Generative Adversarial Networks (GANs) & Restricted Boltzmann Machines (RBMs)

Lecture 23

Changho Suh

October 7, 2021

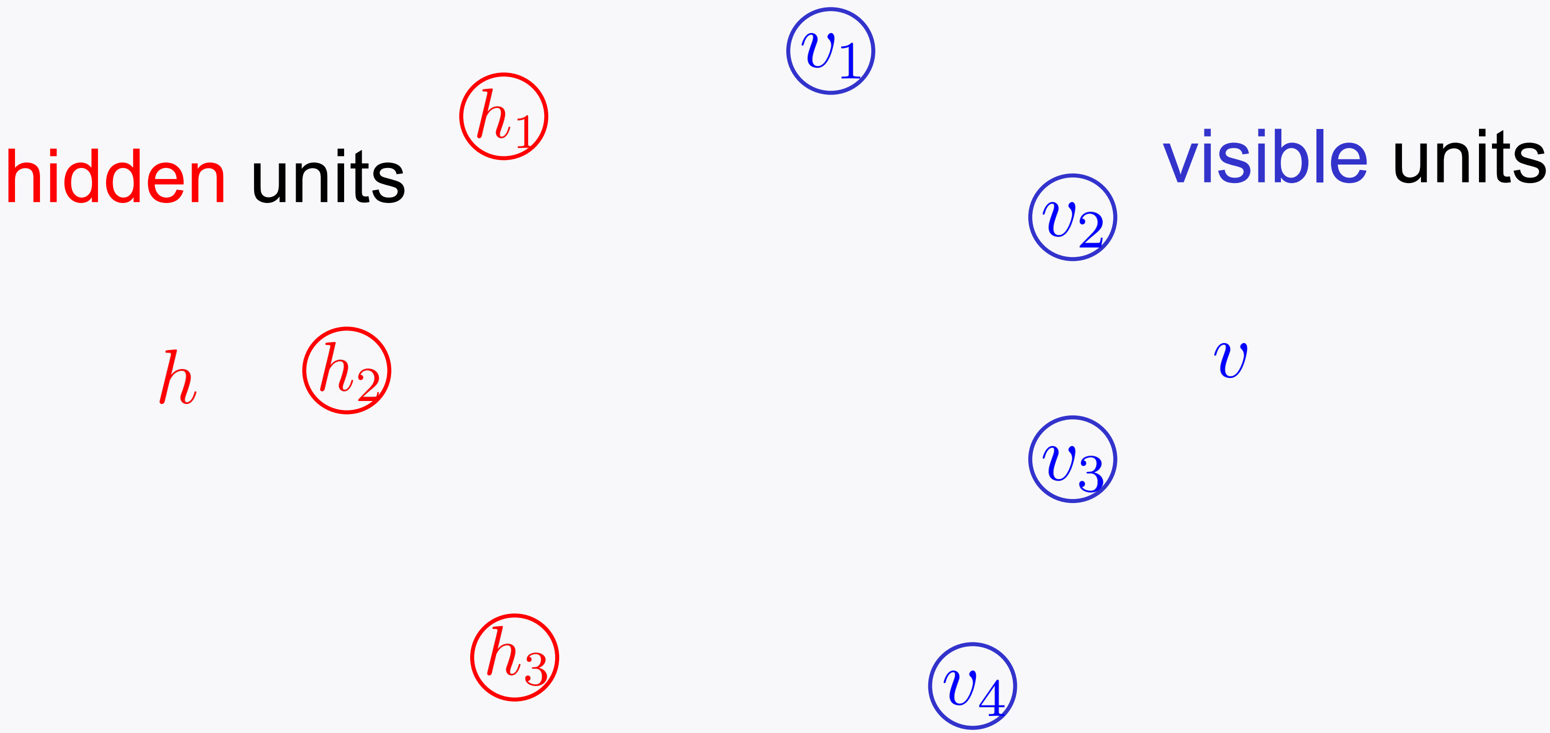
Outline

1. Figure out Boltzmann Machine (BM) and then RBM.

