Hidden Nodes

Visible Nodes

Hidden Nodes →

Visible Nodes →

Energy Function

$$z_i = b_i + \sum_j s_j w_{ij}$$

Probability Form

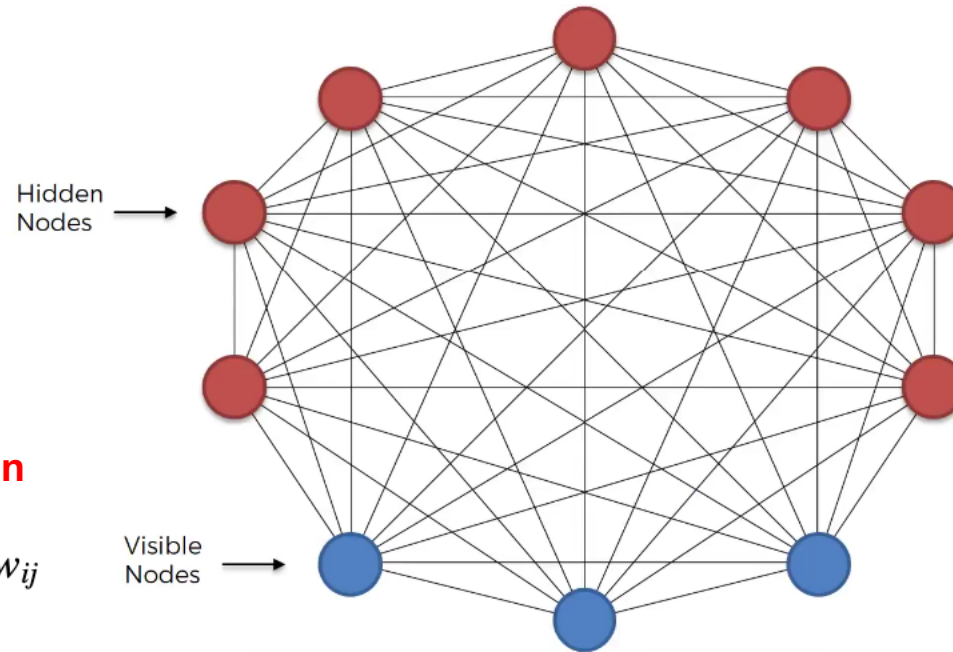$$prob(s_i = 1) \;\; = \;\; \frac{1}{1 + e^{-z_i}}$$

