STA2700 - Graphical Models - Assignment 3

Luis Correia - Student No. 1006508566

November 19th 2020

Question 1

Let X and Y be two i.i.d R.V.'s with entropy $H(X)$. Prove that

$$Pr(X = Y) \geq 2^{-H(X)}$$

{*Solution.*}

Let's remember that, if X and Y shares the same entropy function, this means theu have the same probability distribution since, by definition, the entropy function is deterministic number and function of the probability distribution of X.

That being said, we have:

$$P(X = Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(X = x, Y = y) \cdot \mathbb{1}(x, y) \tag{1}$$

where

$$\mathbb{1}(x, y) = \begin{cases} 1, \text{ if } x = y \\ 0, \text{ if } x \neq y \end{cases} \tag{2}$$

As X and Y are independent, from (1) we have:

$$\begin{aligned} P(X = Y) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(X = x) \cdot P(Y = y) \cdot \mathbb{1}(x, y) \\ &= \sum_{x \in \mathcal{X}} P(X = x) \cdot \left[\sum_{y \in \mathcal{Y}} P(Y = y) \cdot \mathbb{1}(x, y) \right] \\ &= \sum_{x \in \mathcal{X}} P(X = x) \cdot P(X = x) \\ &= \mathbb{E}\Big[P(X) \Big] \\ &= 2^{\log_2 \{\mathbb{E}[P(X)]\}} \end{aligned}$$

$$\implies P(X = Y) = 2^{\log_2 \{\mathbb{E}[P(X)]\}} \tag{3}$$

Let's remember that the entropy $H(X)$ can be calculated, by definition, as follows:

$$H(X) = \sum_{x \in \mathcal{X}} P(X) \log_2 \left(\frac{1}{P(X)} \right) = \mathbb{E}\Big\{ \log_2 \left[\frac{1}{P(X)} \right] \Big\} \tag{4}$$

As $\log_2 \dfrac{1}{P(X)}$ is a convex function, we can apply *Jensen's Inequality* on (4). Then we have:

$$\begin{aligned} H(X) &= \mathbb{E}\Big\{ \log_2 \left[\frac{1}{P(X)} \right] \Big\} \\ &\geq \log_2 \left\{ \mathbb{E}\left[\frac{1}{P(X)} \right] \right\} \\ &= -\log_2 \mathbb{E}\Big[P(X) \Big] \end{aligned}$$

$$\implies \log_2 \mathbb{E}\Big[P(X) \Big] \geq -H(X) \tag{5}$$

Now, substituting in (5) in (3) we have:

$$\begin{aligned} P(X = Y) &= 2^{\log_2 \{ \mathbb{E}[P(X)] \}} \\ &\geq 2^{-H(X)} \end{aligned}$$

$$\implies P(X = Y) \geq 2^{-H(X)} \tag{6}$$

Question 2

Give an example of two PMF's p and q with $\mathcal{X} = \{0, 1\}$ such that

$$\mathcal{D}(p \parallel q) = \mathcal{D}(q \parallel p)$$

The case $p = q$ is trivial, we need a non-trivial example.

{*Solution.*}

The relative entropy or **Kullback-Leibler** divergence between two probability distributions $p(X)$ and $q(X)$ defined that are defined over the same alphabet $\mathcal{X}$ is:

$$\mathcal{D}_{KL}(p \parallel q) = \sum_{x \in \mathcal{X}} p(x) \cdot \log \frac{p(x)}{q(x)} \tag{7}$$

For this question, let's consider the following non-trivial example:

$$p(x) = \begin{cases} 0.4, \text{ if } x = 0 \\ 0.6, \text{ if } x = 1 \end{cases} \tag{8}$$

... and

$$q(x) = \begin{cases} 0.6, \text{ if } x = 0 \\ 0.4, \text{ if } x = 1 \end{cases} \tag{9}$$

Calculating the relative entropy $D(p \parallel q)$ we have:

$$\begin{aligned} \mathcal{D}(p \parallel q) &= \sum_{x \in \{0,1\}} p(x) \cdot \log \frac{p(x)}{q(x)} \\ &= p(0) \cdot \log \frac{p(0)}{q(0)} + p(1) \cdot \log \frac{p(1)}{q(1)} \\ &= 0.4 \cdot \log \frac{0.4}{0.6} + 0.6 \cdot \log \frac{0.6}{0.4} \\ &= 0.116993 \end{aligned}$$

