

DL Programming Assignment: 4

Priyesh V

April 28, 2017

1 RBM Training

I have trained the RBMs with contrastive divergence (CD) algorithm for steps 1, 5 and 10. I trained the model only for 50 epochs and monitored the Mean Squared Error (MSE) between the original image and the deterministically generated image from CD. The MSE seems to reduce even at 50 epochs and seem to continue reducing beyond 50 epochs. Below is a plot for the same. With 50 epochs, the RBM is able to generate 10 separable clusters for the 10 classes of MNIST. Training with CD-5 (5 steps) and CD-10 (10 steps) gives similar reduction in MSE.

The settings used for reporting results:

Hidden units: 500

Learning rate: 0.0001

momentum rate: 0.9

weight decay rate: 0.0002

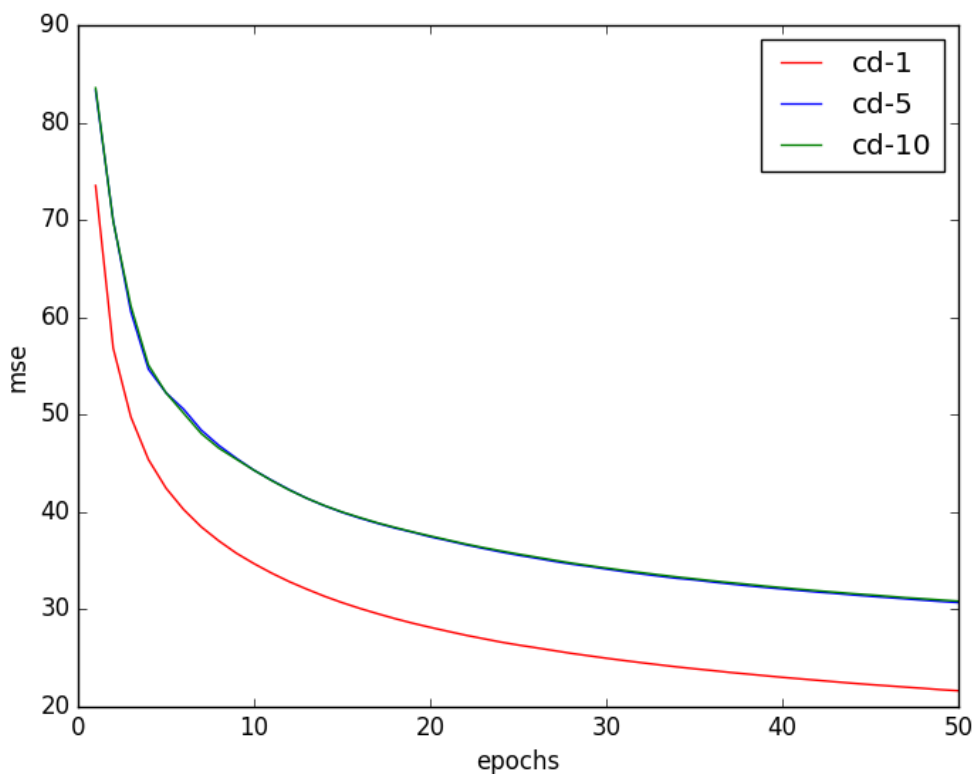


Figure 1: Epochs Vs MSE for varying CD steps

2 T-SNE plots

The T-SNE algorithm were run for 200 epochs with perplexity set to 50. I have reported results for 2 runs with CD-1. Run-1 gives me clear 10 seperable clusters. Run-2 and other two subsequent plots for CD-5 and CD-10 were run with slightly different setup where I use the sampled Hidden value instead of the deterministic Hidden value for CD. These runs give 11 clusters, where 1 is split into 2 clusters.

Observations:

Clusters for 3 and 6 seem to lie closer to each other (Fig 2) and 8 also appearing together in certain cases (Fig 2,4,5) as they have similar shapes.

With increase in CD steps, you seem to get better seperation between the classes. The t-sne plots would have been better if they had been trained for more epochs.

