机器学习白板推导系列

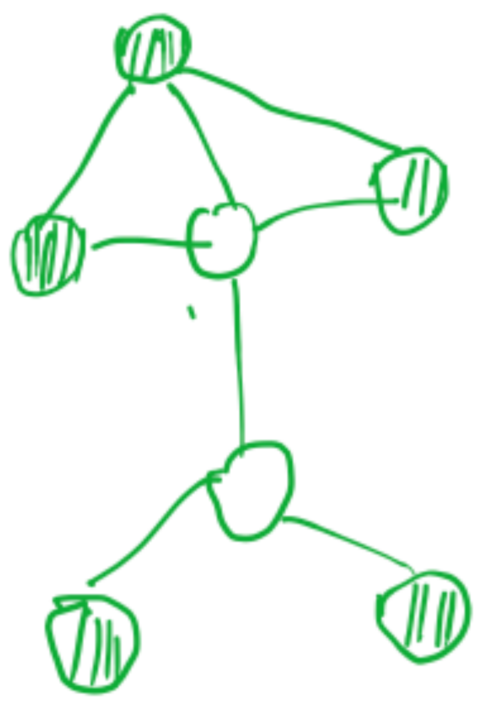
受限玻尔兹曼机-RBM

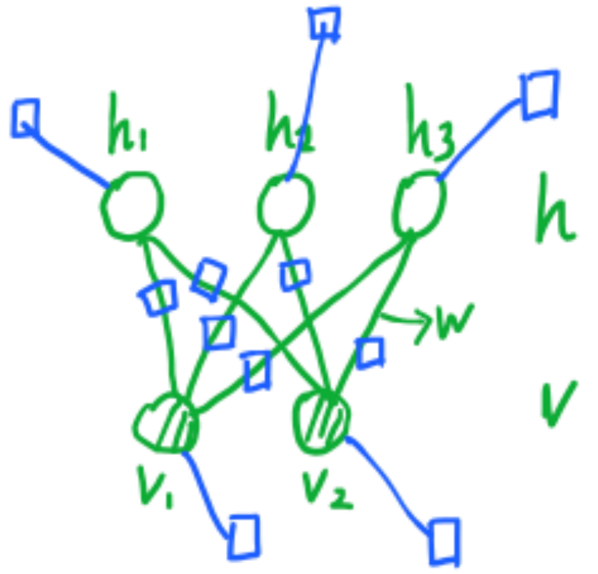
RBM - Representation

RBM → Restricted Boltzmann Machine

Boltzmann Dist:

统计物理：一个物理系统





Boltzman Machine : Markov Random Field with hidden nodes.

Nodes → R.V. → $\begin{cases} \text{observed variable}: v \\ \text{hidden variable}: h \end{cases}$

因子分解: (Hammersley Clifford Theorem)

→ 基于最大团. (C_i: 最大团, $\psi_i(x_{c_i})$: 势函数 potential function

Z: (规一化因子) $\underbrace{\text{partition function}}_{\text{配分函数}}$

$$P(x) = \frac{1}{Z}\prod_{i=1}^{K}\psi_i(x_{c_i})$$

s.t. : ψ_i 严格大于 0

$$Z = \sum_x \prod_{i=1}^{K}\psi_i(x_{c_i}) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_p}\prod_{i=1}^{K}\psi_i(x_{c_i})$$

→ $\psi_i(x_{c_i}) = \exp\{-E(x_{c_i})\}$ → E: energy function 能量函数

→ $P(x) = \frac{1}{Z}\prod_{i=1}^{K}\psi_i(x_{c_i}) = \underbrace{\frac{1}{Z}\exp\{-\sum_{i=1}^{K}E(x_{c_i})\}}_{\text{指数族分布}}$

→ $\underbrace{P(x) = \frac{1}{Z}\exp\{-E(x)\}}_{\text{Boltzmann Distribution (Gibbs Distribution)}}$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} h \\ v \end{pmatrix} \quad h = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix} \quad v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \quad p = m+n$$

