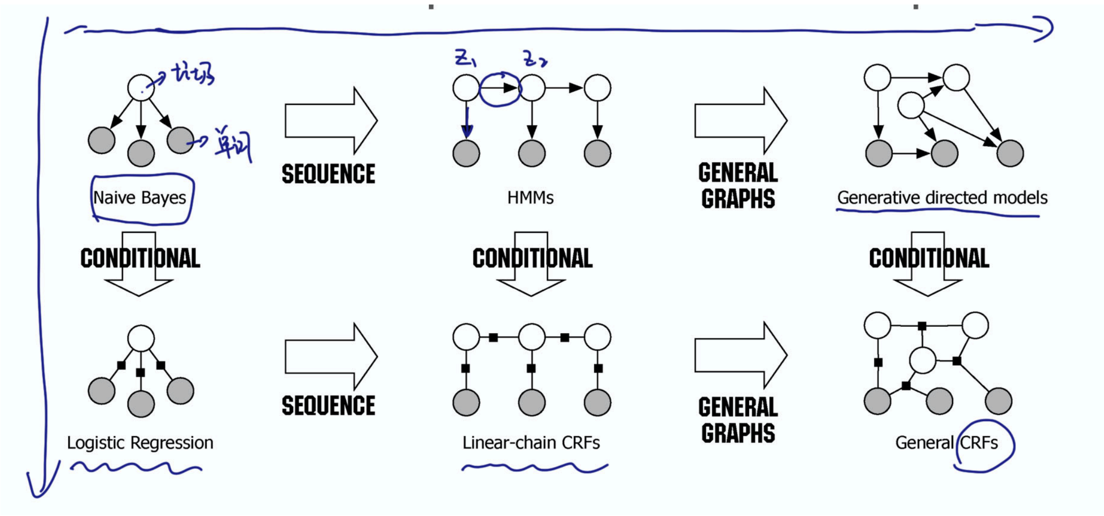
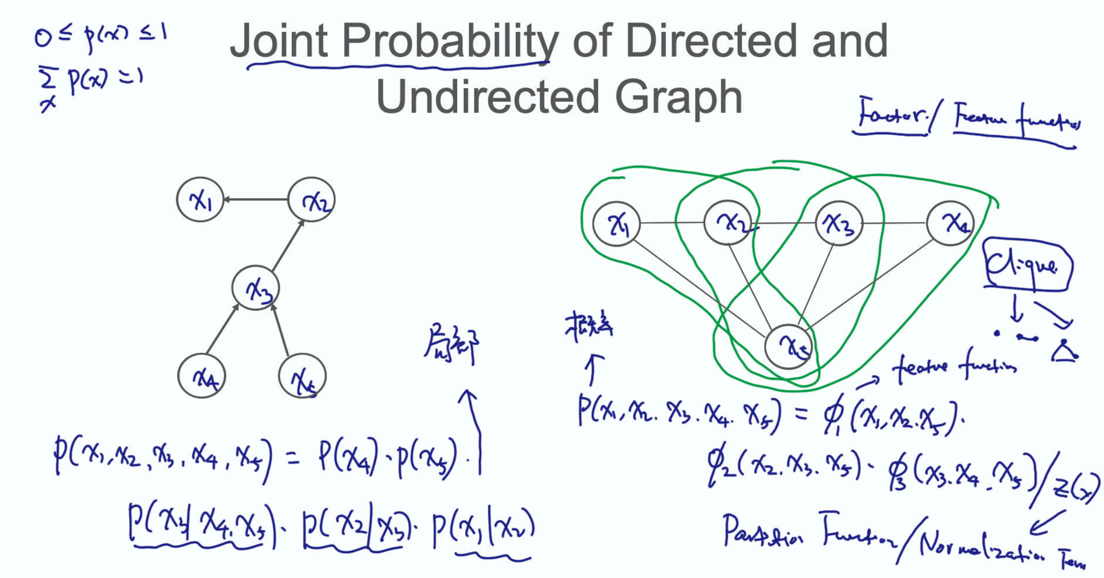
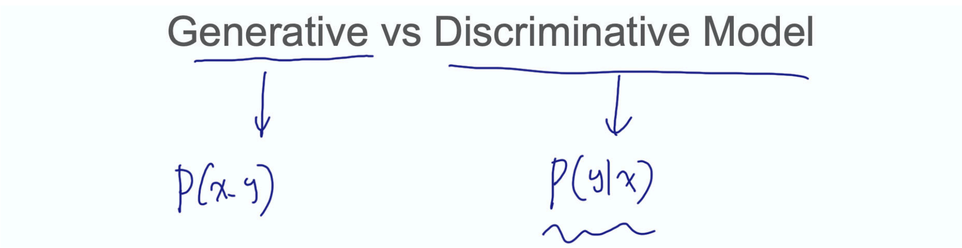
1. 有向图与无向图



