

# Gated Softmax Classification:

Say, you trained  $K$  class-specific Restricted Boltzmann Machines and you would like to combine the  $K$  RBMs for classification. Unlike with, say, mixtures of Gaussians, you cannot simply use Bayes' rule, because each RBM will have a different partition function.

As it turns, however, there *is* in fact a principled way to combine the RBMs for classification:

Just *think of* the set of  $K$  class-specific RBMs as a single conditional distribution  $p(\text{inputs, hidden}|\text{class})$ . Now compute  $p(\text{class}|\text{inputs})$ , integrating over the hidden. The partition functions cancel and you can compute both the probability and the derivatives with respect to all the RBMs parameters in polynomial time. This is the "gated softmax classifier" [[pdf, NIPS2010](#)].

An equivalent view of the model is a log-bilinear classifier (or bilinear logistic regression), whose hidden may be viewed "style" variables that capture within-class variability.

It can be shown that a model with  $K$  latent variables is exactly the same as a mixture of  $2^K$  logistic regression models with weight-sharing. This makes it possible to train a mixture of about 100,000,000,000,000,000,000 linear classifiers and apply it to test data in closed form. An implementation of the model in Python is provided below.

## Code

The following two Python modules implement two versions of the model. Both modules make use of GPUs via V. Mnih's `cudamat` package (linked below).

[gatedSoftmaxCuda.py](#)

The basic, "unfactored" model.

[gatedSoftmaxFactoredCuda.py](#)

The "factored" model, whose parameter tensor is represented by low-rank matrices. This makes it possible to represent invariances using shared basis functions as described in the paper.

Prerequisites: [numpy](#), [cudamat](#).

The bottom of each file (the `__name__ == '__main__'` clause) contains example code that instantiates and applies the models to dummy data.

## Errata

The gradient in the NIPS 2010 [paper](#) contains a bug, which is corrected [here](#).

## References

2010 Memisevic, R. Zach, C., Hinton, G., Pollefeys M.

**Gated Softmax Classification**

Neural Information Processing Systems (NIPS) 2010. [[pdf](#)]