Boltzmann Machine : 问题 : Inference $\begin{cases} \text{精确} \to \text{untractable} \\ \text{近似} \to \text{计算量太大} \end{cases}$

简化

→ Restricted Boltzmann Machine : (h, v 之间有连结, h, v 内部无连结)

$$P(x) = \frac{1}{Z}\exp\{-E(x)\}$$

$$\|$$

$$P(v,h) = \frac{1}{Z}\exp\{-E(v,h)\} = \frac{1}{Z}\exp\{h^T w v + \alpha^T v + \beta^T h\} = \underbrace{\frac{1}{Z}\exp\{\underset{\text{边}}{h^T w v}\}\cdot\exp(\underset{\text{点}}{\alpha^T v})\cdot\exp(\underset{\text{点}}{\beta^T h})}_{\text{factor graph view}}$$

$$E(v,h) = -(h^T w v + \alpha^T v + \beta^T h)$$

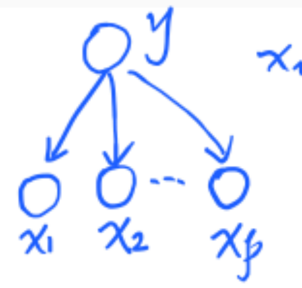
$$\text{RBM' pdf} \longrightarrow P(x) = P(v, h) = \frac{1}{Z} \exp(h^T W v) \cdot \exp(\alpha^T v) \cdot \exp(\beta^T h)$$

$$= \frac{1}{Z} \underbrace{\prod_{i=1}^{m} \prod_{j=1}^{n} \exp(h_i w_{ij} v_j)}_{\text{edge}} \underbrace{\prod_{j=1}^{n} \exp(\alpha_j v_j)}_{\text{node } v} \underbrace{\prod_{i=1}^{m} \exp(\beta_i h_i)}_{\text{node } h}$$