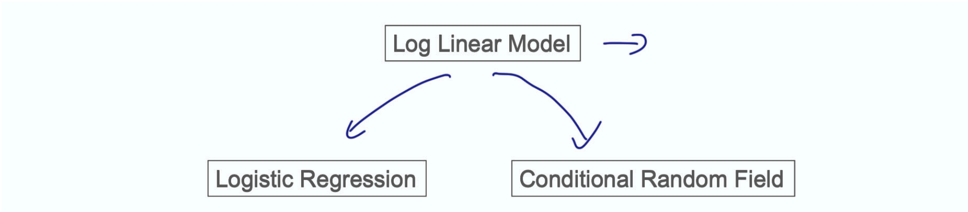


拆分有向图：clique

生成模型和判别模型



2. Log-Linear Model



2.1 标准数学公式

Log Linear Model

$$p(y|x;w) = \frac{\exp \sum_{j=1}^{J} \boxed{w_j} \boxed{F_j(x,y)}}{\boxed{Z(x,w)}}$$

w_j → 参数

$F_j(x,y)$ → Feature Function

$Z(x,w)$ → Normalization/Partition Function

→ Logistic Regression

2.2 log-linear model 与多元逻辑回归

Multinomial Logistic Regression

$\boxed{C=3}$

$$p(y|x;w) = \frac{\exp \sum_{j=1}^{J} w_j \boxed{F_j(x,y)}}{Z(x,w)}$$

$$F_j(x,y) = \underbrace{x_i}_{d} \cdot I(y=c) \qquad x \in \mathbb{R}^d,\ y \in \{1,2,\dots c\}$$

$I(y=c)$ → Indicator Function

$\boxed{j=1,2,\dots 3d}$ ← $w \in \mathbb{R}^{3d}$

$y=1$ 时	$y=2$ 时	$y=3$ 时
$F_1(x,y)=x_1$	$F_1(x,y)=0$	$F_1(x,y)=0$
$F_2(x,y)=x_2$	$F_2(x,y)=0$	$F_2(x,y)=0$
$\vdots$	$\vdots$	$\vdots$
$F_d(x,y)=x_d$	$F_d(x,y)=0$	
$F_{d+1}(x,y)=0$	$F_{d+1}(x,y)=x_1$	$F_{d+1}(x,y)=0$
$\vdots$	$\vdots$	$\vdots$
$F_{2d}(x,y)=0$	$F_{2d}(x,y)=x_d$	$F_{2d}(x,y)=0$
	$F_{2d+1}(x,y)=0$	$F_{2d+1}(x,y)=x_1$
$\vdots$	$\vdots$	$\vdots$
$F_{3d}(x,y)=0$	$F_{3d}(x,y)=0$	$F_{3d}(x,y)=x_d$