2. Study how RBM can serve as a generative model.

3. Study a couple of concepts regarding RBM.

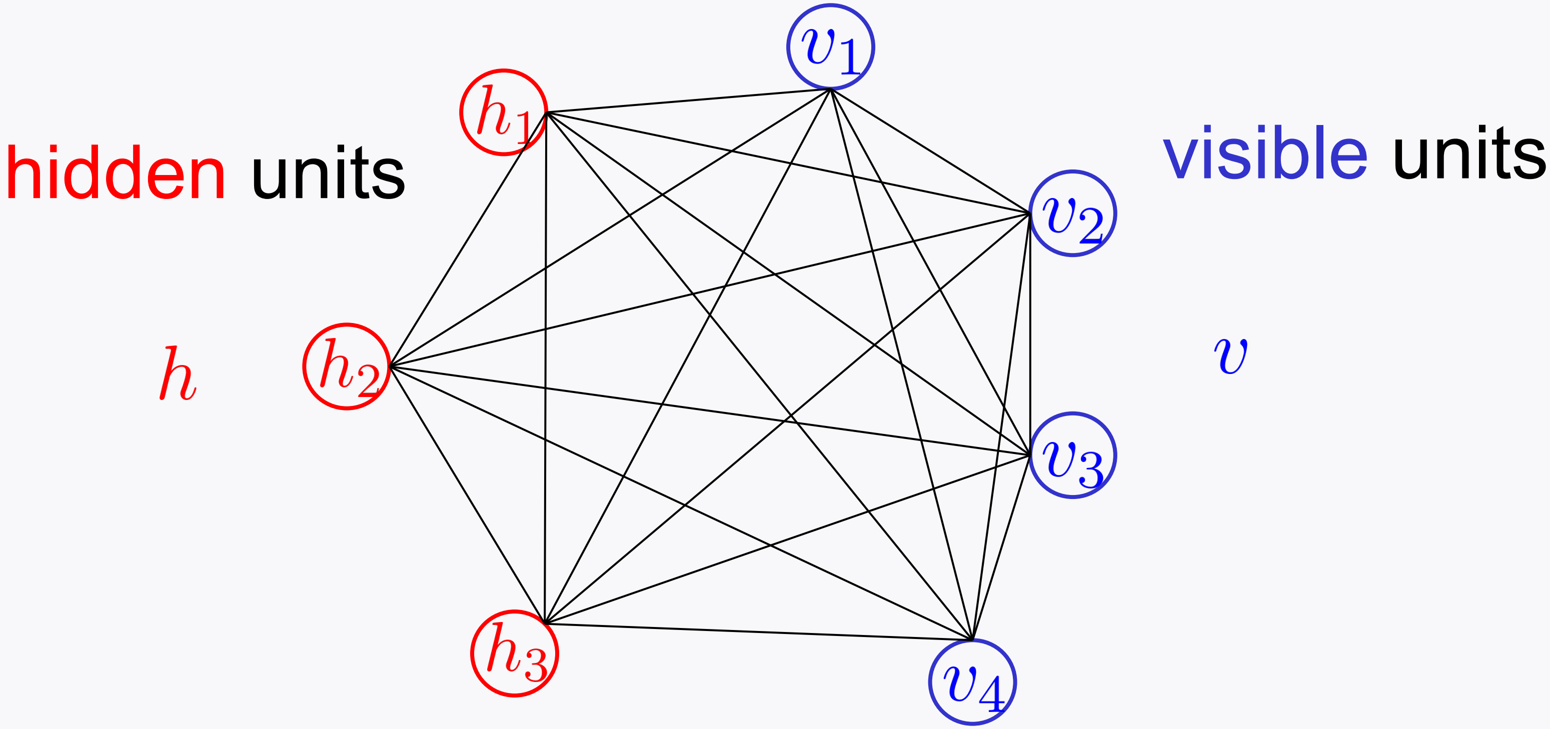
Boltzmann Machine (BM)



Represent relationship between nodes often via edges.

