# MFML2 - Homework 6

Group 7:
Thanh Nam Nguyen,
Lorenz Hermann Huber,
Alexander Dmitrievic Efremov,
Noah Frederik Krumbüge

November 30, 2025

## 6.1

(a) We prove by induction on $d \in \mathbb{N}$ that there exists a Gray code $G_d$ of length $2^d$ on $\{0,1\}^d$.

For $d = 1$ set
$$G_1 := (g_0, g_1) := (0, 1).$$
Then $\{g_0, g_1\} = \{0,1\} = \{0,1\}^1$, and the Hamming distance satisfies $d_H(g_1, g_0) = 1$, so $G_1$ is a Gray code in dimension 1.

Assume that for some $d \in \mathbb{N}$ there exists a Gray code
$$G_d = (g_0, g_1, \dots, g_{2^d-1})$$
on $\{0,1\}^d$, i.e. $\{g_0, \dots, g_{2^d-1}\} = \{0,1\}^d$ and $d_H(g_i, g_{i-1}) = 1$ for all $i = 1, \dots, 2^d - 1$.

We construct a Gray code $G_{d+1}$ on $\{0,1\}^{d+1}$ as follows:
$$G_{d+1} := \big((g_0, 0), \dots, (g_{2^d-1}, 0), (g_{2^d-1}, 1), (g_{2^d-2}, 1), \dots, (g_0, 1)\big).$$

The first block $(g_0, 0), \dots, (g_{2^d-1}, 0)$ contains all vectors in $\{0,1\}^{d+1}$ whose last coordinate is 0, since $\{g_0, \dots, g_{2^d-1}\} = \{0,1\}^d$. The second block $(g_{2^d-1}, 1), \dots, (g_0, 1)$ analogously contains all vectors whose last coordinate is 1. Thus every vector in $\{0,1\}^{d+1}$ appears exactly once in $G_{d+1}$.

For $i = 1, \dots, 2^d - 1$,
$$d_H\big((g_i, 0), (g_{i-1}, 0)\big) = d_H(g_i, g_{i-1}) = 1,$$
since appending the same last coordinate does not change Hamming distance and $G_d$ is a Gray code.

Between the last element of the first block and the first element of the second block we have
$$d_H\big((g_{2^d-1}, 0), (g_{2^d-1}, 1)\big) = 1,$$
because these vectors differ only in the last coordinate.

For $i = 2^d - 1, \dots, 1$, we consider consecutive vectors $(g_i, 1)$ and $(g_{i-1}, 1)$ in the reversed order. Again,
$$d_H\big((g_i, 1), (g_{i-1}, 1)\big) = d_H(g_i, g_{i-1}) = 1,$$

since reversing the sequence does not affect the pairwise Hamming distances and the last coordinate is the same.

Thus every pair of consecutive vectors in $G_{d+1}$ has Hamming distance 1.

Thus $G_{d+1}$ is a Gray code on $\{0,1\}^{d+1}$. By induction, there exists a Gray code $G_d$ for all $d \in \mathbb{N}$.

(b) Let $d \in \mathbb{N}$ and let

$$G_d = (g_0, g_1, \dots, g_{2^d-1})$$

be a Gray code on $\{0,1\}^d$, i.e. $\{g_0, \dots, g_{2^d-1}\} = \{0,1\}^d$ and $d_H(g_i, g_{i-1}) = 1$ for all $i = 1, \dots, 2^d - 1$.

Let $\pi \in \mathrm{Sym}(d)$ be a permutation of the coordinate indices and define

$$\pi(g_i) := \big(g_{i,\pi(1)}, \dots, g_{i,\pi(d)}\big), \qquad \pi(G_d) := \big(\pi(g_0), \dots, \pi(g_{2^d-1})\big).$$

We first show that consecutive vectors in $\pi(G_d)$ still have Hamming distance 1. For any $i = 1, \dots, 2^d - 1$ we have

$$d_H(\pi(g_i), \pi(g_{i-1})) = \big|\{k : g_{i,\pi(k)} \neq g_{i-1,\pi(k)}\}\big| = \big|\{j : g_{i,j} \neq g_{i-1,j}\}\big| = d_H(g_i, g_{i-1}) = 1.$$

Thus the "distance = 1" property is invariant under permutation of coordinates.

Next, we verify that $\pi(G_d)$ still visits every corner of the $d$-dimensional unit cube exactly once. Since $\{g_0, \dots, g_{2^d-1}\} = \{0,1\}^d$ and $\pi$ is a bijection on $\{0,1\}^d$, the set

$$\{\pi(g_0), \dots, \pi(g_{2^d-1})\}$$

is again equal to $\{0,1\}^d$.

