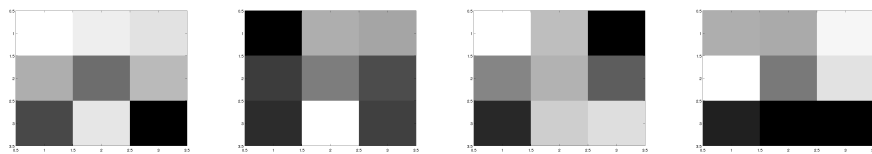

Figure 5: The "filters" learnt by the hidden layer for a model with 4 hidden nodes