Hidden Nodes →

Visible Nodes →

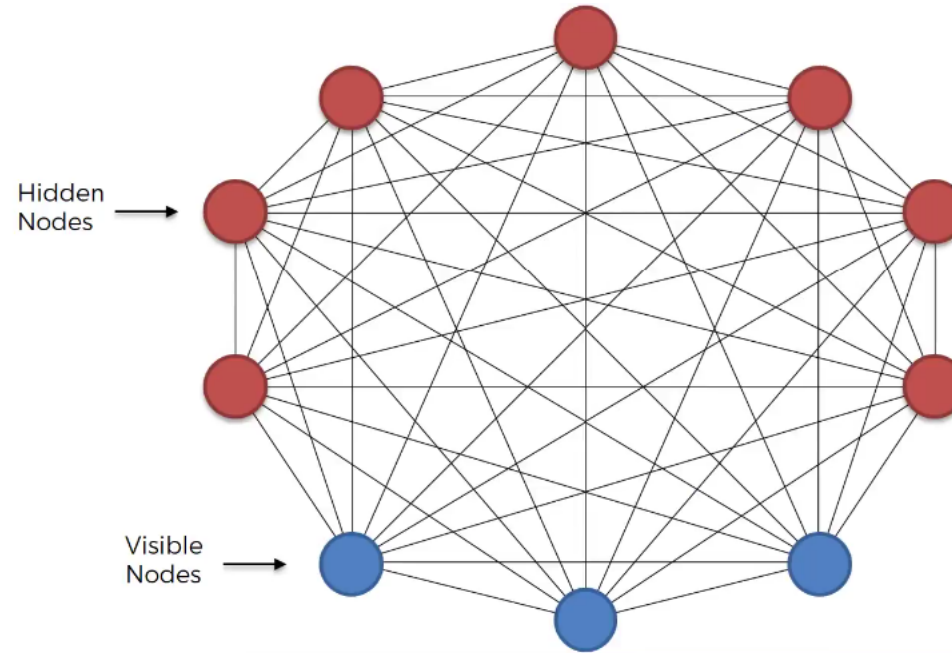**Energy Function**

$$z_i = b_i + \sum_j s_j w_{ij}$$

Probability Form

$$prob(s_i = 1) = \frac{1}{1 + e^{-z_i}}$$

Energy State of v

$$E(\mathbf{v}) = -\sum_i s_i^{\mathbf{v}} b_i - \sum_{i<j} s_i^{\mathbf{v}} s_j^{\mathbf{v}} w_{ij}$$

$$P(\mathbf{v}) = e^{-E(\mathbf{v})} / \sum_{\mathbf{u}} e^{-E(\mathbf{u})}$$

Hidden Nodes →

Visible Nodes →

Energy State of v

$$E(\mathbf{v}) = -\sum_i s_i^{\mathbf{v}} b_i - \sum_{i<j} s_i^{\mathbf{v}} s_j^{\mathbf{v}} w_{ij}$$

$$P(\mathbf{v}) = e^{-E(\mathbf{v})} / \sum_{\mathbf{u}}{}' e^{-E(\mathbf{u})}$$

Learning without hidden layers

$$\left\langle \frac{\partial \log P(\mathbf{v})}{\partial w_{ij}} \right\rangle_{\text{data}} = \langle s_i s_j \rangle_{\text{data}} - \langle s_i s_j \rangle_{\text{model}}$$
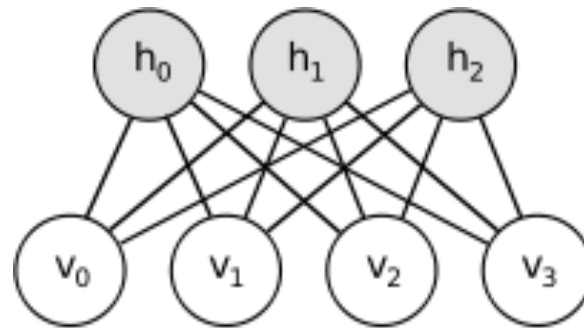
Energy Based Model Perspective
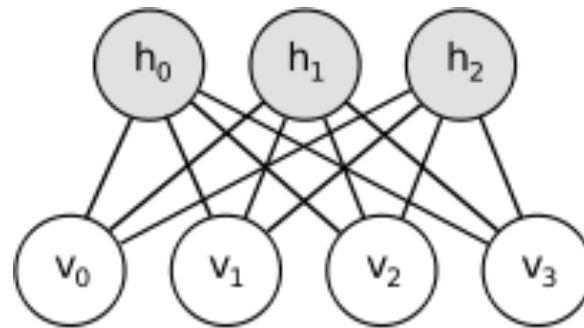
RBM Energy Form

