Probabilistic Modelling and Reasoning, Tutorial Question Sheet 5 (for Week 8)

School of Informatics, University of Edinburgh

Instructor: Amos Storkey
Prob. Sheet. ID: 27114.1

A Boltzmann Machine takes the form

$$P(\mathbf{x}|\mathbf{W}, \mathbf{a}, \mathbf{b}) = \frac{1}{Z} \exp\left(\frac{1}{2}\mathbf{x}^T\mathbf{W}\mathbf{x} + \mathbf{b}^T\mathbf{x}\right)$$

1. Show that the probability $P(x_i = 1|x_{rest}) = 1/(1 + \exp(-\mathbf{w}_i^T\mathbf{x} - b_i))$ where $\mathbf{w}_i^T$ is the ith row of $\mathbf{w}$. Write a matlab function `samplebm(w,b,n)` to generate from a Boltzmann machine with weight matrix w, and b a column vector of biases. The function should return a $d * n$ matrix, where n is the number of samples. You should do the sampling by choosing a random initialisation and Gibbs sampling (can be done in 6 lines of clean Matlab code). This will help you understand the working of the Boltzmann machine, and the working of Gibbs sampling. If you do this you will probably remember both.

   Now download the mnist data from the tutorial website. Run the code on the PMR website. It will set a weight matrix directly (using a Hopfield Learning Rule rather than a Boltzmann Learning rule). This enables you to see the process of sampling from the Boltzmann Machine. What do you notice?

2. Now build a Gibbs sampling process for the restricted Boltzmann Machine with 784 visible units and 784 hidden units, and implement a simple batch contrastive divergence rule. Run this on the mnist training data from the website. Then sample for the hidden units, and run another RBM on that. Keep track of all the weight matrices. Finally run a simple classifier on all the features and test your classification performance on the mnist test set. Alternatively if you are short of time, find read and try out code that does this...

3. Show that we can write the same model in the same form but where the binary variables have a $\{-1, 1\}$ representation instead. What is the transformation for the parameters in one form to the parameters in the other?

4. Show that we can add a diagonal term to the $\mathbf{W}$ without affecting the model (hint - it is easiest to do this in the $\{-1, 1\}$ representation.

5. Derive the learning rule for a restricted Boltzmann machine with zero biases where both $\mathbf{x}$ and $\mathbf{y}$ are visible variables.

6. What is a restricted Boltzmann Machine with one hidden unit? (Draw and manipulate the graphical model).

7. Consider a variant of the restricted Boltzmann machine:

   $$P(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x} + \mathbf{x}^T\mathbf{W}\mathbf{y} + \mathbf{b}^T\mathbf{y}\right) \quad (1)$$

   where $\mathbf{x}$ is real valued and $\mathbf{y} = \pm 1$.

Write the content of the exponent in the form $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}(\mathbf{y}))^T(\mathbf{x}-\boldsymbol{\mu}(\mathbf{y}))+F(\mathbf{y})$. What are $\boldsymbol{\mu}(\mathbf{y})$ and $F(\mathbf{y})$. Rewrite (1) using this exponent, separating out the terms with $\mathbf{x}$ in and the terms without. Notice that the terms with $\mathbf{x}$ in are Gaussian and integrate over them to get a constant. What are you left with? Compare this with a Boltzmann Machine.

8. Now rewrite (1) as

$$P(\mathbf{x}^*, \mathbf{y}) = \frac{1}{Z'} \exp\left(-\frac{1}{2}(\mathbf{x}^*)^T(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{x}^*\right) \prod_i \exp\left(x_i^* y_i + b_i y_i\right) \tag{2}$$

where $\mathbf{x}^* = \mathbf{W}^T\mathbf{x}$. Marginalise over the y_i to get the induced distribution over the $\mathbf{x}^*$. This shows how we can represent the Boltzmann distribution in terms of real valued augmented variables.