$$p(y=1|x;w) = \exp \sum_{j=1}^{3d} w_j F_j(x,y) / Z(x,w)$$

$$= \exp\left[\sum_{j=1}^{d} w_j x_j + 0\right] / Z(x,w)$$

$$= \exp\left[\sum_{j=1}^{d} w_j x_j\right] / Z(x,w)$$

$$p(y=2|x;w) = \exp\left[\sum_{j=d+1}^{2d} w_j \cdot x_{j-d}\right] / Z(x,w)$$

$$p(y=3|x;w) = \exp\left[\sum_{j=2d+1}^{3d} w_j \cdot x_{j-2d}\right] / Z(x,w)$$

$w \in \mathbb{R}^{3d}$ $\quad\Sigma$

$$w = (\underbrace{w_1, w_2 \dots w_d}_{w^{(1)}}\ \underbrace{w_{d+1} \dots w_{2d}}_{w^{(2)}}\ \underbrace{w_{2d+1} \dots w_{3d}}_{w^{(3)}})^T$$

$$p(y=1|x;w) = \exp(w^{(1)T} \cdot x) / Z(x,w)$$

$$p(y=2|x;w) = \exp(w^{(2)T} \cdot x) / Z(x,w)$$

$$p(y=3|x;w) = \exp(w^{(3)T} \cdot x) / Z(x,w)$$

$$\boxed{Z(x,w)} = \exp(w^{(1)T} \cdot x) + \exp(w^{(2)T} \cdot x) + \exp(w^{(3)T} \cdot x)$$

$$\Rightarrow p(y=1|x;w) = \frac{e^{w^{(1)T} \cdot x}}{\sum_{c=1}^{3} e^{w^{(c)T} \cdot x}} \qquad p(y=2|x;w) = \frac{e^{w^{(2)T} \cdot x}}{\sum_{c=1}^{3} e^{w^{(c)T} \cdot x}} \qquad p(y=3|x;w) = \frac{e^{w^{(3)T} \cdot x}}{\sum_{c=1}^{3} e^{w^{(c)T} \cdot x}}$$

↳ 多元 Logistic Regression

$$\boxed{F_j(x,y)} = x_i \cdot I(y=c)$$
$$= \boxed{A_a(x)} \cdot B_b(y)$$

$\{w_j\}$

$A_1(x)$：是否单词x是大写还是小写？
$A_2(x)$：单词x的长度
$A_3(x)$：单词x的前缀是否"xx"

3. CRF

3.1 log-linear model for sequential data

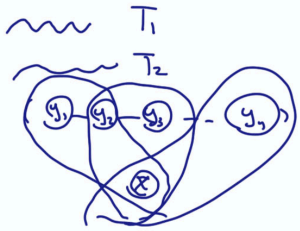
$$p(y|x;w) = \frac{1}{Z(x,w)} \cdot \exp \sum_{j=1}^{J} w_j F_j(x,y)$$

$\bar{x}$: sequence of observation

$\bar{y}$: sequence of tages

$$p(\bar{y}|\bar{x};w) = \frac{1}{Z(\bar{x},w)} \exp \sum_{j=1}^{J} w_j \boxed{F_j(\bar{x},\bar{y})}$$
$$= \frac{1}{Z(\bar{x},w)} \exp \sum_{j=1}^{J} w_j \sum_{i=2}^{n} \boxed{f_j(y_{i-1}, y_i, \bar{x}, i)}$$

linear-chain CRF



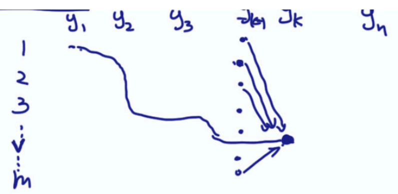
3.2 inference problem

Given $\{w\}$, $\bar{x}$, 求出 $\bar{y}$

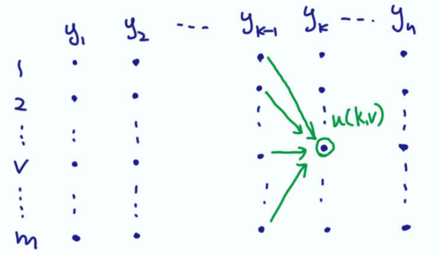
$$\hat{y} = \arg\max_{\bar{y}} p(\bar{y}|\bar{x};w)$$
$$= \arg\max_{\bar{y}} \sum_{j=1}^{J} w_j F_j(\bar{x},\bar{y})$$
$$= \arg\max_{\bar{y}} \sum_{j=1}^{J} w_j \sum_{i=2}^{n} f_j(y_{i-1}, y_i, \bar{x}, i)$$
$$= \arg\max_{\bar{y}} \sum_{i=2}^{n} g_i(y_{i-1}, y_i)$$

