Classification with Restricted Boltzmann Machines

Projects in Machine Learning and AI

Fritjof Wolf
Katarzyna Tarnowska

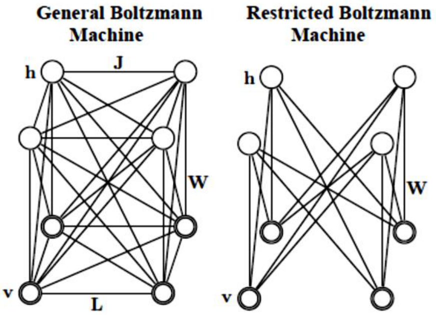
Technische Universitat Berlin

2015

Table of Contents

Table of Contents

Boltzman Machine and Restricted Boltzmann Machine



- The top layer shows stochastic binary hidden units and the bottom layer shows stochastic binary visible units
- Some text comparing General BM
- In a Restricted Boltzmann Machine the joints between hidden units and also between visible units are disconnected

Table of Contents

Sample frame title

Model of a Restricted Boltzmann Machine as a complete bipartite graph. Restricted Boltzmann Machine can be regarded as stochastic neural network, where the nodes and edges correspond to neurons and synaptic commections, respectively.

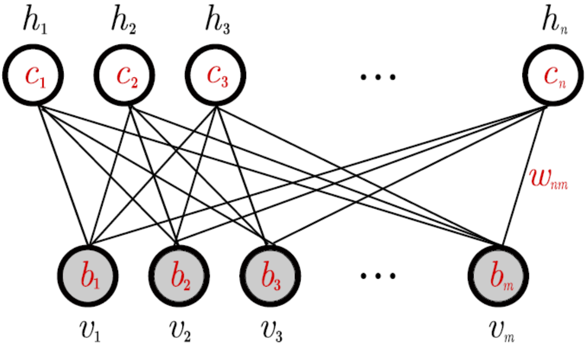


Figure The network graph of an RBM with n hidden and m visible units, Source: A.Fischer, Ch.Igel: Training Restricted Boltzmann Machines: An Introduction

Some useful math equations

$$E(v,h) = \sum_{i=1}^{V} \frac{(v_i - b_i^v)^2}{2\sigma_i^2} - \sum_{j=1}^{H} b_j^h h_j - \sum_{i=1}^{V} \sum_{j=1}^{H} \frac{v_i}{\sigma_i} h_j w_i j \quad (1)$$

$$p(v,h) = \frac{e^{-E(v,h)}}{\sum_x \sum_k e^{-E(x,k)}} \quad (2)$$

Table of Contents

Contrastive Divergence with Gibbs sampling

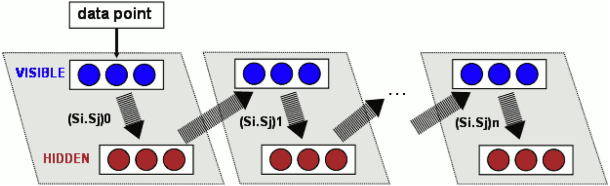
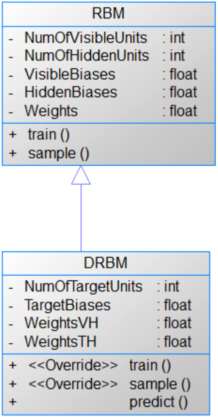
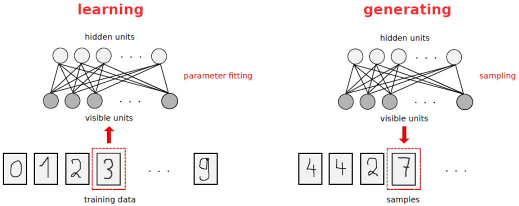


Table of Contents

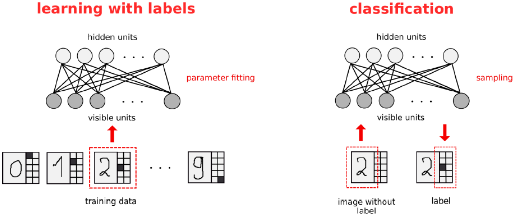
Generative and Discriminative models of RBM





GRBM

Source: A.Fischer, Ch.Igel: Training Restricted Boltzmann Machines: An Introduction



DRBM

Source: A.Fischer, Ch.Igel: Training Restricted Boltzmann Machines: An Introduction

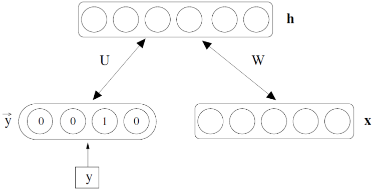
Generative model

- Train phase - fitting RBM parameters so that to model distribution of the training data
- Training performed on inputs from user-given class
- Sample phase - trained RBM used to generate samples from learned distribution
- Shows reconstructed image for the user-specified digit