$$E(v, h) = -b'v - c'h - h'Wv$$

$$\mathcal{F}(v) = -b'v - \sum_i \log \sum_{h_i} e^{h_i(c_i + W_i v)}.$$

Energy Based Model Perspective

Energy Based Model Function

RBM Energy Form

$$p(x) = \frac{e^{-E(x)}}{Z}. \quad Z = \sum_x e^{-E(x)}$$

$$E(v, h) = -b'v - c'h - h'Wv$$

$$\mathcal{F}(x) = -\log \sum_h e^{-E(x,h)}$$ **Free Energy Function**

$$\mathcal{F}(v) = -b'v - \sum_i \log \sum_{h_i} e^{h_i(c_i + W_i v)}.$$

Probability Perspective
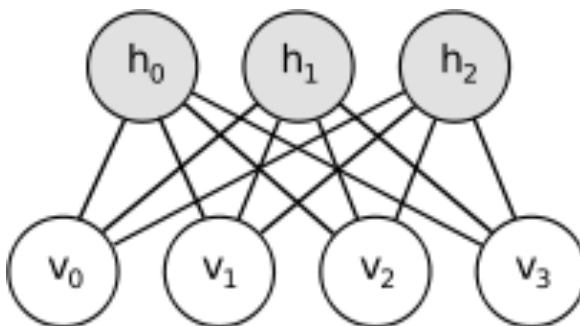
RBM Energy Form

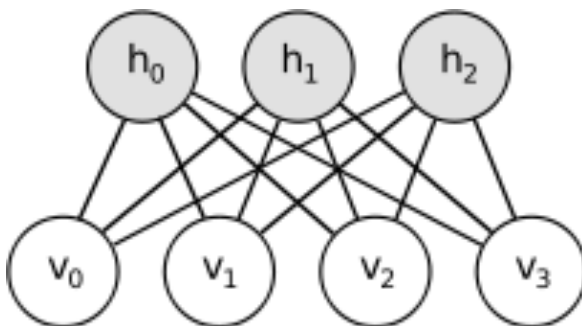$$E(v, h) = -b'v - c'h - h'Wv$$

$$\mathcal{F}(v) = -b'v - \sum_i \log \sum_{h_i} e^{h_i(c_i + W_i v)}.$$

Conditional Probability

$$p(h|v) = \prod_i p(h_i|v)$$

$$p(v|h) = \prod_j p(v_j|h).$$

Probability Perspective

$$p(h|v) = \prod_i p(h_i|v)$$

$$p(v|h) = \prod_j p(v_j|h).$$

$$P(v_j = 1|h) = sigm(b_j + W'_j h)$$

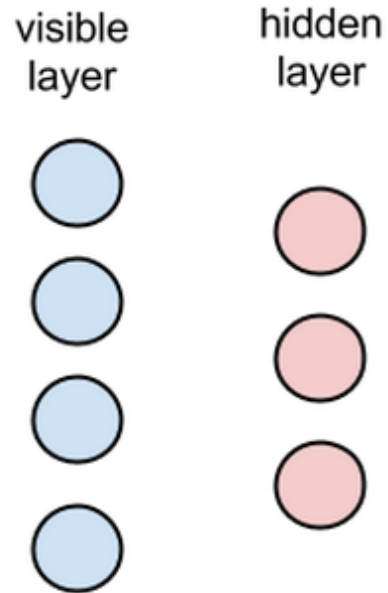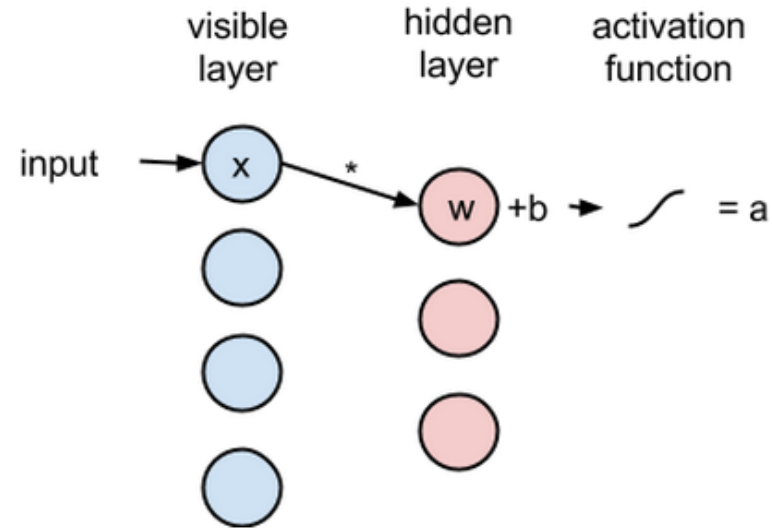$$P(v_j = 1|h) = sigm(b_j + W'_j h)$$