$$g_i(y_{i-1}, y_i) = \sum_{j=1}^{J} w_j f_j(y_{i-1}, y_i, \bar{x}, i)$$

$$\hat{y} = \arg\max_{\bar{y}} \sum_{i=2}^{n} g_i(y_{i-1}, y_i)$$



$u(k,v)$: $\boxed{\text{score}}$ of the best sequence of tags from position 1 to k where kth tag is v



$u(k,v)$: score of the best sequence of tags from position 1 to k where k-th tag is v

$$u = \begin{array}{c} 1 \cdots k \cdots n \\ \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} \end{array} \begin{array}{c} \vdots \\ m \end{array}$$

$$\boxed{u(k,v)} = \max_{y_1 \cdots y_{k-1}} \sum_{i=1}^{k-1} g_i(y_{i-1}, y_i) + g_k(y_{k-1}, v)$$

$$= \max_{y_{k-1}} \left[\max_{y_1 \cdots y_{k-2}} \sum_{i=1}^{k-2} g_i(y_{i-1}, y_i) + g_{k-1}(y_{k-2}, y_{k-1}) \right] + g_k(y_{k-1}, v)$$

$$= \max_{u} \left[\boxed{u(k-1, u)} + g_k(u, v) \right]$$

3.3 参数估计

Estimation of W

Log-Linear ①
Logistic crf ②

$$p(y|x;w) = \frac{\exp \sum_{j=1}^{J} w_j F_j(x,y)}{Z(x,w)}$$

$$\frac{\partial}{\partial w_j} \log p(y|x;w) = \frac{\partial}{\partial w_j} \left[\sum_{j=1}^{J} w_j F_j(x,y) - \log Z(x,w) \right]$$

$$= F_j(x,y) - \frac{1}{Z(x,w)} \frac{\partial}{\partial w_j} Z(x,w)$$

$$Z(x,w) = \sum_{y'} \exp \sum_{j=1}^{J} w_j F_j(x,y')$$

$$\boxed{\frac{\partial}{\partial w_j} Z(x,w)} = \frac{\partial}{\partial w_j} \sum_{y'} \exp \sum_{j=1}^{J} w_j F_j(x,y')$$

$$= \sum_{y'} \frac{\partial}{\partial w_j} \left[\exp \sum_{j'} w_{j'} F_{j'}(x,y') \right]$$

$$= \sum_{y'} \exp \sum_{j'} w_{j'} F_{j'}(x,y') \cdot \frac{\partial}{\partial w_j} \sum_{j'} w_{j'} F_{j'}(x,y')$$

$$= \boxed{\sum_{y'} \left[\exp \sum_{j'} w_{j'} F_{j'}(x,y') \right] \cdot F_j(x,y)}$$

$$\frac{\partial}{\partial w_j} \log p(y|x;w) = F_j(x,y) - \frac{1}{Z(x,w)} \cdot \sum_{y'} F_j(x,y') \left[\exp \sum_{j'} w_{j'} F_{j'}(x,y') \right]$$

$$= F_j(x,y) - \sum_{y'} F_j(x,y') \cdot \left(\frac{\exp \sum_{j'} w_{j'} F_{j'}(x,y')}{Z(x,w)} \right)$$

$$= F_j(x,y) - \sum_{y'} F_j(x,y') \cdot p(y'|x;w)$$

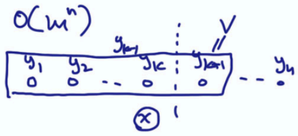
$$= F_j(x,y) - \mathop{E}_{y' \sim p(y'|x;w)} \left\{ F_j(x,y') \right\} \quad ①$$

$$z(\bar{x}, w) = \sum_{\bar{y}} \exp \sum_{j=1}^{J} w_j F_j(\bar{x}, \bar{y})$$

$$= \sum_{\bar{y}} \exp \sum_{j=1}^{J} w_j \sum_{i=2}^{n} f_j(y_{i-1}, y_i, \bar{x}, i) = \sum_{\bar{y}} \exp \sum_{i=2}^{n} g_i(y_{i-1}, y_i)$$

