1 The Effects of Learning Rate on RBM Training

1.1 Experimental Set-up

The experiments are carried out on the TIMIT corpus. 27D FBank features are extracted from the waveforms. A window of 11 frames are adopted as the RBM's input and for the hidden layer, there are 400 latent units.

The training is done without weight restoring due to the increasing of errors. The learning rate keeps constant during the training, *i.e.* no learning rate decay involved during the learning. A fixed number of epochs are executed, *i.e.* in this experiment 100 epochs. The weight cost for update is set to 0.00002. For weight decay, the first 5 epochs take the value of 0.5, while the following epochs use 0.9.

1.2 Experiment Results

The lowest reconstruction errors on the whole training set for different learning rates are listed in Table 1.2.

Table 1: *Comparison of Reconstruction Error performance of using different Learning Rate*

Learning Rate	Lowest Error Epoch	Lowest Error
1	5	1764473.38
0.9	6	1705593.75
0.8	5	1669502.50
0.7	5	1670629.00
0.6	6	1665445.25
0.5	5	1642355.62
0.4	5	1693728.75
0.3	30	1663507.25
0.2	59	1562948.75
0.1	91	1518075.12
0.01	99	1615933.62
0.001	9	1740395.25

1.3 Trend of Reconstruction Errors during Training

From the experimental results, we can see that with larger learning rates, the reconstruction error first reaches a minimum region, then the error will increase and keep dazzling around that value. One possible reason may due to the momentum change at epoch 5, as more of the cases, the error increases around epoch 5. Following are some more detailed figures about the error changes during the training.

With smaller learning rates, we need more epochs to reach a minima, but it is more robust than larger learning rates.

1.3.1 Learning Rate - 1

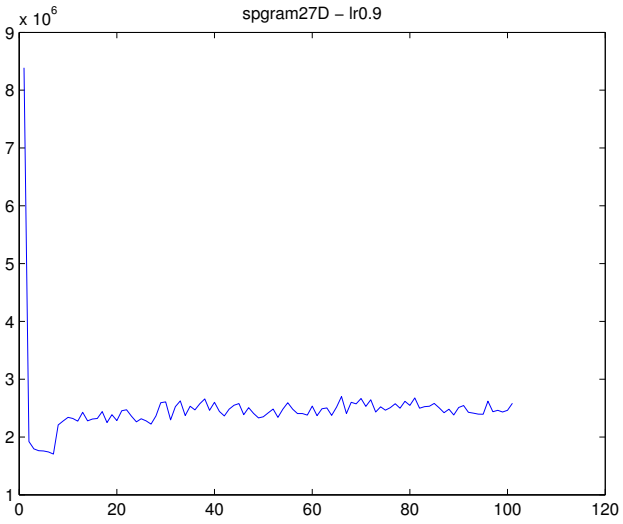

Figure 1: Learning Rate - 1.

1.3.2 Learning Rate - 0.4

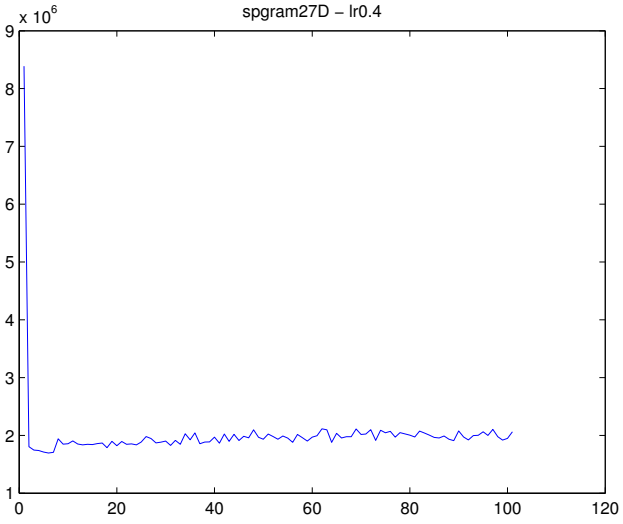

Figure 2: Learning Rate - 0.4.

1.3.3 Learning Rate - 0.3

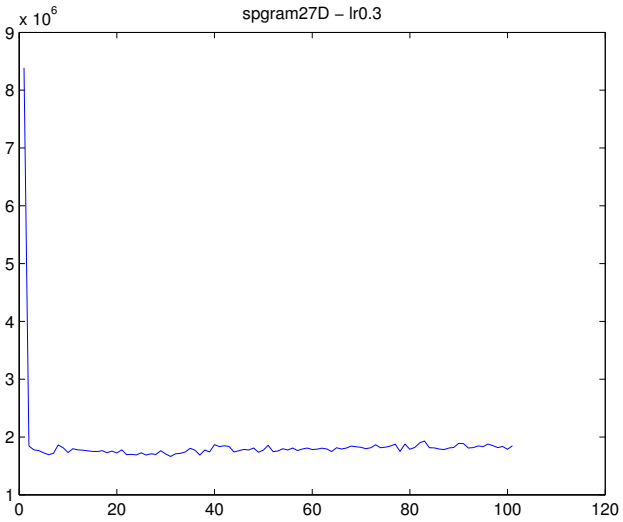

Figure 3: Learning Rate - 0.3.