RBM for classification

Training method:

- Models a joint distribution of inputs (x) and target classes (y)
- Has two sets of visible units and two weight matrices: between x and h (W) and between y and h (U)



- Computes gradients for a mini-batch
- Updates weights with mean gradients and user-defined learning rate
- Iterates through whole dataset in defined number of epochs

RBM for classification

Prediction method:

- Fix the visible variables corresponding to the image
- Sample target variables corresponding to the labels in chosen number of iterations
- Return probabilities of each class
- Choose the class with highest probability
- Perform for each datapoint in testset
- Compare with original labels
- Count wrong predictions and accuracy

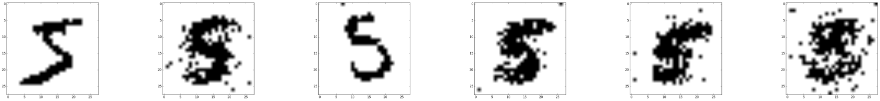
Table of Contents

Testing methodology and assumptions

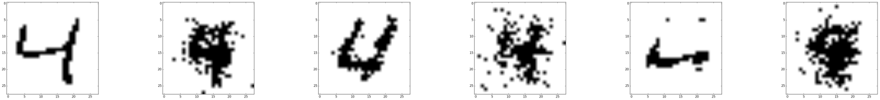
- Tested dataset - MNIST - handwritten digit images (28x28 pixels = 784 features) and their labels (0..9)
- Dataset divided into training (50000), validation (10000) and test (10000) subsets
- Reducing to binary problem(binarization threshold = 0.5)

Testing generative model

- Experiments on different numbers of hidden units
- First image is original, next are for hidden units in size of: 200;300;400;500;700)



- Generating "4": first image is original, next are for hidden units in size of: 100;200;300;400;500)

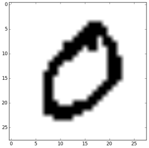
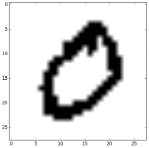
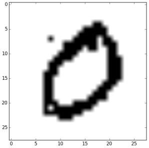


Remark

Different digit classes have different optimal hyperparameters

Testing generative model - joint-probabilities approach

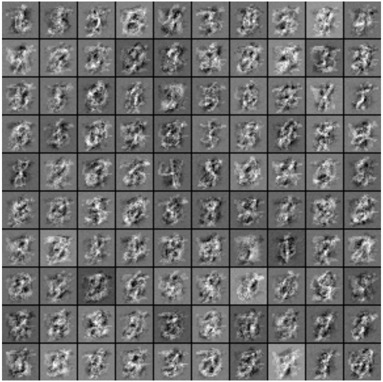
- Learning in mini-batches improved reconstruction and performance
- Momentum parameter for weight update other than 0.0 worsened results
- Optimized results for reconstruction after 500 epochs of training were very good (first is original, second with momentum=0.0, third with momentum=0.5):

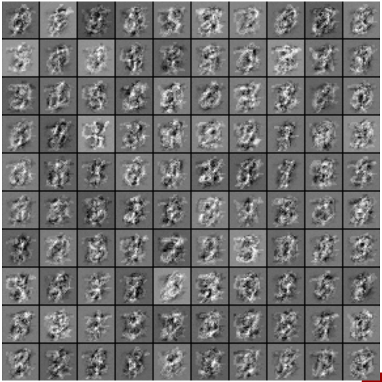




- For 500-epoch training MSE falls below 1.0 - in about 30 minutes

Monitoring progress of learning

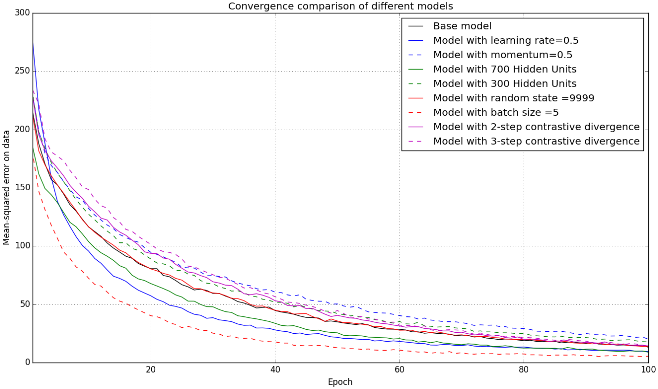
Visualization of the weights of RBM (after 100. and 500. epoch)