Suppose: All of the hidden and visible units are arbitrarily correlated.

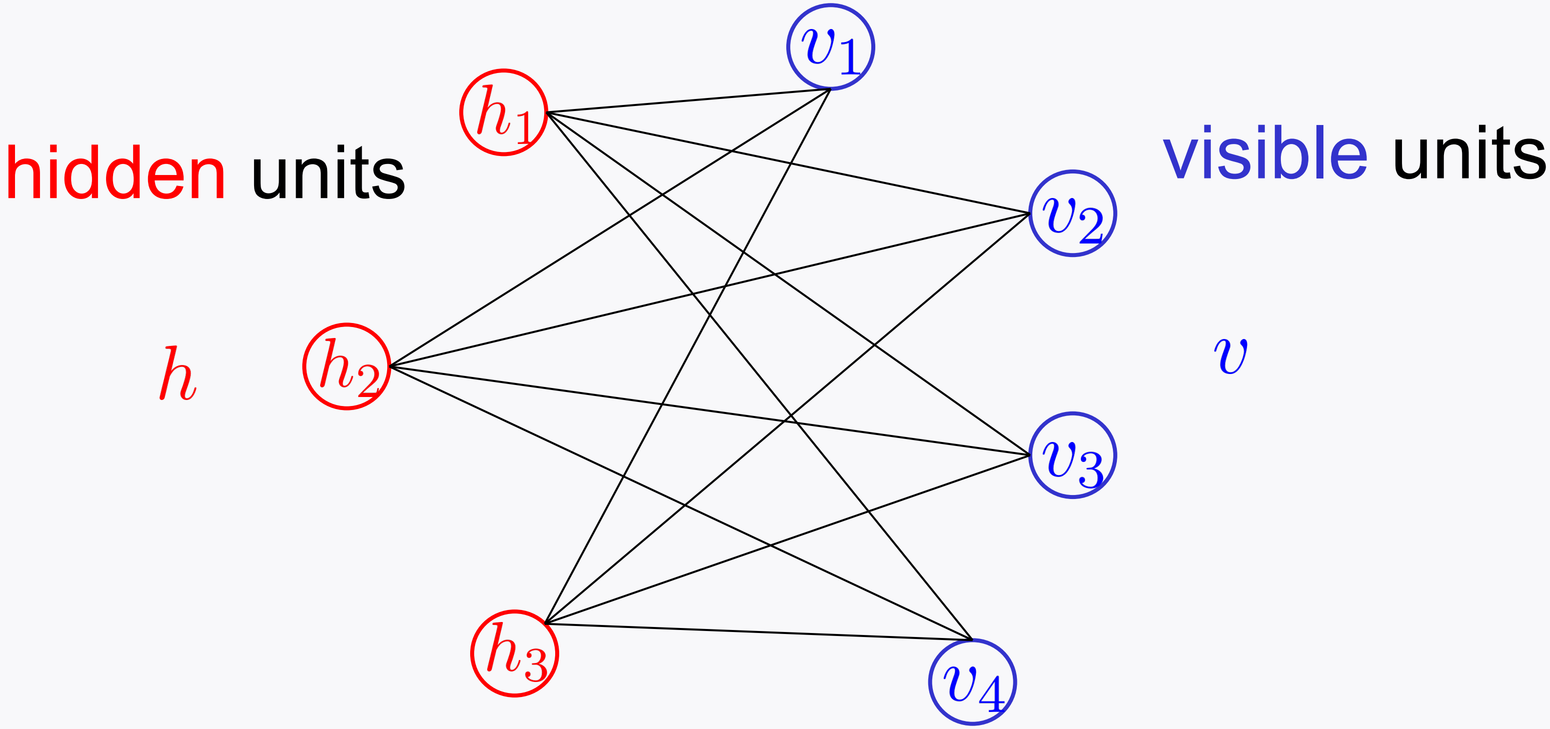
Boltzmann Machine (BM)



