CSE 5526: Homework 5 Solution

Problem 1. Following slide 26, we first consider $x_i = -1$.

In this case,

$$P(x_i \to -x_i) = \varphi(v_i) = \frac{1}{1 + \exp(-\frac{v_i}{T})}$$

In the other case when $x_i = 1$,

$$P(x_i \to -x_i) = 1 - \varphi(v_i) = 1 - \frac{1}{1 + \exp(-\frac{v_i}{T})} = \frac{1}{1 + \exp(\frac{v_i}{T})}$$

Thus,

$$P(x_i \to -x_i) = \frac{1}{1 + \exp(x_i \frac{v_i}{T})} = \frac{1}{1 + \exp(\frac{x_i}{T} \sum_j w_{ij} x_j)}$$

On the other hand, such a flip leads to the following energy change:

$$\Delta E_i = -\frac{1}{2} \sum_k \sum_j w_{jk} x'_k x'_j + \frac{1}{2} \sum_k \sum_j w_{jk} x_k x_j = 2x_i \sum_j w_{ji} x_j$$

Where x'_k and x'_j are equal to x_k except for being negated at bit i

Therefore:

$$P(x_i \to -x_i) = \frac{1}{1 + \exp(\frac{\Delta E_i}{2T})}$$

where ΔE_i is the energy change due to the flip.

OR, following slide 25,

The probability of flipping from 1 to -1 is

$$P(x_i = 1 \to x_i = -1|\boldsymbol{x}_{\sim i}) = P(x_i = -1|\boldsymbol{x}_{\sim i}) = \frac{\frac{1}{Z}\exp\left(-\frac{E_i^-}{T}\right)}{\frac{1}{Z}\exp\left(-\frac{E_i^+}{T}\right) + \frac{1}{Z}\exp\left(-\frac{E_i^-}{T}\right)}$$

$$= \frac{1}{1 + \exp\left(\frac{1}{T}(E_i^- - E_i^+)\right)} = \frac{1}{1 + \exp\left(\frac{\Delta E_i}{T}\right)}$$

Where in this case $\Delta E_i = E_{\text{final}} - E_{\text{initial}} = E_i^- - E_i^+$

And the probability of flipping from -1 to 1 is

$$P(x_i = -1 \to x_i = 1|\boldsymbol{x}_{\sim i}) = P(x_i = 1|\boldsymbol{x}_{\sim i}) = \frac{\frac{1}{Z}\exp\left(-\frac{E_i^+}{T}\right)}{\frac{1}{Z}\exp\left(-\frac{E_i^-}{T}\right) + \frac{1}{Z}\exp\left(-\frac{E_i^+}{T}\right)}$$

$$= \frac{1}{1 + \exp\left(\frac{1}{T}(E_i^+ - E_i^-)\right)} = \frac{1}{1 + \exp\left(\frac{\Delta E_i}{T}\right)}$$

Where in this case $\Delta E_i = E_{\text{final}} - E_{\text{initial}} = E_i^+ - E_i^-$

So in both cases,

$$p(x_i \to -x_i) = \frac{1}{1 + \exp\left(\frac{\Delta E_i}{T}\right)}$$

Problem 2. The infinitely deep belief net is equivalent to a single RBM because they both describe the same distribution over visible units, sampling from them is equivalent, and the same maximum likelihood learning rule is used to train both networks. The depth in the belief net corresponds to steps of alternating Gibbs sampling in the RBM.