RBM Free Energy under Sigmoid

$$\mathcal{F}(v) = -b'v - \sum_i \log(1 + e^{(c_i + W_i v)}).$$
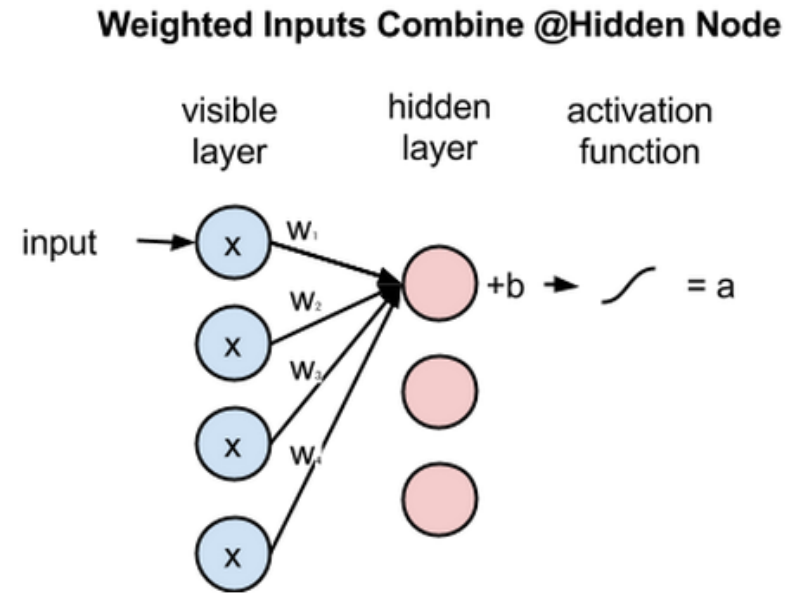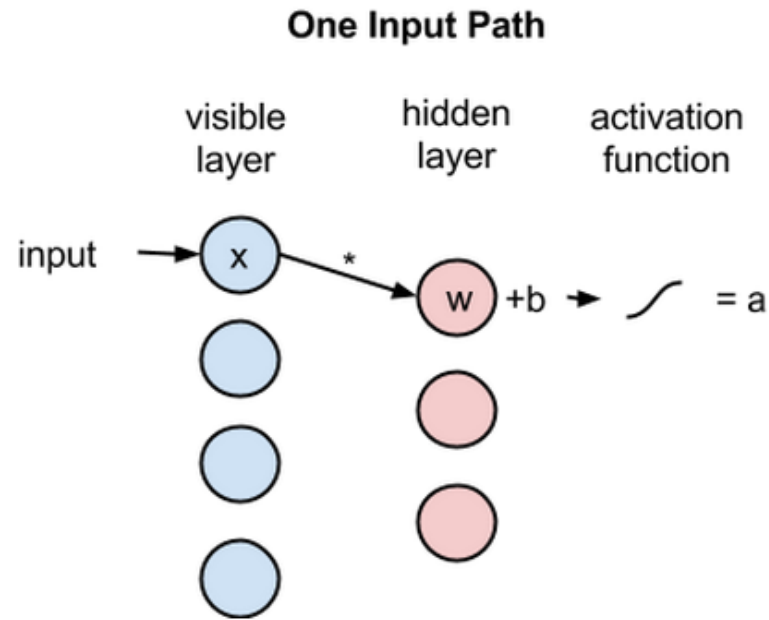
Deep Network Perspective