The above picture captures such arbitrary distribution:

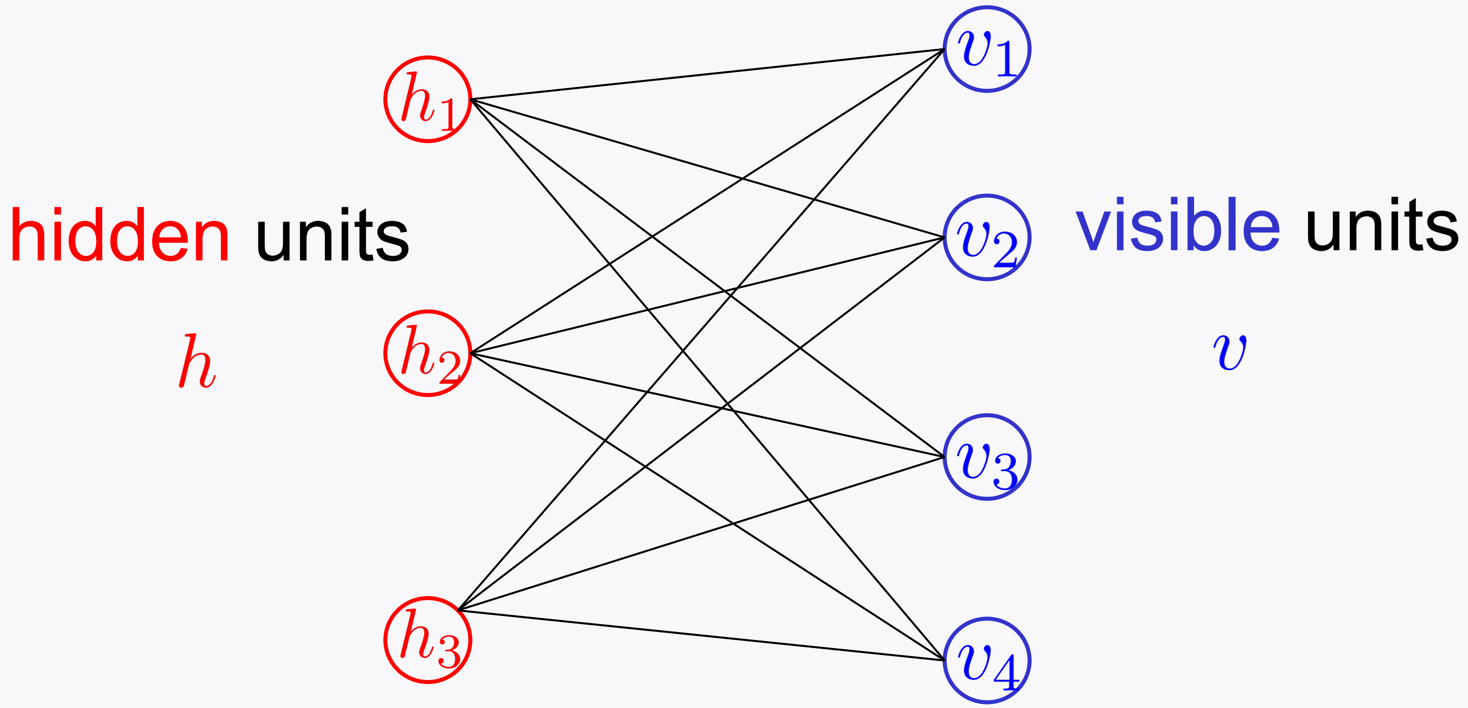
$$\mathbb{P}(h, v)$$

Restricted Boltzmann Machine (RBM)



