机器学习白板推导系列

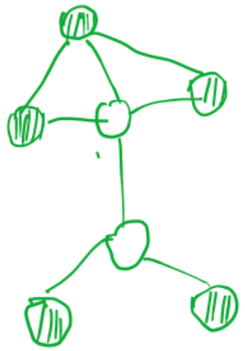
受限玻尔兹曼机-RBM

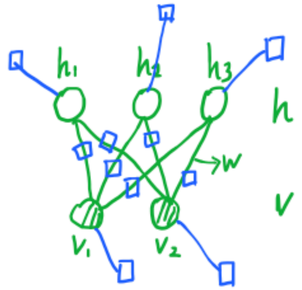
RBM - Representation

RBM
Restricted Boltzmann Machine

Boltzmann Dist:

统计物理：一个物理系统





Boltzmann Machine：Markov Random Field with hidden nodes.

Nodes → R.V. → { observed variable：v ; hidden variable：h }

因子分解：(Hammersley Clifford Theorem)

→ 基于最大团.（C_i：最大团，$\psi_i(x_{ci})$：势函数 potential function；Z：(规一化因子) partition function 配分函数）

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

s.t.：ψ_i 要格大于0

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

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

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

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

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 { 精确 → untractable ; 近似 → 计算量太大 }

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

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

$$\|$$

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

（边 / 点）

factor graph view

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

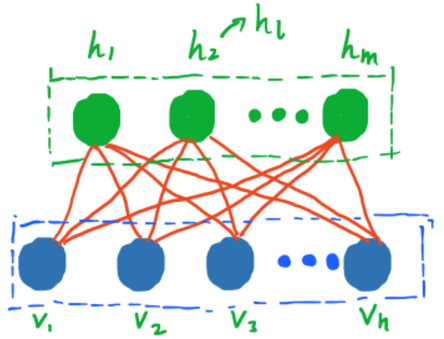
RBM' pdf $\longrightarrow$ $P(x) = P(v,h) = \frac{1}{Z}\exp(h^T W v)\cdot\exp(a^T v)\cdot\exp(b^T h)$

$$= \frac{1}{Z}\prod_{i=1}^{m}\prod_{j=1}^{n}\exp(h_i W_{ij} v_j)\prod_{j=1}^{n}\exp(a_j v_j)\prod_{i=1}^{m}\exp(b_i h_i)$$

edge — node v — node h

factor

RBM-Inference



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

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

$\Updownarrow$ $\Updownarrow$

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

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

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

Δ $\times$ Δ

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

求$P(h|v)$

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

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

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

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

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

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

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

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

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

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

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

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

目的：Inference → marginal → $P(v)$

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

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

$$W=[W_{ij}]_{m\times n}=\begin{pmatrix}-W_1-\\-W_2-\\\vdots\\-W_m-\end{pmatrix}$$

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

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

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

$$=\exp(a^Tv)\cdot\sum_{h_1}\cdots\sum_{h_m}\exp\{h_1W_1v+\beta_1h_1\}\cdot\exp\{h_2W_2v+\beta_2h_2\}\cdots\exp\{h_mW_mv+\beta_mh_m\}/Z$$

$$=\exp(a^Tv)\sum_{h_1}\exp\{h_1W_1v+\beta_1h_1\}\cdots\sum_{h_m}\exp\{h_mW_mv+\beta_mh_m\}$$

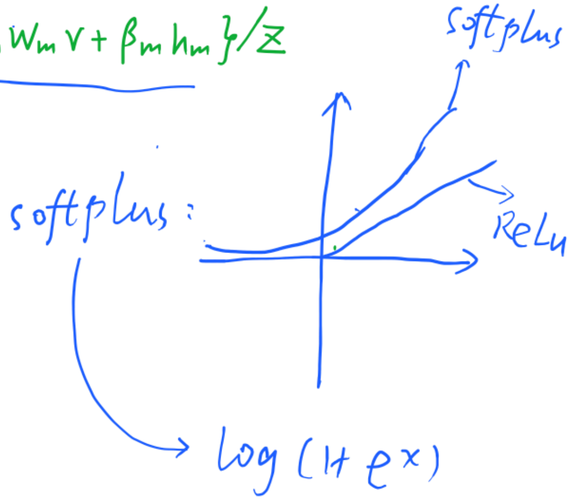
$$=\exp(a^Tv)\underbrace{(1+\exp\{W_1v+\beta_1\})}_{\exp\log}\cdots(1+\exp\{W_mv+\beta_m\})$$

$$=\exp(a^Tv)\cdot\exp\{\log(1+\exp\{W_1v+\beta_1\}\cdots\exp\{\log(1+\exp\{W_mv+\beta_m\})\}$$

$$=\exp\left(a^Tv+\sum_{i=1}^m\underbrace{\log(1+\exp\{W_iv+\beta_i\})}_{\text{softplus}}\right)$$

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

$$P(v)=\exp(a^Tv+\sum_{i=1}^m\text{softplus}(W_iv+\beta_i))$$ 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