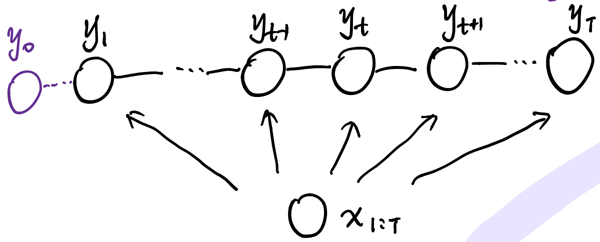
CRF 概率密度的参数形式

HMM — MEMM — CRF （合理性）

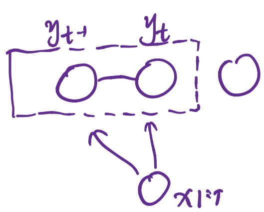
条件 → 判别式模型 $P(Y|X)$

随机场 → 无向图模型



$$P(Y|X)=\frac{1}{Z}\exp\sum_{i=1}^{K}F_i(x_{c_i})$$

$$=\frac{1}{Z}\exp\sum_{t=1}^{T}\underbrace{F(y_{t-1},y_t,x_{1:T})}_{\Delta}$$



$$=\frac{1}{Z}\exp\sum_{t=1}^{T}\left[\sum_{k=1}^{K}\lambda_k f_k(y_{t-1},y_t,x_{1:T})+\sum_{l=1}^{L}\eta_l g_l(y_t,x_{1:T})\right]$$

MRF 因子分解: $x\in\mathbb{R}^p$

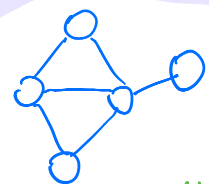
$$P(x)=\frac{1}{Z}\prod_{i=1}^{K}\underbrace{\psi_i(x_{c_i})}_{>0}$$

势函数

$$=\frac{1}{Z}\prod_{i=1}^{K}\exp[-E_i(x_{c_i})]$$

能量函数

$$=\frac{1}{Z}\exp\sum_{i=1}^{K}F_i(x_{c_i})$$



K个最大团

C_i: 最大团

t时刻

$$F(y_{t-1},y_t,x_{1:T})=\Delta_{y_{t-1},x_{1:T}}+\Delta_{y_t,x_{1:T}}+\Delta_{y_{t-1},y_t,x_{1:T}}$$

状态函数　转移函数

$$=\Delta_{y_t,x_{1:T}}+\Delta_{y_{t-1},y_t,x_{1:T}}$$

$$\Delta_{y_{t-1},y_t,x_{1:T}}=\sum_{k=1}^{K}\lambda_k f_k(y_{t-1},y_t,x_{1:T})$$

$$\Delta_{y_t,x_{1:T}}=\sum_{l=1}^{L}\eta_l g_l(y_t,x_{1:T})$$

f_k, g_l 给定的特征函数

λ_k, η_l 参数

CRF 概率密度的向量形式

$$P(Y|X) = \frac{1}{Z} \exp \sum_{t=1}^{T} \Big[\underbrace{\sum_{k=1}^{K} \lambda_k f_k(y_{t-1}, y_t, x_{1:T})} + \sum_{l=1}^{L} \eta_l \, g_l(y_t, x_{1:T}) \Big]$$

$$P(Y=y \mid X=x) = \frac{1}{Z(x,\lambda,\eta)} \exp \sum_{t=1}^{T} \Big[\lambda^T \cdot f(y_{t-1}, y_t, x) + \eta^T \cdot g(y_t, x) \Big]$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \quad \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_K \end{pmatrix} \quad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_L \end{pmatrix} \quad f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_K \end{pmatrix} = f(y_{t-1}, y_t, x) \quad g = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_L \end{pmatrix} = g(y_t, x)$$

$$= \frac{1}{Z(x,\lambda,\eta)} \exp \Big(\lambda^T \cdot \sum_{t=1}^{T} f(y_{t-1}, y_t, x) + \eta^T \sum_{t=1}^{T} g(y_t, x) \Big)$$

$$\theta = \begin{pmatrix} \lambda \\ \eta \end{pmatrix}_{K+L} \qquad H = \begin{pmatrix} \sum_{t=1}^{T} f \\ \sum_{t=1}^{T} g \end{pmatrix}_{K+L}$$

$$P(Y=y \mid X=x) = \frac{1}{Z(x,\theta)} \exp \underbrace{\theta^T \cdot H(y_t, y_{t-1}, x)}_{\langle \theta, H \rangle}$$

$P(Y=y \mid X=x) \rightarrow P(y|x)$

CRF 几个问题

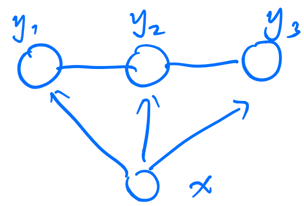
Given training data : $\{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$ 　　x, y 均是 T维

$\hat{\theta} = \arg\max \prod_{i=1}^{N} P(y^{(i)}|x^{(i)})$

Learning : parameter estimation ②

Inference :
- marginal prob : $\Rightarrow P(y_t|x)$ ① $\rightarrow P(y_t = i|x)$
- condinal prob : 生成模型
- MAP Inference : decoding $\Rightarrow \hat{y} = \arg\max P(y|x)$ ③

  $y = y_1 y_2 \cdots y_T$



HMM : decoding $\rightarrow$ Viterbi $\rightarrow$ 动态规划

CRF

CRF 概率计算

Given $P(Y=y|X=x)$，求 $P(y_t=i|x)$

$$P(y|x)=\frac{1}{Z}\prod_{t=1}^{T}\psi_t(y_{t-1},y_t,x)$$

$$P(y_t=i|x)=\sum_{y_1,y_2\cdots y_{t-1}y_{t+1}\cdots y_T}\qquad P(y|x)=\sum_{y_{<1,t-1>}}\sum_{y_{<t+1,T>}}\frac{1}{Z}\prod_{t'=1}^{T}\psi_{t'}(y_{t'-1},y_{t'},x)$$