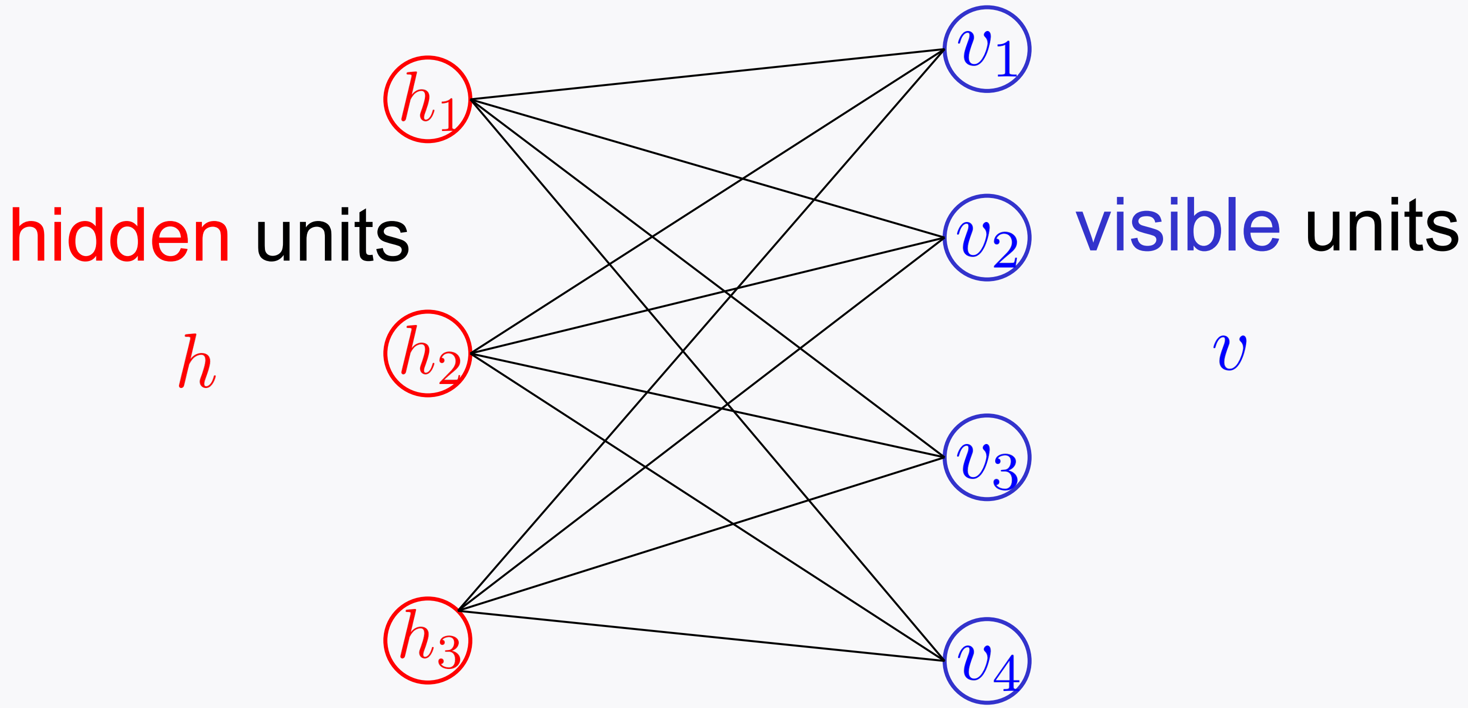
RBM is a simplified BM in which there is no edge within hidden units as well as within visible units.

Representation of RBM



Usually it is represented a **bipartite** graph.

