Homework 2

Deep Learning 2024 Spring

Due 11:59pm, 2024/4/13

1 True or False

Problem 1. Generative models can be used to both classify and generate images.

Problem 2. We can train an energy-based model without knowing the explicit density function (or normalizing factor Z).

2 Q&A

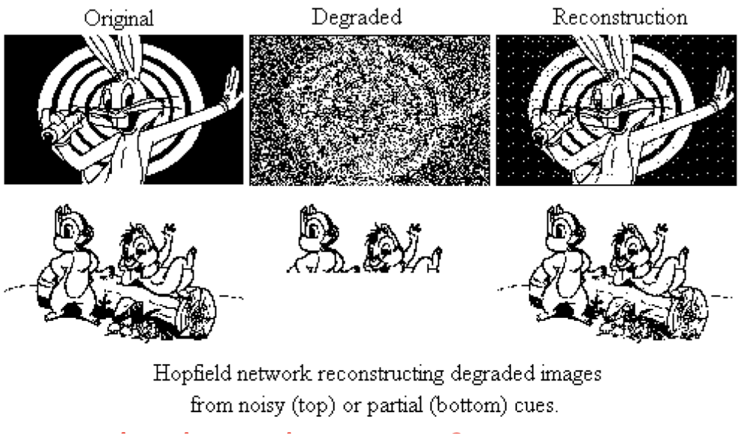

Figure 1: Noisy image (top row) and masked image (bottom row).

Problem 3. (Hopfield Network) Answer the following questions about the Hopfield network.

1. Figure 1 shows two types of degraded images: noisy image and masked image. Design an appropriate process to retrieve stored patterns using the Hopfield network for each case respectively.

   (Hint: The unmasked part of a masked image is the same as ground truth. By contrast, most pixels of a noisy image are different from the ground truth.)

2. Suppose we are using the Hebbian learning rule to train Hopfield networks. Prove that with N orthogonal patterns y_p (y_p is a N-dim vector) for $p = 1, \dots, N$, the Hopfield network can memorize all 2^N patterns. The Hebbian learning rule is given by $W = \frac{1}{N}\sum_p y_p y_p^T$.

Problem 4. (Boltzman Machine) Consider a fully connected Boltzman machine. We remark the visible units as v, the hidden units as h, and all units $y = (v, h)$. The joint probability of v and h is given by

$$\mathbb{P}(v, h) = \frac{\exp(y^T W y)}{\sum_{y'} \exp(y'^T W y')}. \tag{1}$$

And the marginal probability of v is given by

$$\mathbb{P}(v) = \sum_h \mathbb{P}(v, h). \tag{2}$$

We aim to maximize the log-likelihood, and the loss is given by

$$L(W) = -\frac{1}{|P|} \sum_{v \in P} \log \mathbb{P}(v). \tag{3}$$

Prove that the gradient of Eq. (3) has the following form:

$$\nabla_W \text{loss}(W) = -\frac{1}{|P|} \sum_{v \in P} \left(\mathbb{E}_{h|v} \left[y y^T \right] - \mathbb{E}_{y'} \left[y' y'^T \right] \right).$$

Problem 5. (Gaussian RBM) Consider a restricted Boltzman machine with a single hidden layer and the following energy function $\mathcal{E}_{W,b} : \mathbb{R}^{N_h + N_v} \to \mathbb{R}$:

$$\mathcal{E}_{W,b}(v, h) = \frac{1}{2}(v - b)^T (v - b) - v^T W h$$

where W, b are trainable parameters, v is visible continuous-value units (i.e., $v \in \mathbb{R}^{N_v}$), and h is hidden discrete-value units (i.e., $h \in \{-1, 1\}^{N_h}$).

1. Derive the conditional distribution $\mathbb{P}(v|h)$.
2. Derive the gradient of b if we train this model based on the maximum log-likelihood principle. (Hint: Your answer should contain the form of an expectation.)

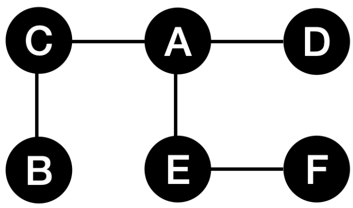

Figure 2: An example of undirected probabilistic model.

Problem 6. (Undirected Probabilistic Model) A **graphical model** or **probabilistic graphical model** or **structured probabilistic model** is a probabilistic model for which a graph expresses the conditional dependence structure between random variables. [1] In an **undirected graphical model**, an edge implies dependence between the corresponding random variables. Figure 2 shows an example, where the joint probability

[1] https://en.wikipedia.org/wiki/Graphical_model

distribution can be factorized as

$$\mathbb{P}(A, B, C, D, E, F) = \frac{1}{Z} f_{AD}(A, D) f_{AC}(A, C) f_{AE}(A, E) f_{BC}(B, C) f_{EF}(E, F)$$

for some non-negative functions f_{AB}, f_{AC}, f_{AD}, f_{AE}, f_{BC}, and f_{EF}, and a normalizing factor Z (also called partition function).

1. Are D and F independent?

2. Write down the unnormalized conditional distribution $\mathbb{P}(B, E|A)$. "unnormalized" means you can omit the normalizing factor. Are B and E independent given A?

3. We model $\mathbb{P}(A, B, C, D, E, F)$ as a Boltzman distribution

$$\mathbb{P}(A, B, C, D, E, F) \propto \exp(-\mathcal{E}(A, B, C, D, E, F))$$

   where $\mathcal{E}$ is the energy function. Show that the energy function can be expressed by the following factorization:

$$\mathcal{E}(A, B, C, D, E, F) = \mathcal{E}_{AC}(A, C) + \mathcal{E}_{AD}(A, D) + \mathcal{E}_{AE}(A, E) + \mathcal{E}_{BC}(B, C) + \mathcal{E}_{EF}(E, F)$$