factor

M-Representation

Naive Bayes ← NB → 朴素贝叶斯假设 →



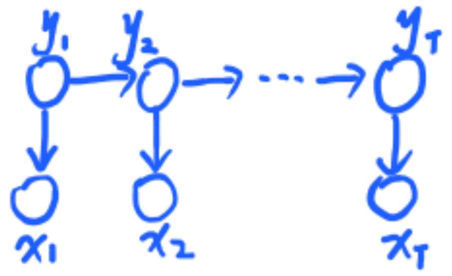
$x_i \perp x_j \mid y$, $i \neq j$

Gaussian Mixture Model ← GMM → 引入隐变量 →

$y \sim$ 离散的, K个选择
$x|y \sim$ Gaussian Dist

y 变 → Seq

State Space Model ← SSM { HMM / Kalman Filter / Particle Filter } →

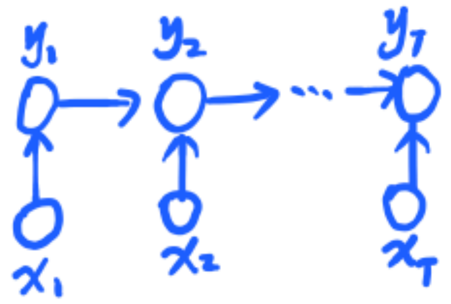


LR → MEM

Maximum Entropy Markov Model ← MEMM

{ 引入隐变量
两个假设 { 齐次Markov / 观测独立 }

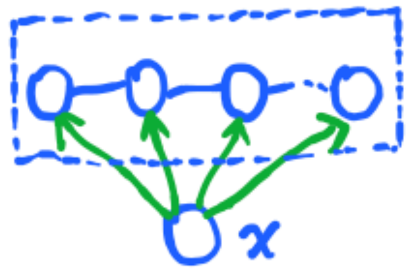
{ 判别模型
打破了观测独立 } →

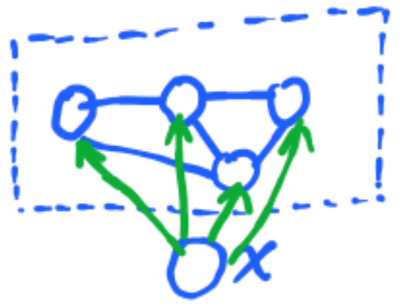


Conditional Random Field ← CRF

{ 判别模型
无向(MRF)
打破了齐次Markov }

Linear Chain-CRF ← LC-CRF

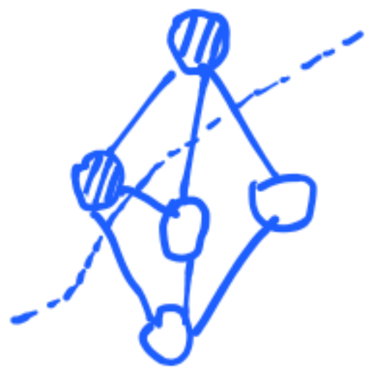




$y \mid x \sim$ MRF

Boltzmann Machine ← BM

{ 无向
引入了隐变量
PDF: 指数族 }



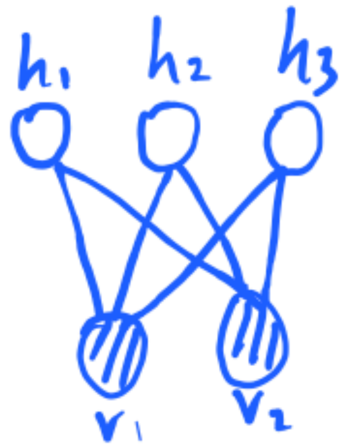
$x = \{v, h\}$

$p(x) = \frac{1}{Z} \exp\{-E(x)\}$

能量函数

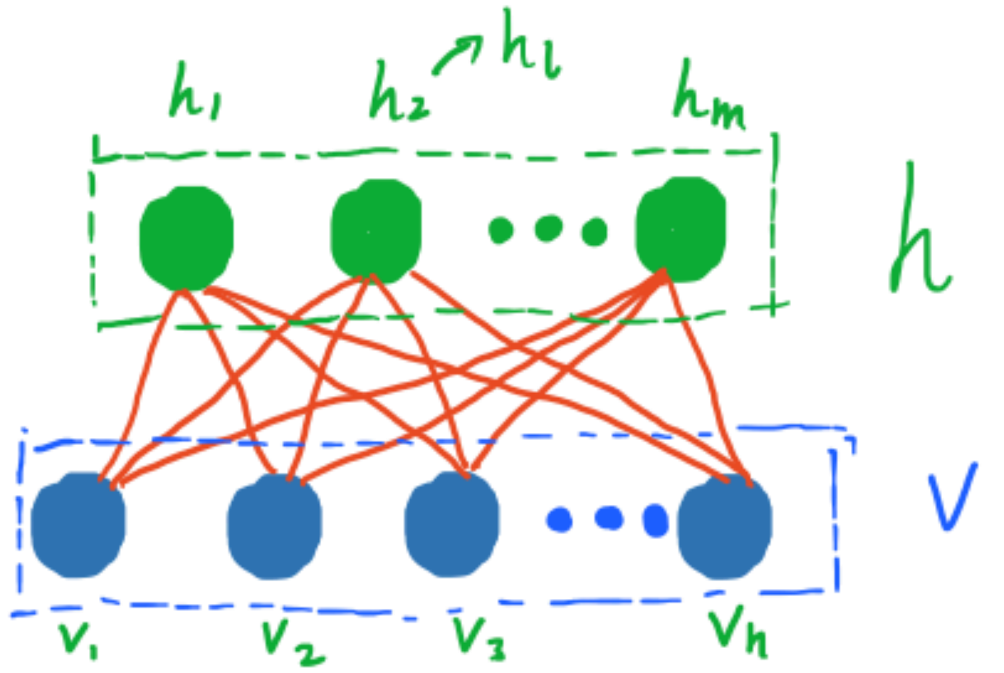
Restricted BM ← RBM

{ BM
条件独立性 }



概率图模型 {
方向(有向/无向) → 边
离散/连续/混 → 点
条件独立性 → 边
隐变量 → 点
指数族分布 → 结构
}

RBM-Inference



$x=\begin{pmatrix}x_1\\x_2\\\vdots\\x_p\end{pmatrix}=\begin{pmatrix}h\\v\end{pmatrix}\quad h=\begin{pmatrix}h_1\\h_2\\\vdots\\h_m\end{pmatrix}\quad v=\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}\quad p=m+n$