Now, $D(q \parallel p)$:

$$\begin{aligned} \mathcal{D}(q \parallel p) &= \sum_{x \in \{0,1\}} q(x) \cdot \log \frac{q(x)}{p(x)} \\ &= q(0) \cdot \log \frac{q(0)}{p(0)} + q(1) \cdot \log \frac{q(1)}{p(1)} \\ &= 0.6 \cdot \log \frac{0.6}{0.4} + 0.4 \cdot \log \frac{0.4}{0.6} \\ &= 0.116993 \end{aligned}$$

... and we have $\mathcal{D}(p \parallel q) = \mathcal{D}(q \parallel p)$.

Question 3

Consider a $1D$ homogeneous antiferromagnetic Ising model with periodic boundary conditions. We denote the coupling parameter by J, which is a negative real number. Let N be a number of particles in the model (i.e., the number of variable nodes in the corresponding factor graph).

Inthe thermodynamic limit (i.e., as $N \to \infty$) and in the low-temperature limit (i.e., $\beta J \to -\infty$), how many valid configurations does the model have?

{*Solution.*}

In order to calculate the number of valid configurations we need to derive from *free energy of the lattice* which depends on the partition function and, after applying the limit for $N \to \infty$ the result is the thermodynamic limit:

$$F = F(\beta, E, N) = \lim_{N\to\infty} \frac{1}{N} \log_2 Z(\beta, \{J_{i,j}\}, N) \tag{10}$$

In the present case, the anti-ferromagnetic Ising Model states the coupling parameter constant and equal to J, which is a negative real number.

Then, the model configuration can be represented by the sequence $\{X_i\}_{i=1}^N \in \mathcal{X}^N$, with $\mathcal{X} = \{-1, +1\}$. As our model has periodic boundary conditions, it means the energy function $E(\boldsymbol{x})$ can be written as:

$$E(\boldsymbol{x}) = -J\Big(\sum_{i=1}^{N-1} x_i \cdot x_{i+1} + x_N \cdot x_1\Big) \tag{11}$$

This implies, in our particular case, that the partition function becomes:

$$Z(\beta, J, N) = \sum_{\boldsymbol{x}\in\mathcal{X}^N} e^{-\beta E(\boldsymbol{x})} = \sum_{\boldsymbol{x}\in\mathcal{X}^N} \exp\Big[\beta J\big(\sum_{i=1}^{N-1} x_i \cdot x_{i+1} + x_N \cdot x_1\big)\Big] \tag{12}$$

Note that the last term represents the cyclic characteristic of the model with periodic boundary conditions.

In order to facilitate the algebraic manipulations, we will map the x_i's into a more convenient representation using $\tau_i = x_i \cdot x_{i+1}$, with $1 \le i \le N-1$ and $\tau_N = x_N \cdot x_1$.

Then (11) can be rewritten as:

$$E(\boldsymbol{\tau}) = -J\sum_{i=1}^{N} \tau_i \tag{13}$$

Note also that, due to this representation and the periodic boundary conditions, we also have:

$$\prod_{i=1}^{N} \tau_i = \prod_{i=1}^{N} x_i^2 = 1 \tag{14}$$

Then, substituting (13) in (12) we have:

$$Z(\beta, J, N) = \sum_{\{\tau\}} \exp\Big(\beta J \sum_{i=1}^{N} \tau_i\Big) \cdot \mathbb{1}\Big(\prod_{i=1}^{N} \tau_i, 1\Big) \tag{15}$$

... where $\mathbb{1}$ is an indicator function.

