What is sparse coding and why it works

① Undirected graphical models (Markov Random Fields, Energy models)

② Restricted Boltzmann Machine (Generative vs discriminative = undirected vs directed?)

③ Deep Belief Networks

④ Hidden nodes in terms of visible nodes? (Sigmoid function)

⑤ Sparse Coding

CRASH COURSE ON GRAPHICAL MODELS

$X_1, X_2, \dots, X_s$ random variables; $X = (X_1, \dots, X_s)$

$X_i \in \{1, 2, \dots, k_i\}$

$S \subseteq \{1, 2, \dots, s\}$, $X_S := (X_i)_{i \in S}$

$G = (V, E)$ directed acyclic graph

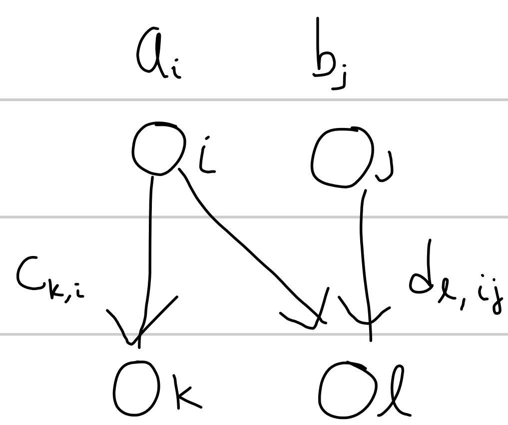
$pa(i) = \{j : j \to i \text{ is an edge}\}$ parents of node i

A prob dist on X factors according to G if

$$p(X = x) = \prod_{i=1}^{s} p(X_i = x_i \mid X_{pa(i)} = x_{pa(i)})$$

Directed Graphical Model on G is family of prob dists which factor according to G. BAYESIAN NETWORK.

Parameters $\theta_{x, x_{pa(i)}} = P(X_i = x_i \mid X_{pa(i)} = x_{pa(i)})$



$P_{ijkl} = a_i c_{k,i} b_j d_{l,ij}$

$i, j, k, l \in \{0, 1\}$

$c_{0,i} + c_{1,i} = 1 \quad \forall i \qquad a_0 + a_1 = 1$

$d_{0,ij} + d_{1,ij} = 1 \quad \forall i, j \qquad b_0 + b_1 = 1$

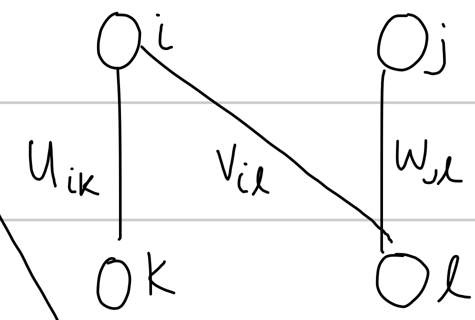
CAUSAL STRUCTURE

$G=(V,E)$ undirected, simple graph (no loop, multiedges)

A clique is a complete subgraph. $\mathcal{C}$ set of (max) cliques.

A prob dist factors according to G if

$$p(X=x) = \frac{1}{Z(W)} \prod_{C \in \mathcal{C}} W^{(c)}_{x_c}$$



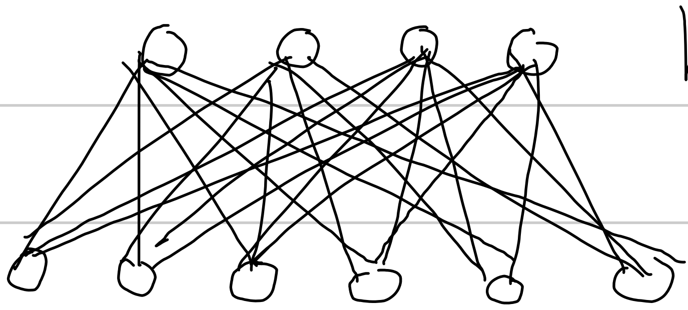
$$P_{ijkl} = \frac{1}{Z} U_{ik} V_{il} W_{jl}$$

Undirected Graphical Model on G is family of all prob dists which factor according to G. MARKOV RANDOM FIELD. ENERGY MODELS.