$P(x)=\frac{1}{Z}\exp\{-E(x)\}$

$\Updownarrow \qquad \Updownarrow$

$P(h,v)=\frac{1}{Z}\exp\{-E(h,v)\}$

$E(h,v)=-(h^TWv+\alpha^Tv+\beta^Th)$

$=-(\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}h_iw_{ij}v_j}_{\Delta}+\underbrace{\sum_{j=1}^{n}\alpha_jv_j}_{x}+\underbrace{\sum_{i=1}^{m}\beta_ih_i}_{\Delta})$

目的: Inference $\rightarrow$ posterior $\rightarrow P(h|v),\ P(v|h)$

求$P(h|v)$

$\boxed{P(h|v)=\prod_{l=1}^{m}P(h_l|v)}$

同理可得 $\Rightarrow P(v|h)=\prod_{l=1}^{n}P(v_l|h)$

$P(h_l=1|v)=P(h_l=1|h_{-l},v)=\frac{P(h_l=1,h_{-l},v)}{P(h_{-l},v)}=\frac{P(h_l=1,h_{-l},v)}{P(h_l=1,h_{-l},v)+P(h_l=0,h_{-l},v)}$

$E(h,v)=-(\underbrace{\sum_{\substack{i=1\\i\neq l}}^{m}\sum_{j=1}^{n}h_iw_{ij}v_j}_{\Delta_1}+\underbrace{h_l\sum_{j=1}^{n}w_{lj}v_j}_{\Delta_2}+\underbrace{\sum_{j=1}^{n}\alpha_jv_j}_{\Delta_3}+\underbrace{\sum_{\substack{i=1\\i\neq l}}^{m}\beta_ih_i}_{\Delta_4}+\underbrace{\beta_lh_l}_{\Delta_5}$

令 $\quad \Delta_2+\Delta_5=h_l\left(\sum_{j=1}^{n}w_{lj}v_j+\beta_l\right)=h_l\cdot H_l(v)$

$\bar{H}_l(h_{-l},v)=\Delta_1+\Delta_3+\Delta_4$

$\therefore E(h,v)=h_l\cdot H_l(v)+\bar{H}_l(h_{-l},v)$

分子 $=P(h_l=1,h_{-l},v)=\frac{1}{Z}\exp\{H_l(v)+\bar{H}_l(h_{-l},v)\}$

分母 $=\frac{1}{Z}\exp\{H_l(v)+\bar{H}_l(h_{-l},v)\}+\frac{1}{Z}\exp\{\bar{H}_l(h_{-l},v)\}$

$P(h_l=1|v)=\frac{1}{1+\exp\{\bar{H}_l(h_{-l},v)-H_l(v)-\bar{H}_l(h_{-l},v)\}}$

$=\frac{1}{1+\exp\{-H_l(v)\}}=\sigma(H_l(v))=\sigma\left(\sum_{j=1}^{n}w_{lj}v_j+\beta_l\right)$

RBM $\Rightarrow$ 神经网络 sigmoid $\rightarrow \sigma(x)=\frac{1}{1+e^{-x}}$

目的: Inference → marginal → $P(v)$

$$h=\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix} \quad h_i \in \{0,1\}$$

$$W=[W_{ij}]_{m\times n} = \begin{pmatrix} — w_1 — \\ — w_2 — \\ \vdots \\ — w_m — \end{pmatrix}$$

$$P(v)=\sum_h P(h,v)=\sum_h \frac{1}{Z}\exp\{-E(h,v)\}=\sum_h \frac{1}{Z}\exp\{+(h^TWv+\alpha^Tv+\beta^Th)\}$$