It follows that we can rewrite (15) in the following way:

$$\begin{aligned} Z(\beta, J, N) &= \sum_{\{\tau\}} \exp\Big(\beta J \sum_{i=1}^{N} \tau_i\Big) \cdot \Big(1 + \prod_{i=1}^{N} \tau_i\Big) \\ &= \sum_{\{\tau\}} \Bigg[\exp\Big(\beta J \sum_{i=1}^{N} \tau_i\Big) + \exp\Big(\beta J \sum_{i=1}^{N} \tau_i\Big) \cdot \prod_{i=1}^{N} \tau_i \Bigg] \\ &= \sum_{\{\tau\}} \Bigg[\prod_{i=1}^{N} \exp\big(\beta J \tau_i\big) + \prod_{i=1}^{N} \tau_i \exp\big(\beta J \tau_i\big) \Bigg] \end{aligned}$$

$$\implies Z(\beta, J, N) = \sum_{\{\tau\}} \Bigg[\prod_{i=1}^{N} \exp\big(\beta J \tau_i\big) + \prod_{i=1}^{N} \tau_i \exp\big(\beta J \tau_i\big) \Bigg] \tag{16}$$

Note that $\tau_i = \pm 1$ then we can simplify (16) as follows:

$$\begin{aligned} Z(\beta, J, N) &= \prod_{i=1}^{N} \sum_{\tau_i = \pm 1} \exp\big(\beta J \tau_i\big) + \prod_{i=1}^{N} \sum_{\tau_i = \pm 1} \tau_i \exp\big(\beta J \tau_i\big) \\ &= \prod_{i=1}^{N} \Big(e^{\beta J} + e^{-\beta J}\Big) + \prod_{i=1}^{N} \Big(e^{\beta J} - e^{-\beta J}\Big) \\ &= \prod_{i=1}^{N} 2\cosh\big(\beta J\big) + \prod_{i=1}^{N} 2\sinh\big(\beta J\big) \\ &= 2^N \Big(\cosh\big(\beta J\big)^N + \sinh\big(\beta J\big)^N \Big) \end{aligned}$$

$$\implies Z(\beta, J, N) = 2^N \cosh\big(\beta J\big)^N \Big(1 + \tanh\big(\beta J\big)^N\Big) \tag{17}$$

Now, to calculate the thermodynamic limit F, applying (17) in (10), we have:

$$\begin{aligned} F(\beta, E, N) &= \lim_{N\to\infty} \frac{1}{N} \ln Z(\beta, \{J_{i,j}\}, N) \\ &= \lim_{N\to\infty} \frac{1}{N} \ln \Big[2^N \cosh\big(\beta J\big)^N \Big(1 + \tanh\big(\beta J\big)^N\Big) \Big] \\ &= \lim_{N\to\infty} \frac{1}{N} \Big[\ln 2^N + \ln \cosh\big(\beta J\big)^N + \ln\Big(1 + \tanh\big(\beta J\big)^N\Big) \Big] \\ &= \lim_{N\to\infty} \frac{1}{N} \Big[N \cdot \ln 2 + N \cdot \ln\cosh\big(\beta J\big) + \ln\Big(1 + \tanh\big(\beta J\big)^N\Big) \Big] \\ &= \ln 2 + \ln\cosh\big(\beta J\big) + \underbrace{\lim_{N\to\infty} \frac{1}{N} \ln\Big(1 + \tanh\big(\beta J\big)^N\Big)}_{=0} \\ &= \ln 2 + \ln\cosh\big(\beta J\big) \end{aligned}$$

$$\implies F(\beta, E, \infty) = \ln 2 + \ln\cosh\big(\beta J\big) \tag{18}$$

Finally, to calculate the number of valid configurations in the low-temperature limit, we need to calculate the limit of (18) when $\beta J \to -\infty$, then we have:

$$\begin{aligned}
F(-\infty, \infty) &= \lim_{\beta J \to -\infty} \Big(\ln 2 + \ln\cosh\big(\beta J\big)\Big) \\
&= \ln 2 + \underbrace{\lim_{\beta J \to -\infty} \Big(\ln\cosh\big(\beta J\big)\Big)}_{=0} \\
&= \ln 2.
\end{aligned}$$

Therefore, the number of valid configurations at the thermodynamic limit and low-temperature limit of this homogeneous anti-ferromagnetic Ising model is ln(2).

Question 4

Consider a $1D$ Ising Model with free boundary conditions of size N. The Hamiltonian (the energy function) is given by

$$\mathcal{H}(\boldsymbol{x}) = -J \sum_{(i,j) neighbors} x_i x_j - B \sum_{1 \leq k \leq N} x_k \tag{19}$$

where J is the coupling parameter and B denotes the presence of an external field.