CORRELATION STRUCTURE

BOLTZMANN MACHINE = Undirected graphical model with binary variables

RESTRICTED BOLTZMANN MACHINE = Boltzmann machine on bipartite graphs



Write parameters $\omega_{ij} = e^{W_{ij}}$, $\beta_i = e^{b_i}$, $\gamma_j = e^{c_j}$

$h_i \in \{0,1\}$, $v_j \in \{0,1\}$

$$p(v,h) = \frac{1}{Z} \prod_{i,j} \omega_{ij}^{h_i v_j} \prod_i \beta_i^{h_i} \prod_j \gamma_j^{v_j}$$

$$= \frac{1}{Z} e^{\sum_{ij} h_i W_{ij} v_j + \sum_i b_i h_i + \sum_j c_j v_j}$$

$$= \frac{1}{Z} \exp(h^T W v + b^T h + c^T v)$$

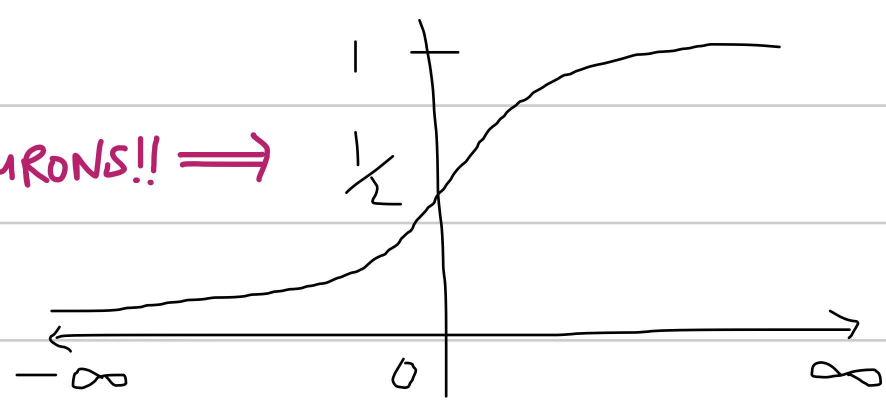
$$p(h_i=1 \mid v) = p(v, h_i=1) / p(v)$$

$$= \sum_{h: h_i=1} p(v,h) \Big/ \sum_h p(v,h)$$

$$= \text{Sigmoid}(c_i + W_i v)$$

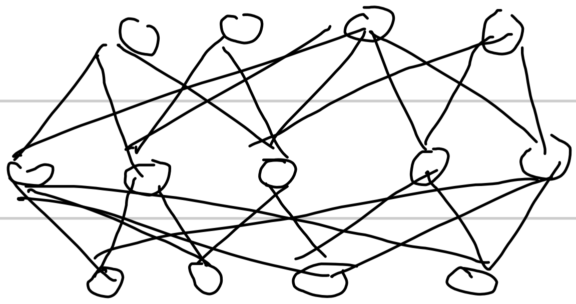
$$\text{Sigmoid}(z) = \frac{1}{1+e^{-z}}$$

ACTIVATION OF NEURONS!! $\Longrightarrow$



DEEP BOLTZMANN MACHINES

Stacking restricted boltzmann machines



DEEP BELIEF NETWORKS

Learning weights one layer at a time

HISTORY OF NEURAL NETWORKS

① Perceptrons ~ 1960s

Hand-coded weights, directed graphical models

② 2nd Gen neural networks ~ 1985

Learned weights using back-propagation

③ Deep Learning ~ 1996

Undirected graphical models

Restricted Boltzmann Machines

Sparse coding