$y_t\in S$

$$=\frac{1}{Z}\ \Delta_{左}\cdot\Delta_{右}$$

$$\Delta_{左}=\sum_{y_{<1:t-1>}}\psi_1(y_0,y_1,x)\,\psi_2(y_1,y_2,x)\cdots\psi_{t-1}(y_{t-2},y_{t-1},x)\cdot\psi_t(y_{t-1},y_t=i,x)$$

$$\Delta_{右}=\sum_{y_{<t+1,T>}}\psi_{t+1}(y_t=i,y_{t+1},x)\cdot\cdots\psi_T(y_{T-1},y_T,x)$$

$$\Delta_{左}=\sum_{y_{t-1}}\psi_t(y_{t-1},y_t=i,x)\sum_{y_{t-2}}\psi_{t-1}(y_{t-2},y_{t-1},x)\cdots\psi_2(y_1,y_2,x)\cdot\sum_{y_0}\psi_1(y_0,y_1,x)$$

$$\Delta_{左}\rightarrow\alpha_t(i):\ \underbrace{y_0\ y_1\ y_2\cdots y_{t-1}}_{\text{所有势函数}}\ \underbrace{y_t=i}_{\text{左半部势函数}}$$

$$\alpha_{t-1}(j):\ \underbrace{y_0\ y_1\ y_2\cdots y_{t-2}}_{\text{所有势函数}}\ y_{t-1}=j$$

左半部势函数

$$\alpha_t(i)=\sum_{j\in S}\psi_t(y_{t-1}=j,y_t=i,x)\,\alpha_{t-1}(j)$$

$$\Delta_{左}=\alpha_t(i)$$

$$\Delta_{右}=\beta_t(i)$$

$$\therefore P(y_t=i|x)=\frac{1}{Z}\alpha_t(i)\,\beta_t(i)$$

CRF Learning

$$\hat{\theta} = \arg\max \prod_{i=1}^{N} P(y^{(i)} | x^{(i)})$$

N: size of traing data

$$\Downarrow$$

$$\hat{\lambda}, \hat{\eta} = \arg\max_{\lambda,\eta} \prod_{i=1}^{N} P(y^{(i)} | x^{(i)})$$

$$P(y|x) = \frac{1}{Z(x,\lambda,\eta)} \exp \sum_{t=1}^{T} \left[\lambda^T \cdot f(y_{t-1}, y_t, x) + \eta^T \cdot g(y_t, x) \right]$$

$$= \arg\max_{\lambda,\eta} \log \prod_{i=1}^{N} P(y^{(i)} | x^{(i)}) = \arg\max_{\lambda,\eta} \sum_{i=1}^{N} \log P(y^{(i)} | x^{(i)})$$

$$= \arg\max_{\lambda,\eta} \sum_{i=1}^{N} \left(-\log Z(x^{(i)}, \lambda, \eta) + \sum_{t=1}^{T} \left[\lambda^T \cdot f(y_{t-1}^{(i)}, y_t^{(i)}, x^{(i)}) + \eta^T \cdot g(y_t^{(i)}, x^{(i)}) \right] \right)$$

f

$$\triangleq \arg\max_{\lambda,\eta} L(\lambda, \eta, x^{(i)})$$

方法: 梯度上升 $\nabla_\lambda L$, $\nabla_\eta L$

$$\nabla_\lambda L = \sum_{i=1}^{N} \left(\sum_{t=1}^{T} f(y_{t-1}, y_t, x^{(i)}) - \nabla_\lambda \underbrace{\log Z(x^{(i)}, \lambda, \eta)}_{\text{log-partition function}} \right)$$

$$\nabla_\lambda \log Z(x^{(i)}, \lambda, \eta) = E\left[\sum_{t=1}^{T} f(y_{t-1}, y_t, x^{(i)}) \right]$$

$$= \sum_{y} P(y | x^{(i)}) \sum_{t=1}^{T} f(y_{t-1}, y_t, x^{(i)})$$

$$= \sum_{t=1}^{T} \left(\sum_{y} P(y | x^{(i)}) \cdot f(y_{t-1}, y_t, x^{(i)}) \right)$$

$$= \sum_{t=1}^{T} \sum_{y_{<1:t-2>}} \sum_{y_{t-1}} \sum_{y_t} \sum_{y_{<t+1,T>}} P(y | x^{(i)}) \cdot f(0)$$

$$= \sum_{t=1}^{T} \sum_{y_{t-1}} \sum_{y_t} \left(\sum_{y_{<1,t-2>}} \sum_{y_{<t+1,T>}} P(y | x^{(i)}) f(0) \right)$$

$$= \sum_{t=1}^{T} \sum_{y_{t-1}} \sum_{y_t} \underbrace{P(y_{t-1}, y_t, x^{(i)})}_{A(y_{t-1}, y_t)} f(y_{t-1}, y_t, x^{(i)})$$

梯度上升算法

$$\begin{cases} \lambda^{(t+1)} = \lambda^{(t)} + step \cdot \nabla_\lambda L(\lambda^{(t)}, \eta^{(t)}) \\ \eta^{(t+1)} = \eta^{(t)} + step \cdot \nabla_\eta L(\lambda^{(t)}, \eta^{(t)}) \end{cases}$$

$$\nabla_\lambda L = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(f(y_{t-1}, y_t, x^{(i)}) - \sum_{y_{t-1}} \sum_{y_t} A(y_{t-1}, y_t) \cdot f(y_{t-1}, y_t, x^{(i)}) \right)$$