(a) For $N = 4$. Draw the factor graph. Show that the graph has pairwise factors (which depends on two variables) and unary factors (which depends on only one variable);

(b) Compute the free energy per site in the thermodynamic limit, i.e.,

$$f = \lim_{N \to \infty} \frac{\ln(Z)}{N} \tag{20}$$

{*Solution.*}

item (a)

For $N = 4$ we can represent the factor graph as follows:

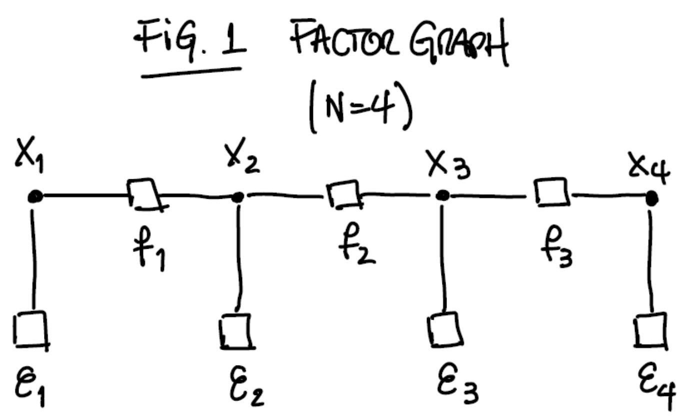

Figure 1: Graphical Model

Where the factors can be represented as function of variables as follows:

- $f_1(x_1, x_2) = x_1 \cdot x_2$
- $f_2(x_2, x_3) = x_2 \cdot x_3$
- $f_3(x_3, x_4) = x_3 \cdot x_4$

... and

- $\epsilon_1(x_1) = -B \cdot x_1$

- $\epsilon_2(x_2) = -B \cdot x_2$
- $\epsilon_3(x_3) = -B \cdot x_3$
- $\epsilon_4(x_4) = -B \cdot x_4$

item (b)

Let's consider the one-dimension Ising Model with free boundary conditions of size N.

Using the Hamiltonian represented in (19), we can write the partition function $Z(\beta, J, B, N)$ as follows:

$$Z(\beta, J, B, N) = \sum_{\boldsymbol{x} \in \mathcal{X}^N} \exp\Big(\beta J \sum_{i=1}^{N-1} x_i \cdot x_{i+1} + \beta B \sum_{j=1}^{N} x_j\Big) \tag{21}$$

Differently of what was done in **Question 3**, in this question we will rewrite the partition function in order to simplify the calculation of (21) by designing a factorization of each component in a way they can be grouped.

$$\begin{aligned}
Z(\beta, J, B, N) = \sum_{\boldsymbol{x} \in \mathcal{X}^N} \exp\Big(&\beta J(x_1 x_2 + x_2 x_3 + \cdots + x_{N-1} x_N) + \\
&\beta B(x_1 + x_2 + \cdots + x_N)\Big) \\
= \sum_{\boldsymbol{x} \in \mathcal{X}^N} \exp\Big(&\beta J x_1 x_2 + \beta J x_2 x_3 + \cdots + \beta J x_{N-1} x_N + \frac{\beta B}{2} x_1 + \frac{\beta B}{2} x_1 + \\
&\frac{\beta B}{2} x_2 + \frac{\beta B}{2} x_2 + \cdots + \frac{\beta B}{2} x_N)\Big) \\
= \sum_{\boldsymbol{x} \in \mathcal{X}^N} \exp\Big[&\beta J x_1 x_2 + \frac{\beta B}{2}(x_1 + x_2)\Big] \cdot \ldots \cdot \exp\Big[\beta J x_{N-1} x_N + \\
&\frac{\beta B}{2}(x_{N-1} + x_N)\Big] \cdot \exp\Big[\frac{\beta B}{2}(x_N + x_1)\Big]
\end{aligned}$$