Property of RBM



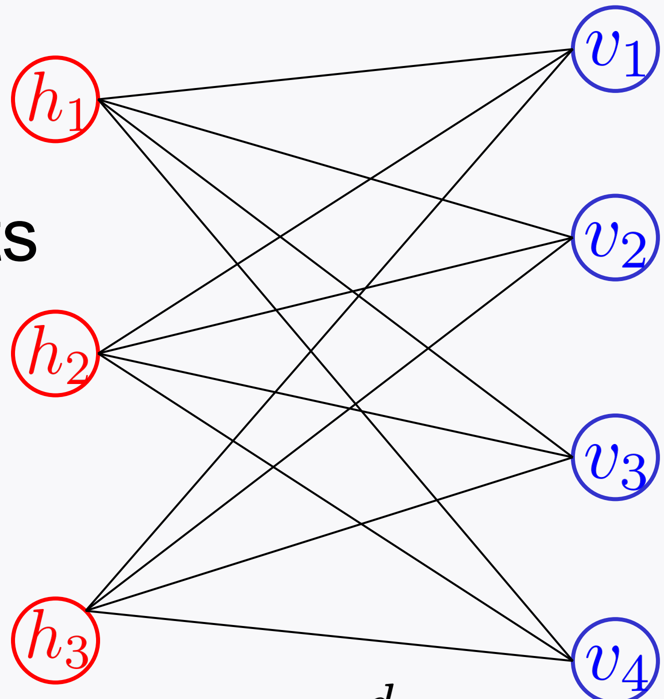
Given the visible units, hidden units are independent:

$$\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$$

Property of RBM

hidden units

h



visible units

v

$$\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$$

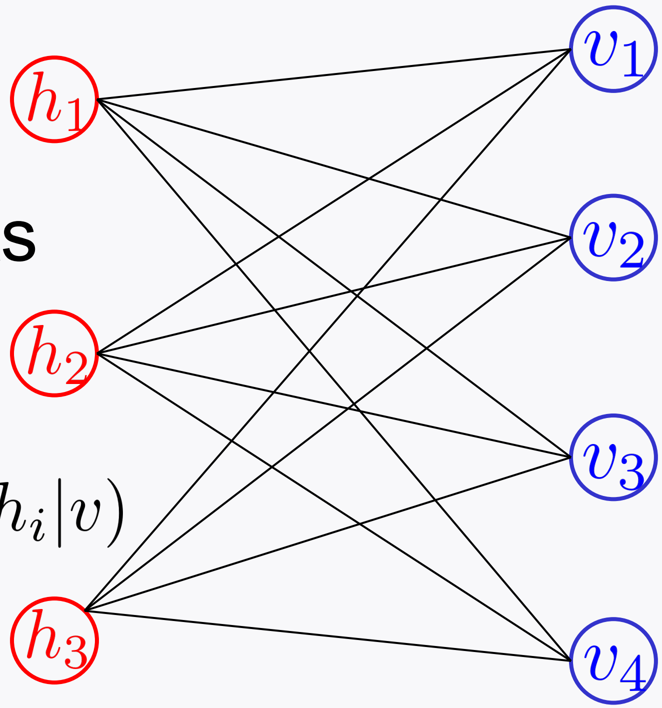
Similarly: $$\mathbb{P}(v|h) = \prod_{i=1}^{n} \mathbb{P}(v_i|h)$$

Main role of RBM

hidden units

h

$$\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$$



visible units

v

$$\mathbb{P}(v|h) = \prod_{i=1}^{n} \mathbb{P}(v_i|h)$$

Can serve as a **generative model**.

When only the visible units are available, can generate hidden units via $\prod_{i=1}^{d} \mathbb{P}(h_i|v)$.

