

## Homework 2

## 1 True or False Questions

## Problem 1

True.

## Problem 2

True.

## 2 Q &amp; A

## Problem 3

(1) For noisy images, we can (i) set all neurons into the given values and then (ii) run evolution until convergence to get the reconstructed image.

For the masked images, we can (i) set the neurons corresponding to the given parts of the image into the given values, then (ii) run the evolution process with **the given neurons fixed** until convergence to get the image.

(2) Let the matrix  $Y$  be such that  $Y_{ij} = (y_i)_j$ , then since  $y_i$  are orthogonal, we know that  $Y$  is an orthogonal matrix. Thus, we have

$$W_{ij} = \frac{1}{N} \sum_{p=1}^N (y_p)_i (y_p)_j = \frac{1}{N} (Y^T Y)_{ij} = \frac{1}{N} \delta_{ij}.$$

This implies that for any  $y$ , the energy is

$$E = -\frac{1}{2} y^T W y = -\frac{1}{2N} y^T y = -\frac{1}{2},$$

so all patterns are memorized.

## Problem 4

2

The gradient is

$$\begin{aligned}
 \nabla_W L(W) &= -\frac{1}{|P|} \sum_{v \in P} \frac{\nabla_W \mathbb{P}(v)}{\mathbb{P}(v)} \\
 &= -\frac{1}{|P|} \sum_{v \in P} \frac{1}{\mathbb{P}(v)} \sum_h \mathbb{P}(v, h) (\nabla_W \log \mathbb{P}(v, h)) \\
 &= -\frac{1}{|P|} \sum_{v \in P} \frac{1}{\mathbb{P}(v)} \sum_h \mathbb{P}(v, h) \left( yy^T - \frac{\sum_{y'} \exp(y'^T W y') y' y'^T}{\sum_{y'} \exp(y'^T W y')} \right) \\
 &= -\frac{1}{|P|} \sum_{v \in P} \left( \mathbb{E}_{h|v} [yy^T] - \sum_{y'} \frac{\exp(y'^T W y')}{\sum_{y''} \exp(y''^T W y'')} y' y'^T \right) \\
 &= -\frac{1}{|P|} \sum_{v \in P} (\mathbb{E}_{h|v} [yy^T] - \mathbb{E}_{y'} [y' y'^T]),
 \end{aligned}$$

so we are done.

## Problem 5

(1) We have the conditional distribution being

$$\begin{aligned}
 \mathbb{P}(v|h) &= \frac{\mathbb{P}(v, h)}{\int_v \mathbb{P}(v, h) dv} \\
 &= \frac{\exp(-\frac{1}{2}(v-b)^T(v-b) + (v-b)^T W h)}{\int_v \exp(-\frac{1}{2}(v-b)^T(v-b) + (v-b)^T W h) dv}.
 \end{aligned}$$

Now, we let  $x = v - b$ ,  $c = Wh$  and do the integration:

$$\begin{aligned}
 \int \exp\left(-\frac{1}{2}x^T x + x^T c\right) dx &= \int \int \cdots \int \exp\left(-\frac{1}{2} \sum_{i=1}^{N_v} x_i^2 + \sum_{i=1}^{N_v} x_i c_i\right) dx_1 dx_2 \cdots dx_n \\
 &= \exp\left(\frac{1}{2} \sum_{i=1}^{N_v} c_i^2\right) \cdot (2\pi)^{\frac{N_v}{2}}.
 \end{aligned}$$

Thus, we have

$$\begin{aligned}
 \mathbb{P}(v|h) &= (2\pi)^{-\frac{N_v}{2}} \exp\left(-\frac{1}{2}h^T W^T W h - \frac{1}{2}(v-b)^T(v-b) + (v-b)^T W h\right) \\
 &= (2\pi)^{-\frac{N_v}{2}} \exp\left(-\frac{1}{2} \|v-b-Wh\|^2\right).
 \end{aligned}$$

(2) The loss function is

3

$$\begin{aligned}
L(b) &= -\frac{1}{|P|} \sum_{v \in P} \log \left( \sum_h \mathbb{P}(v, h) \right) \\
&= -\frac{1}{|P|} \sum_{v \in P} \log \left( \frac{1}{Z} \sum_h \exp \left( -\frac{1}{2}(v-b)^T(v-b) + v^T W h \right) \right) \\
&= -\frac{1}{|P|} \sum_{v \in P} \log \left( \sum_h \exp \left( -\frac{1}{2}(v-b)^T(v-b) + v^T W h \right) \right) + \log Z
\end{aligned}$$

Then we can compute the gradient:

$$\begin{aligned}
\nabla_b L(b) &= -\frac{1}{|P|} \sum_{v \in P} \frac{\sum_h \exp \left( -\frac{1}{2}(v-b)^T(v-b) + v^T W h \right) (v-b)}{\sum_h \exp \left( -\frac{1}{2}(v-b)^T(v-b) + v^T W h \right)} \\
&\quad + \frac{\sum_{v,h} \exp \left( -\frac{1}{2}(v-b)^T(v-b) + v^T W h \right) (v-b)}{Z} \\
&= -\frac{1}{|P|} \sum_{v \in P} (v-b) + \mathbb{E}_{v,h}[v-b] \\
&= -\frac{1}{|P|} \sum_{v \in P} v + \mathbb{E}_{v,h}[v]
\end{aligned}$$

## Problem 6

(1) No. Suppose that the probability distribution is given such that each adjacent neuron must have the same values, then we must have  $D = A = E = F$  for the joint probability  $\mathbb{P}(D, F)$  to be nonzero. Thus,  $D$  and  $F$  must not be independent.

(2) Yes. Ignoring the normalizing factor, we have

$$\begin{aligned}
\mathbb{P}(B, E|A) &= \frac{1}{Z} \int_{CDF} f_{AD}(A, D) f_{AC}(A, C) f_{AE}(A, E) f_{BC}(B, C) f_{EF}(E, F) dC dD dF \\
&= \frac{1}{Z} f_{AE}(A, E) l(A, B) m(A) h(E),
\end{aligned}$$

where  $m, h$  are functions only depending on  $A, E$ , respectively;  $l$  is a function of  $A, B$ . We then find that  $B$  and  $E$  are independent given  $A$ .

(3) We have

$$\log \mathbb{P}(A, B, C, D, E, F) = -\mathcal{E}(A, B, C, D, E, F) + C_2,$$

where  $C_2$  is a constant independent of  $A, B, \dots, F$ . Thus, we can have

4

$$\begin{aligned}\mathcal{E}(A, B, C, D, E, F) &= C_1 - \log f_{AD}(A, D) - \log f_{AC}(A, C) - \log f_{AE}(A, E) - \log f_{BC}(B, C) - \log f_{EF}(E, F) \\ &= \mathcal{E}_{AD}(A, D) + \mathcal{E}_{AC}(A, C) + \mathcal{E}_{AE}(A, E) + \mathcal{E}_{BC}(B, C) + \mathcal{E}_{EF}(E, F).\end{aligned}$$