If we define a function $G(x_i, x_j)$ and $H(x_i, x_j)$ such as

$$G(x_i, x_j) = \exp\big[\beta J x_i x_j + \frac{\beta B}{2}(x_i + x_j)\big] \tag{22}$$

$$H(x_i, x_j) = \exp\big[\frac{\beta B}{2}(x_i + x_j)\big] \tag{23}$$

We can now apply (22) and (23) to rewrite $Z(.)$ as follows:

$$Z(\beta, J, B, N) = \underbrace{\sum_{x_1} \sum_{x_2} \cdots \sum_{x_N} G(x_1, x_2) \cdot G(x_2, x_3) \cdot \ldots \cdot G(x_{N-1}, x_N)}_{\boldsymbol{Z}_1} \cdot \underbrace{\exp\Big[\frac{\beta B}{2}(x_N + x_1)\Big]}_{\boldsymbol{Z}_2}$$

Now using the fact that x_i is either $+1$ or -1, for $1 \le i \le N$ we can configure the matrix $\boldsymbol{G}$ and $\boldsymbol{H}$ using (22) and (23) respectively, as follows:

$$\boldsymbol{G} = \begin{bmatrix} G(+1,+1) & G(+1,-1) \\ G(-1,+1) & G(-1,-1) \end{bmatrix} = \begin{bmatrix} \exp\big[\beta(J+B)\big] & \exp(-\beta J) \\ \exp(-\beta J) & \exp\big[\beta(J-B)\big] \end{bmatrix} \tag{24}$$

and

$$\boldsymbol{H} = \begin{bmatrix} H(+1,+1) & 1 \\ 1 & H(-1,-1) \end{bmatrix} = \begin{bmatrix} \exp\big[\beta B\big] & 1 \\ 1 & \exp\big[-\beta B\big] \end{bmatrix} \tag{25}$$

Observing $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ we noticed that each multiplicative factor $G(.,.)$ when summed over all possible configurations of $x_i x_j$ can be rewritten as a product of matrix in (24) and (25), i.e., the matrices $\boldsymbol{G}^{N-1}$ and $\boldsymbol{H}$ on the summation over x_1 as follows:

$$\begin{aligned} Z(\beta,J,B,N) &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_N} G(x_1,x_2)\cdot G(x_2,x_3)\cdot\ldots\cdot G(x_{N-1},x_N)\cdot H(x_1,x_N) \\ &= \sum_{x_1} G(x_1,x_1)^{N-1}\cdot H(x_1,x_1) \\ &= Tr(\boldsymbol{G}^{N-1}\cdot \boldsymbol{H}) \\ &= Tr(\boldsymbol{P}\boldsymbol{\Lambda}^{N-1}\boldsymbol{P}^T\boldsymbol{H}) \\ &= Tr(\boldsymbol{\Lambda}^{N-1}\boldsymbol{P}^T\boldsymbol{H}\boldsymbol{P}) \end{aligned}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of $\boldsymbol{G}$ and $\boldsymbol{P}$ the matrix of the respective eigenvectors.

We know[1] the eigenvalues can be calculated as follows:

$$\lambda = e^{\beta J}\cosh(\beta B) \pm \left(e^{2\beta J}\sinh^2(\beta B) + e^{-2\beta J}\right)^{1/2} \tag{26}$$

Then the partition function becomes:

$$\begin{aligned} Z(\beta,J,B,N) &= \big(2\cosh(\beta B)\big)^{N-1}\cdot\big(2\cosh(\beta B)\big) \\ &= 2^N\cosh^N(\beta B) \end{aligned}$$

$$Z(\beta,J,B,N) = 2^N\cosh^N(\beta B) \tag{27}$$

Now, using (27) we calculate the thermodynamic limit as follows:

$$\begin{aligned} f &= \lim_{N\to\infty} N^{-1}\ln\big[Z(\beta,J,B,N)\big] \\ &= \lim_{N\to\infty} N^{-1}\ln\big[2^N\cosh^N(\beta B)\big] \\ &= \lim_{N\to\infty} N^{-1}\big\{N\ln 2 + N\ln[\cosh(\beta B)]\big\} \\ &= \ln 2 + \ln[\cosh(\beta B)] \end{aligned}$$

$\Longrightarrow$ The free energy per site in the thermodynamic limit is $\ln 2 + \ln[\cosh(\beta B)]$.

[1] Baxter, R.J.- *Exactly Solved Models in Statistical Mechanics*, Academic Press, 1982