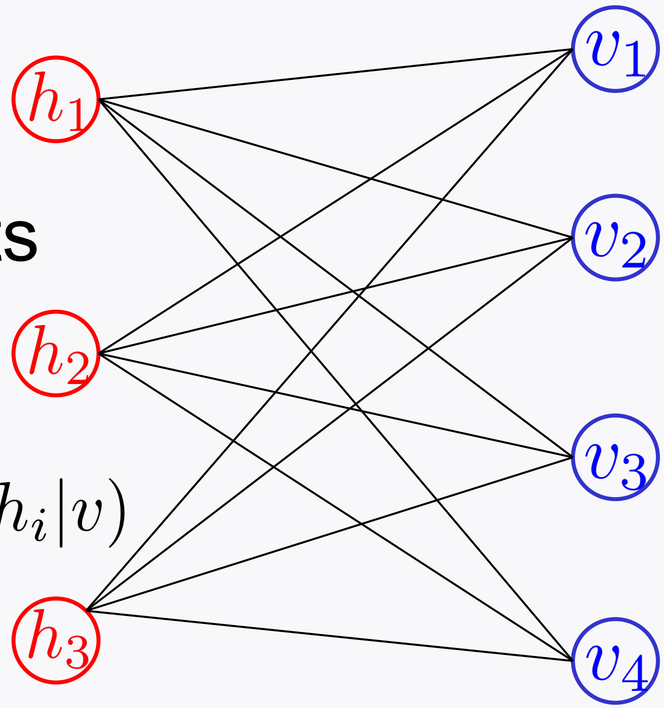
Use of RBMs in practice

modality 2

hidden units

h

$$\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$$



modality 1

visible units

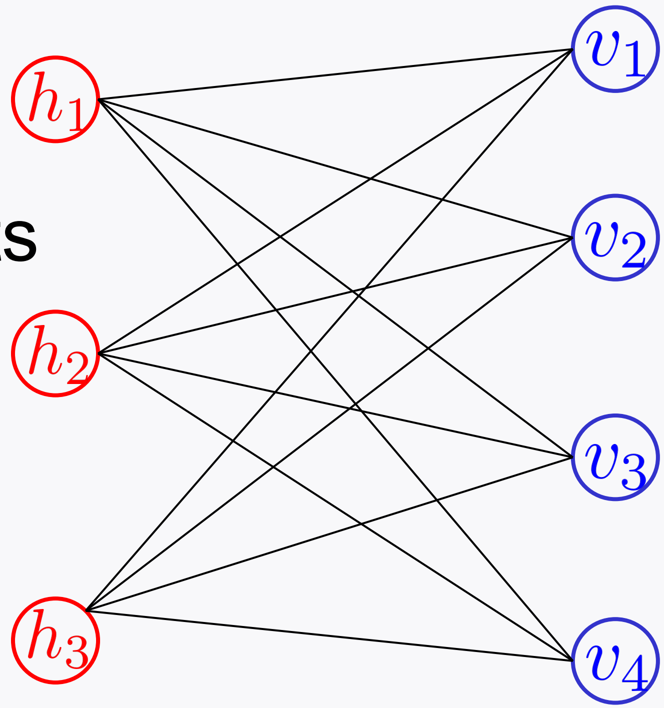
v

$$\mathbb{P}(v|h) = \prod_{i=1}^{n} \mathbb{P}(v_i|h)$$

Probability distribution in RBMs?

hidden units

h



visible units

v

In RBM, the description of probability distribution requires one important concept:

Energy

Energy

Energy is a key function that determines the probability distribution.

It is denoted by $E(v, h)$ and defined such that:

$$\mathbb{P}(v, h) = \frac{e^{-E(v,h)}}{Z} \qquad \text{where } Z = \sum_{v} \sum_{h} e^{-E(v,h)}$$

Interpretation:

