

# Guide to training RBM/Dynamic Boltzmann Dist.

OKE

January 4, 2018

Key:

1. **Softmax/multi-valued**
2. **Layer-wise pre-training:** is very important. Train each RBM as a single layer, then propagate the data upward. Finally, train the whole network.
  - (a) **Self-intersecting trajectories in the pre-training:**
3. **Initial values for weights/biases:** A good value for the initial visible bias is  $\log p_i / (1 - p_i)$  where  $p_i$  is the fraction of training data examples where unit  $i$  is on. A good value for the initial weight is close to 0. A good value for the initial hidden bias is 0. A very negative value (e.g.  $-4$ ) can be used here as well to enforce a kind of sparsity constraint, but this can make the weight values very sensitive.
4. **Nesterov acceleration:** start with a low value 0.5; later 0.9 can be used to improve training.