Monitoring progress of learning

Reconstruction error for 100 epochs

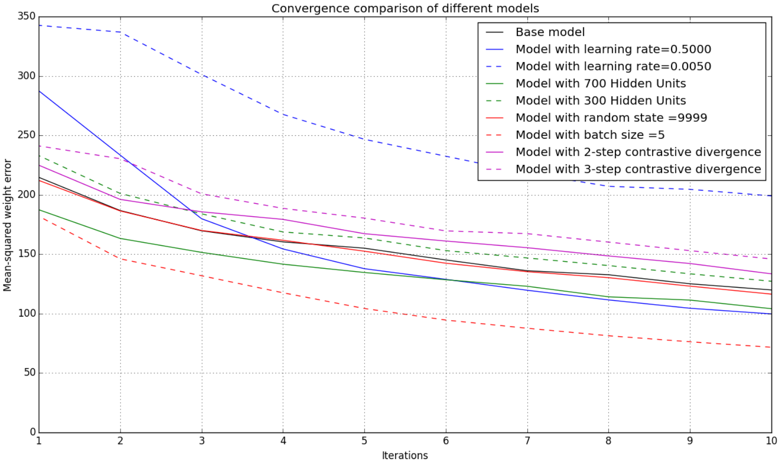


Remark

The reconstruction error on the training set falls rapidly and consistently at the start of learning and then more slowly.

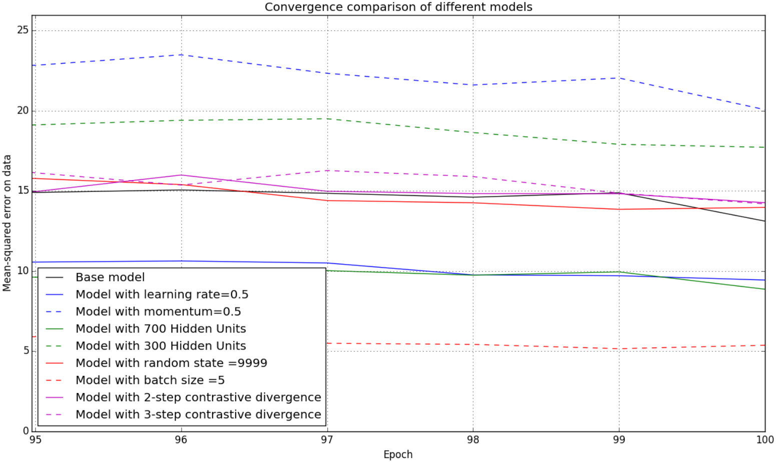
Model selection

Reconstruction error for first 10 epochs



Model selection

Reconstruction error for last 5 epochs



Model selection

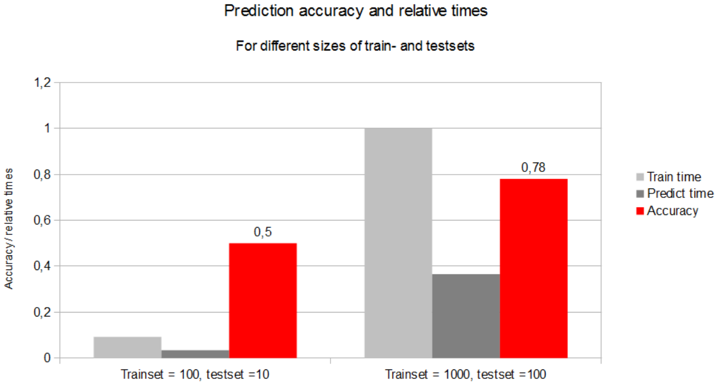
Conclusions

- Learning in smaller mini-batches and increasing number of hidden units (from 400 to 700) improved reconstruction
- However, these changes resulted in longer train time
- Higher learning rate (0.5 instead of 0.05) caused reconstruction error to drop more sharply
- Different random states do not change reconstruction error significantly
- 1-step contrastive divergence is optimal

Testing RBM for prediction (I)

- Optimal hyperparameters for training phase were chosen
- Performing 100 percent accurate classification on training data was trivial
- Classification accuracy on test data possible to achieve in acceptable time was about 80 percent (for 1.5 hour training on 1000- and 0.5 hour testing on 100-dataset, 100 epochs)
- Better results are expected given greater computing power

Testing RBM for prediction (II)



- Improving prediction accuracy on test data requires training a model on greater number of data input and sampling on greater number of test data inputs, but times are prohibitive for personal computing

Plans for further work

- Test-against-all-labels prediction approach
- Optimizing algorithms for best performance
- Testing on gaussian values
- Another dataset, possibly CIFAR

Literature I

Hugo Larochelle, Yoshua Bengio.
Classification using Discriminative Restricted Boltzmann Machines.
Proceedings of the 25th International Conference on Machine Learning, 2008.

Geoffrey Hinton.
A Pratical Guide to Training Restricted Boltzmann Machines.
UTML TR 2010-003, 2010.

Asja Fischer, Christian Igel.
Training Restricted Boltzmann Machines: An Introduction.
UTML TR 2010-003, 2(1):50–100, 2008.

Questions?