1.3.4 Learning Rate - 0.2

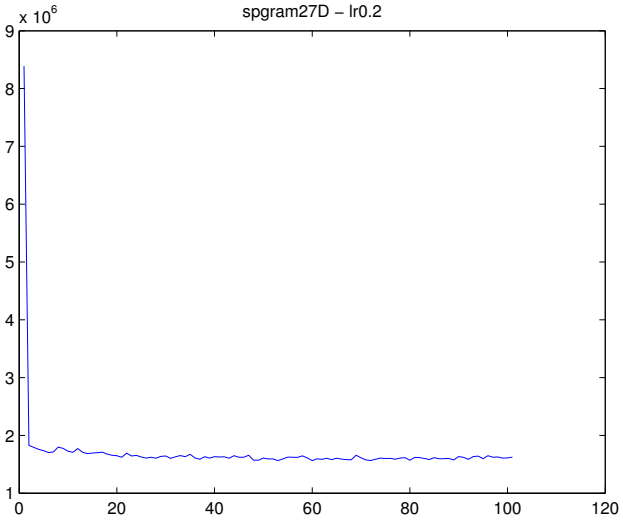

Figure 4: Learning Rate - 0.2.

1.3.5 Learning Rate - 0.1

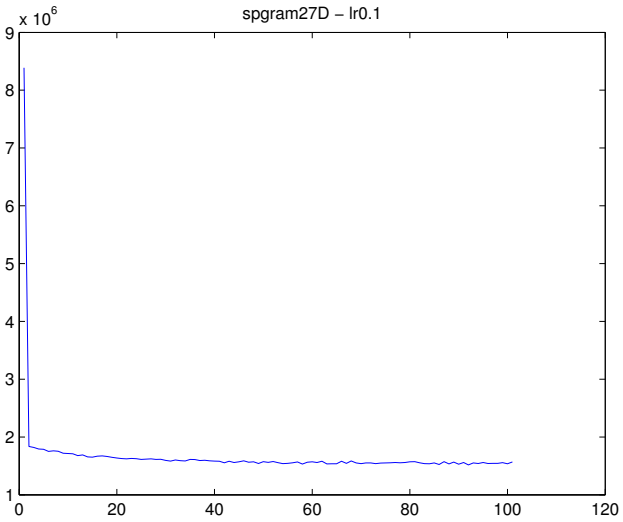

Figure 5: Learning Rate - 0.1.

1.3.6 Learning Rate - 0.01

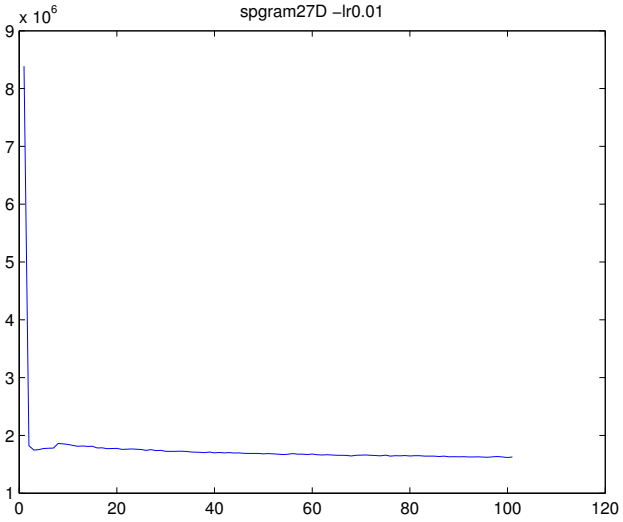

Figure 6: Learning Rate - 0.01.

1.3.7 Learning Rate - 0.001

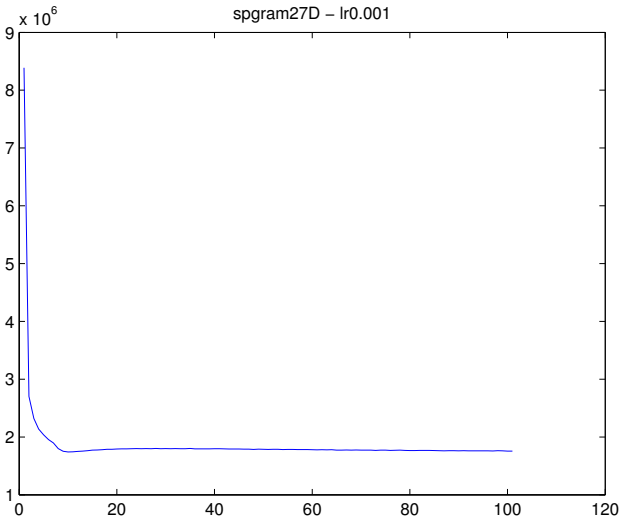

Figure 7: Learning Rate - 0.001.