Lower energy → more stable → more probable

Energy of visible units

$$\mathbb{P}(v) = \sum_h \mathbb{P}(v,h)$$ total probability law

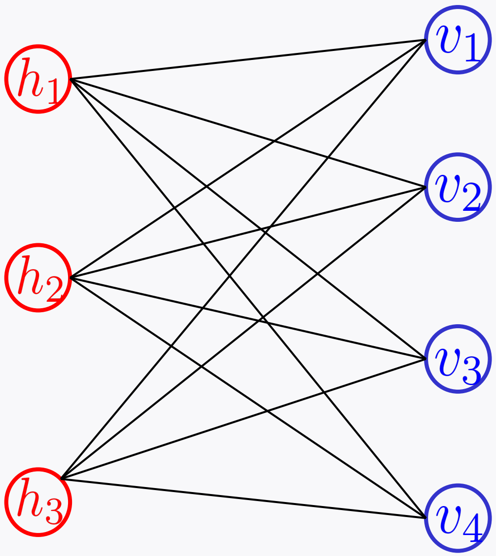
$$= \sum_h \frac{e^{-E(v,h)}}{Z}$$

Wish to find energy of v, say $F(v)$, such that

$$\mathbb{P}(v) = \sum_h \frac{e^{-E(v,h)}}{Z} = \frac{e^{-F(v)}}{Z}$$

$$F(v) = -\log\left(\sum_h e^{-E(v,h)}\right)$$ Called "**Free Energy**".

How can RBM serve as a generative model?



How to compute: $\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$?

In RBM, we define $E(v,h)$ as:

$$E(v,h) := -b^T v - c^T h - h^T W v$$

- c: bias w.r.t. hidden units $\in \mathbb{R}^d$
- W: weight matrix $\in \mathbb{R}^{d \times n}$
- b: bias w.r.t. visible units $\in \mathbb{R}^n$

Computation of $\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$?

$$E(v,h) := -b^T v - c^T h - h^T W v$$

$$\mathbb{P}(v,h) = \frac{e^{-E(v,h)}}{Z}$$

$$\mathbb{P}(h|v) = \frac{\mathbb{P}(v,h)}{\mathbb{P}(v)} = \frac{\mathbb{P}(v,h)}{\sum_h \mathbb{P}(v,h)} = \frac{e^{-E(v,h)}}{\sum_h e^{-E(v,h)}}$$

$$= \frac{e^{b^T v + c^T h + h^T W v}}{\sum_h e^{b^T v + c^T h + h^T W v}}$$

$$= \frac{e^{c^T h + h^T W v}}{\sum_h e^{c^T h + h^T W v}} = \frac{\prod_{i=1}^{d} e^{c_i h_i + h_i W_i v}}{\sum_h e^{c^T h + h^T W v}}$$

$$W = \begin{bmatrix} | \\ W_i \\ | \end{bmatrix}$$

Binary case: $\mathbb{P}(h|v) = \prod_{i=1}^{d} \mathbb{P}(h_i|v)$?

$$\mathbb{P}(h_i|v) \sim e^{c_i h_i + h_i W_i v}$$

$$\mathbb{P}(h_i = 1|v) \sim e^{c_i + W_i v}$$

$$\mathbb{P}(h_i = 0|v) \sim e^{0}$$

$$\mathbb{P}(h_i = 1|v) = \frac{e^{c_i + W_i v}}{1 + e^{c_i + W_i v}}$$

$$= \frac{1}{1 + e^{-c_i - W_i v}} = \sigma(c_i + W_i v)$$

Conditional probabilities

$$\mathbb{P}(h|v) = \frac{e^{c^T h + h^T W v}}{\sum_h e^{c^T h + h^T W v}}$$

$$\mathbb{P}(v|h) = \frac{e^{b^T v + v^T W^T h}}{\sum_v e^{b^T v + v^T W^T h}}$$

Binary case:

$$\mathbb{P}(h_i = 1|v) = \sigma(c_i + W_i v)$$

$$\mathbb{P}(v_i = 1|h) = \sigma(b_i + [W^T]_i h)$$

To compute this, need to figure out:

$$\theta := (W, b, c) \quad \text{parameters}$$

How to train parameters θ?

Look ahead

Will study how to train RBM.