Deep Network Perspective

Deep Network Perspective

Reconstruction

Multiple Hidden Layers

visible layer — hidden layer 1 — activation function — hidden layer 2

input

x $+b \to \int = a$

x $+b \to \int = a$

x $+b \to \int = a$

x

$w_i \dots w_n$

Reconstruction

these biases are new

visible layer — hidden layer 1

$r = b +$

reconstructions are the new output $r = b +$

$r = b +$

$r = b +$

activations are the new input

a

a

a

$w_i \dots w_n$

weights are the same
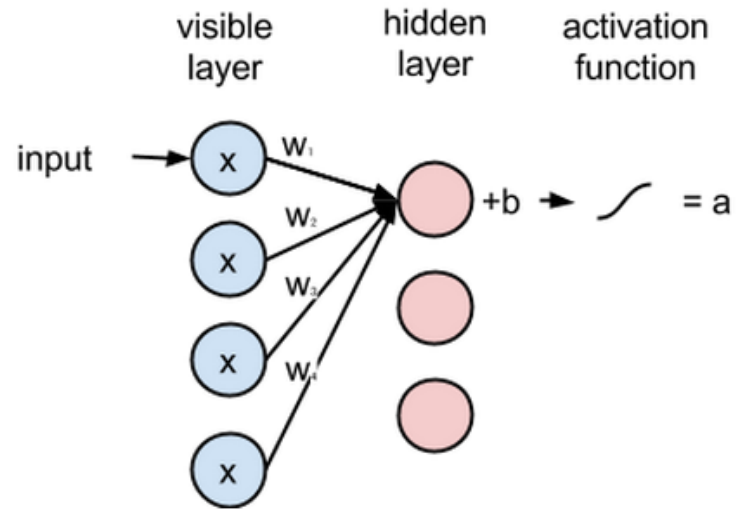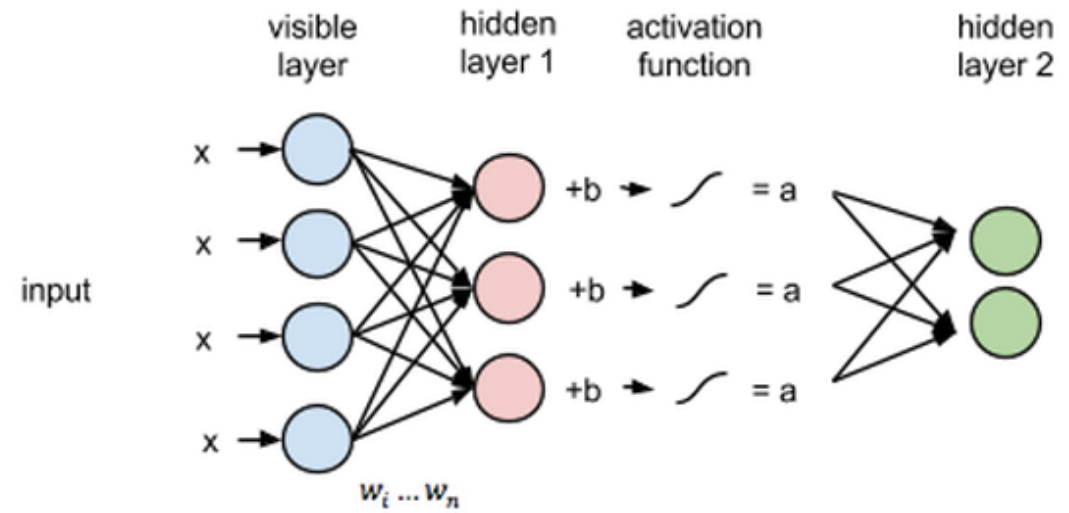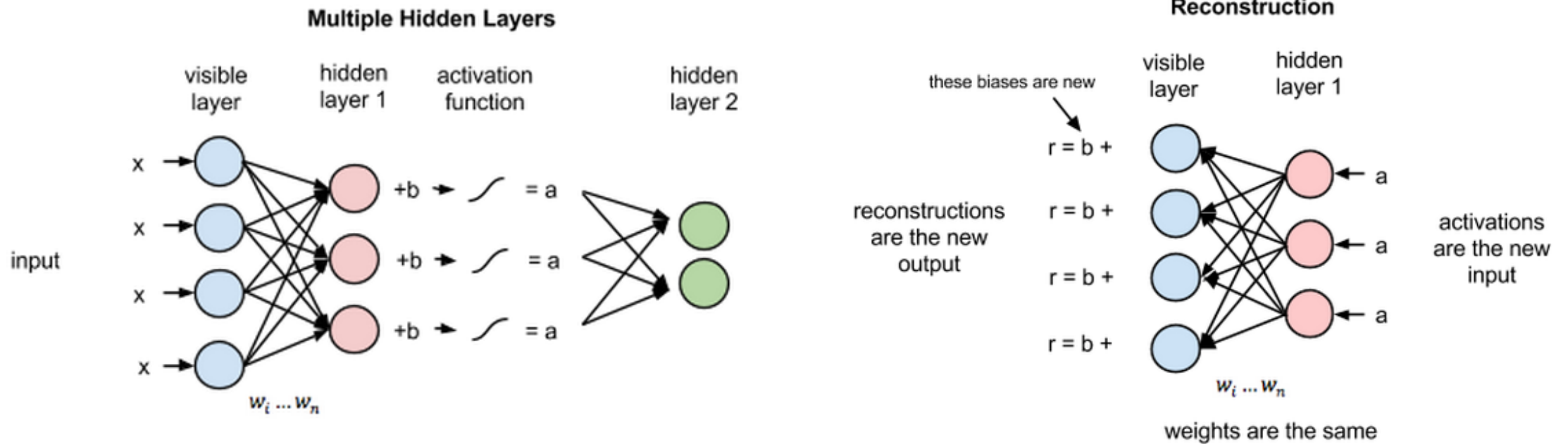
$P(a|x; w) \Longrightarrow P(x, a) \Longleftarrow P(x|a; w)$

Reconstruction