① 考虑所有可能的组合 $O(m^n)$

<ins>Forward Alg:</ins>



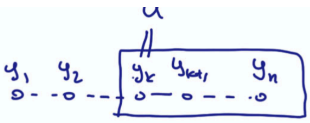
$$\alpha(k+1, v)$$

$$= \sum_{y_1 \cdots y_k} \exp\left[\sum_{i=2}^{k} g_i(y_{i-1}, y_i) + g_{k+1}(y_k, v)\right]$$

$$= \sum_u \left(\sum_{y_1 \cdots y_{k-1}} \exp\left[\sum_{i=2}^{k-1} g_i(y_{i-1}, y_i)\right] \exp\left[g_k(y_{k-1}, u)\right]\right) \exp\left(g_{k+1}(u, v)\right)$$

$$= \sum_u \alpha(k, u) \cdot \exp g_{k+1}(u, v) \Rightarrow \boxed{\alpha(k+1, v) = \sum_u \alpha(k, u) \exp g_{k+1}(u, v)}$$

<ins>Backward Alg:</ins>



$$\beta(u, k) = \sum_v \left[\exp g_{k+1}(u, v)\right] \cdot \beta(v, k+1)$$

$$\boxed{z(\bar{x}, w) = \sum_u \alpha(k, u) \cdot \beta(u, k)} \leftarrow \text{离散型变量.}$$

$\boxed{\text{连续型}} \rightarrow$ Monte Carlo method

$$\boxed{P(y_k = u \mid \bar{x}; w)} = \frac{\alpha(k, u) \cdot \beta(u, k)}{z(\bar{x}, w)}$$

$$P(y_k = u, y_{k+1} = v \mid \bar{x}; w) = \frac{\alpha(k, u)\left[\exp g_{k+1}(u, v)\right] \beta(v, k+1)}{z(\bar{x}, w)} \quad ②$$

$$\frac{\partial}{\partial w_j} \log P(\bar{y} \mid \bar{x}; w) = F_j(\bar{x}, \bar{y}) - E_{y' \sim p(y' \mid \bar{x}; w)}\left[F_j(\bar{x}, y')\right]$$

$$= F_j(\bar{x}, \bar{y}) - E_{\bar{y}}\left[\sum_{i=2}^{n} f_j(y_{i-1}, y_i, \bar{x}, i)\right]$$

$$= F_j(\bar{x}, \bar{y}) - \sum_{i=2}^{n} E_{\bar{y}}\left[f_j(y_{i-1}, y_i, \bar{x}, i)\right]$$

$$= F_j(\bar{x}, \bar{y}) - \sum_{i=2}^{n} E_{y_{i-1}, y_i}\left[f_j(y_{i-1}, y_i, \bar{x}, i)\right]$$

$$= F_j(\bar{x}, \bar{y}) - \sum_{i=2}^{n} \sum_{y_{i-1}} \sum_{y_i} f_j(y_{i-1}, y_i, \bar{x}, i) \cdot P(y_i, y_{i-1} \mid \bar{x}; w) \leftarrow ②$$

$$= F_j(\bar{x}, \bar{y}) - \sum_{i=2}^{n} \underbrace{\sum_{y_{i-1}} \sum_{y_i}}_{O(m^2)} f_j(y_{i-1}, y_i, \bar{x}, i) \cdot \frac{\alpha(i-1, y_{i-1}) \cdot \left[\exp g_i(y_{i-1}, y_i)\right] \cdot \beta(y_i, i)}{z(\bar{x}, w)}$$

$$w_j^{t+1} \leftarrow w_j^t - \eta_t \cdot \frac{\partial}{\partial w_j} \log P(\bar{y} \mid \bar{x}; w)$$