$$=\sum_{h_1}\cdots\sum_{h_m}\exp\{+(h^TWv+\alpha^Tv+\beta^Th)\}/Z$$

$$=\exp(\alpha^Tv)\cdot\sum_{h_1}\cdots\sum_{h_m}\exp\{h^TWv+\beta^Th\}/Z$$

$$=\exp(\alpha^Tv)\cdot\sum_{h_1}\cdots\sum_{h_m}\exp\{\sum_{i=1}^m(h_iw_iv+\beta_ih_i)\}/Z$$

$$=\exp(\alpha^Tv)\cdot\sum_{h_1}\cdots\sum_{h_m}\exp\{h_1w_1v+\beta_1h_1\}\cdot\exp\{h_2w_2v+\beta_2h_2\}\cdots\exp\{h_mw_mv+\beta_mh_m\}/Z$$

$$=\exp(\alpha^Tv)\sum_{h_1}\exp\{h_1w_1v+\beta_1h_1\}\cdots\sum_{h_m}\exp\{h_mw_mv+\beta_mh_m\}$$

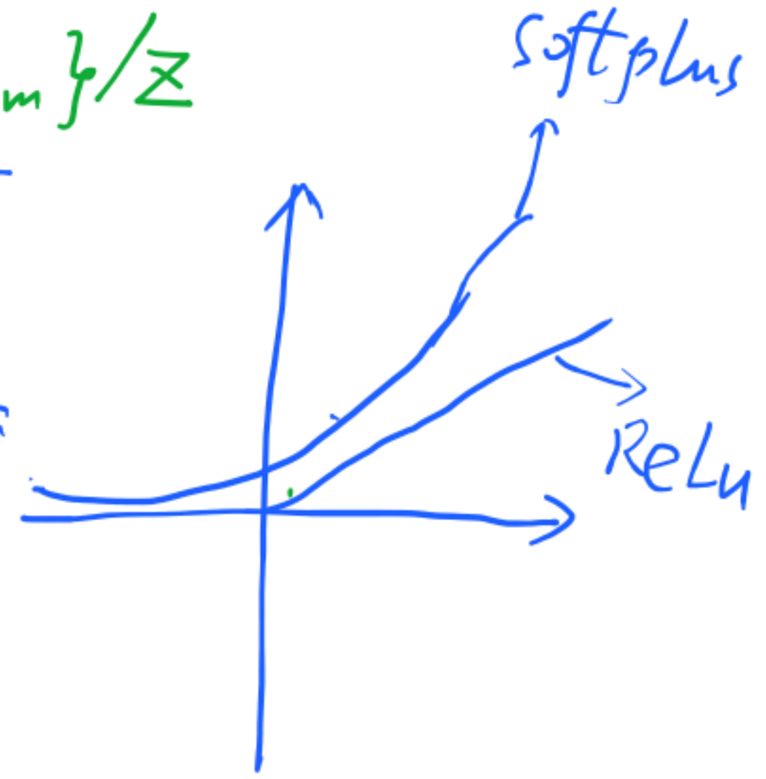
$$=\exp(\alpha^Tv)\underbrace{(1+\exp\{w_1v+\beta_1\})}_{\exp\log}\cdots(1+\exp\{w_mv+\beta_m\})$$

$$=\exp(\alpha^Tv)\cdot\exp\{\log(1+\exp\{w_1v+\beta_1\})\cdots\exp\{\log(1+\exp\{w_mv+\beta_m\})\}$$

$$=\exp\left(\alpha^Tv+\sum_{i=1}^m\underbrace{\log(1+\exp\{w_iv+\beta_i\})}_{\text{soft plus}}\right)$$

$$=\exp(\alpha^Tv+\sum_{i=1}^m \text{softplus}(w_iv+\beta_i)).$$

softplus: $\log(1+e^x)$



$$P(v)=\exp(\alpha^Tv+\sum_{i=1}^m\text{softplus}(w_iv+\beta_i))$$ 其中 w_i是 W 的行向量.

THANK YOU

For questions or suggestions:

EMail:zhoujie338@126.com

Github:
https://github.com/shuhuai007/Machine-Learning-Session