Hence $\pi(G_d)$ satisfies both defining properties of a Gray code: it consists of all $2^d$ binary vectors in $\{0,1\}^d$, and consecutive vectors have Hamming distance 1. Therefore $\pi(G_d)$ is a Gray code for every permutation $\pi \in \mathrm{Sym}(d)$.

## 6.2

We have

$$p(y \mid x) = \frac{p(x,y)}{p(x)}.$$

where

$$\begin{aligned} p(x) = \sum_{y \in \{0,1\}^U} p(x,y) &= \sum_{y \in \{0,1\}^U} \frac{1}{Z(\theta)} \exp(x^\top \Theta y + \theta_x^\top x + \theta_y^\top y) \\ &= \frac{\exp(\theta_x^\top x)}{Z(\theta)} \sum_{y \in \{0,1\}^U} \exp(x^\top \Theta y + \theta_y^\top y) \end{aligned}$$

Therefore

$$p(y \mid x) = \frac{\frac{1}{Z(\theta)} \exp(x^\top \Theta y + \theta_x^\top x + \theta_y^\top y)}{\frac{\exp(\theta_x^\top x)}{Z(\theta)} \sum_{y \in \{0,1\}^U} \exp(x^\top \Theta y + \theta_y^\top y)} = \frac{\exp(x^\top \Theta y) \exp(\theta_y^\top y)}{\sum_{y \in \{0,1\}^U} \exp(x^\top \Theta y) \exp(\theta_y^\top y)}$$

We use

$$\exp(\theta_y^\top y) = \exp(\sum_{j=1}^{U} \theta_{y_j}^\top y_j) = \prod_{j=1}^{U} \exp(\theta_{y_j}^\top y_j)$$

$$\exp(x^\top \Theta y) = \exp(\sum_{j=1}^{U} \sum_{i=1}^{W} x_i \Theta_{i,j} y_j) = \prod_{j=1}^{U} \exp(x^T \Theta_j y_j)$$

Where $\Theta_j$ is the $j_{th}$ column of $\Theta$. Additionally

$$\sum_{y \in \{0,1\}^U} \exp(x^\top \Theta y) \exp(\theta_y^\top y) = \sum_{y_1 \in \{0,1\}} \dots \sum_{y_U \in \{0,1\}} \prod_{j=1}^{U} \exp(x^\top \Theta_j y_j) \exp(\theta_{y_j}^\top y_j)$$

and with

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} b_j$$

$$\sum_{y \in \{0,1\}^U} \exp(x^\top \Theta y) \exp(\theta_y^\top y) = \prod_{j=1}^{U} \sum_{y_j \in \{0,1\}} \exp(x^\top \Theta_j y_j) \exp(\theta_{y_j}^\top y_j)$$

and

$$p(y \mid x) = \prod_{j=1}^{U} \frac{\exp(x^\top \Theta_j y_j) \exp(\theta_{y_j}^\top y_j)}{\sum_{y_j \in \{0,1\}} \exp(x^\top \Theta_j y_j) \exp(\theta_{y_j}^\top y_j)} = \prod_{j=1}^{U} p(y_j \mid x)$$

## 6.3

Choose $k \in \mathbb{N}$ so that the grid spacing $2^{-k}$ satisfies

$$2^{-k} \leq \frac{\varepsilon}{2}.$$

Therefore $k = \lceil \log_2(2/\varepsilon) \rceil$. Then the push-forward $b_k(\mu)$ satisfies

$$W_1(\mu,\, b_k(\mu)) \leq 2^{-k} \leq \frac{\varepsilon}{2}.$$

So it suffices to approximate the discrete measure $\nu := b_k(\mu)$ on the finite grid up to $W_1$-error $\varepsilon/2$. Since [0,1] has diameter 1, we have

$$W_1(\mu, \nu) \leq d_{TV}(\mu, \nu)$$

and with theorem 26.4 we can with $2^{W-1} - 1$ where $W = k = \lceil \log_2(2/\varepsilon) \rceil$ meaning in $\mathcal{O}(\frac{1}{\varepsilon})$ hidden neurons approximate $b_k(\mu)$ with RBM p with

$$W_1(b_k(\mu), p) \leq \frac{\varepsilon}{2}$$

and with triangular inequality

$$W_1(\mu, p) \leq \varepsilon$$

Using $k = \lceil \log_2(2/\varepsilon) \rceil$ visible and in $\mathcal{O}(\frac{1}{\varepsilon})$ hidden neurons, gives us a parameter bound in $\mathcal{O}(\log_2(\frac{1}{\varepsilon})\frac{1}{\varepsilon})$

# Index der Kommentare