

RBM guide

W.J.MacInnes

From the horse's mouth (Hinton, 2010,

A practical guide ...

Link on his website)

November 24, 2017

Training batch size

- Divide data into training batches
 - 10x 10 batches
- Feed one batch at a time before adjusting weights
 - This is more efficient on GPU
- If the classes to be learned are finite (0..9) the minibatch size should have one example of each class
- Otherwise, simply use 10

overfitting

- If the RBM is a simple generative model, you can fit to a higher level
- But It's not supervised, so what error is used to test overfitting?
 - The probability that the model produces a held-out example
 - Stop when this decreases
 - Calculated using 'free energy'
- when using RBMs as part of DLNN, it may be MORE important to underfit!

Learning rate

- Reducing learning rate during training is a common machine learning technique
 - Large jumps in learning at the start
 - Smaller jumps near the end
 - AKA 'Momentum' Choose momentum of conservative .5 if offered
- For RBMs. only reduce learning rate near the end of training
- Rule of thumb, an update should only change the weight by 10^{-3}

Initial values of weights/biases

- Choose SMALL but random initial weight values
- Large random values can increase initial learning, but at the cost of final model fit
- Set hidden biases to 0
- Set visible biases to $\log[t/(1-t)]$
- Others suggest hidden biases as -4 to encourage sparsity

Number of hidden units

- This one is harder to estimate
- Are the training cases very similar?
- How complex are the examples?

Choice of unit

- Binary
 - Default option for input and hidden
- Softmax
 - Select one of K possible values
- Gaussian
 - Natural data doesn't fit binary or logistic very well
 - Speech, natural images often switch to Gaussian
VISIBLE units
 - I suspect saccades will also land here
 - Network instability becomes a problem if both layers are gaussian