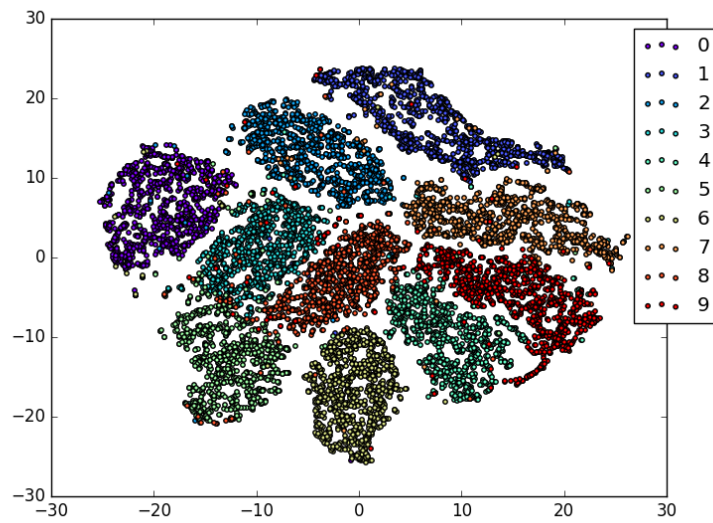


Figure 2: CD-1 embeddings - Run 1

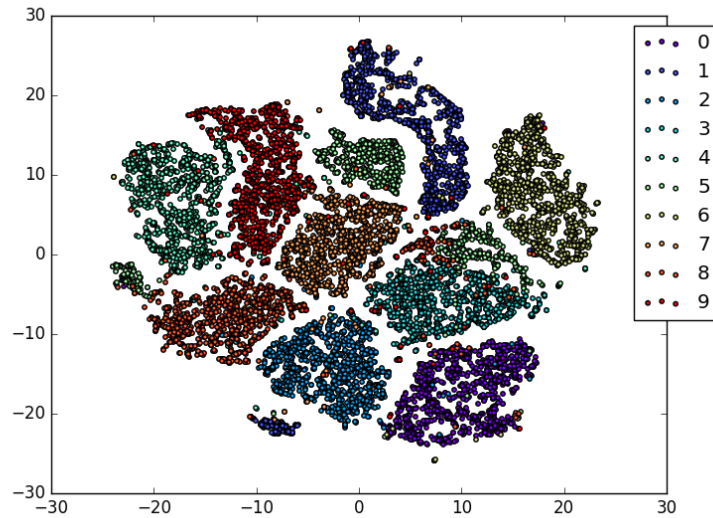


Figure 3: CD-1 embeddings - Run 2

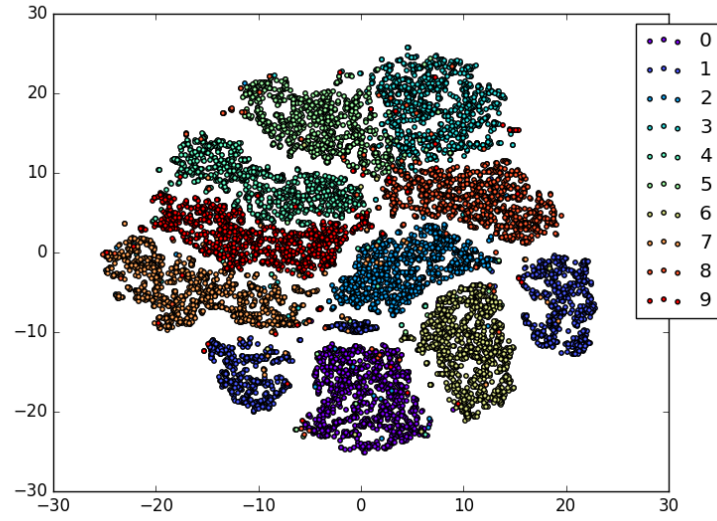


Figure 4: CD-5 embeddings

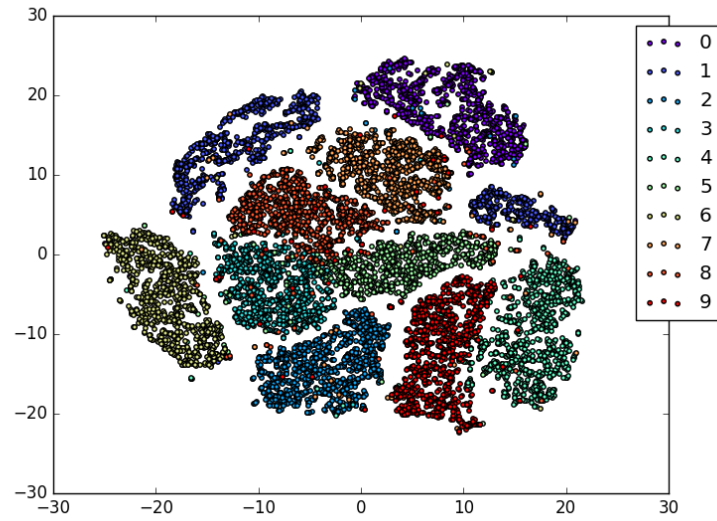


Figure 5: CD-10 embeddings