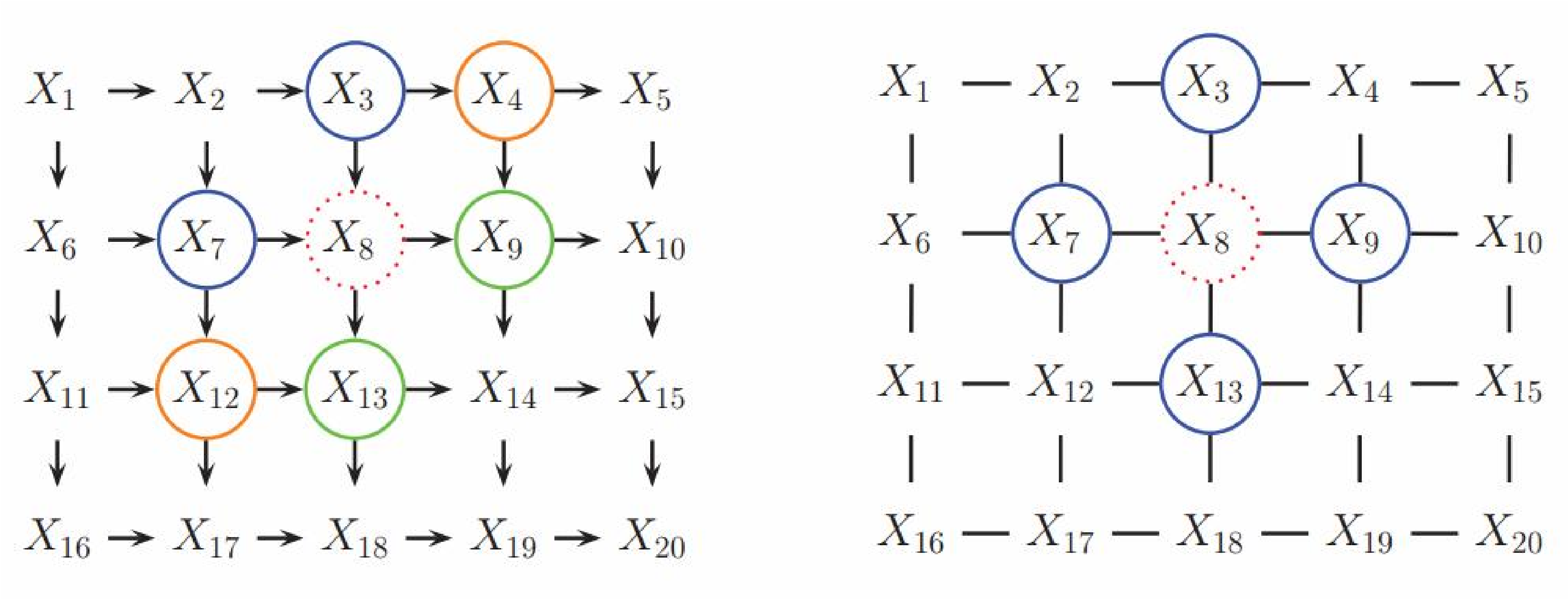
条件随机场

七月算法 邹博

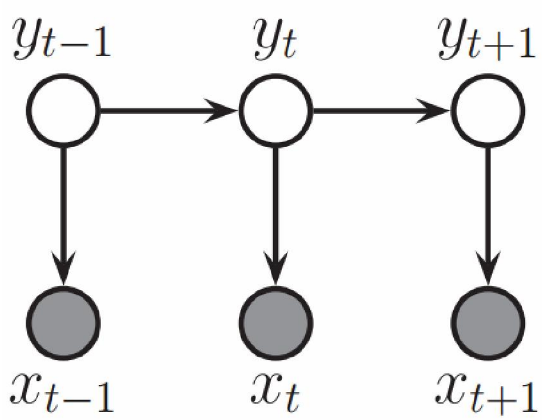
2015年12月13日

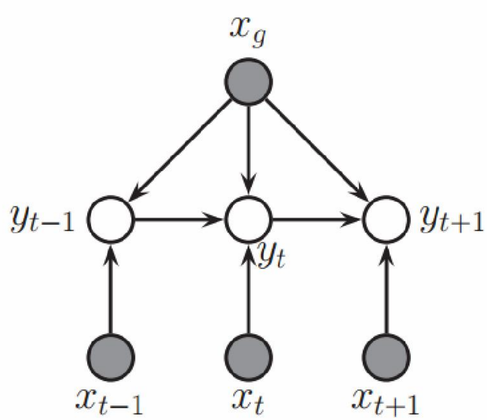
分析对图像像素间建立贝叶斯网络

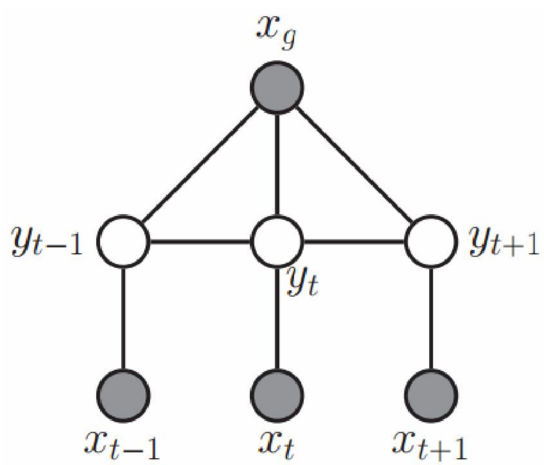
- 考察X8的马尔科夫毯(Markov blanket)

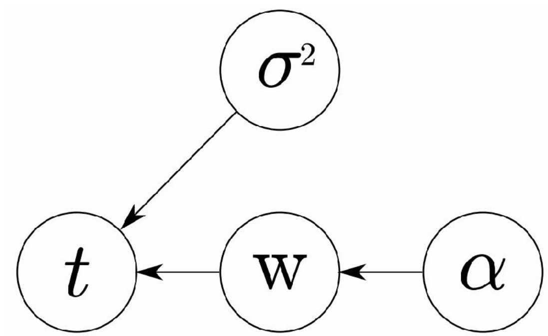


网络模型比较HMM/MEMM/CRF/RVM









条件随机场举例

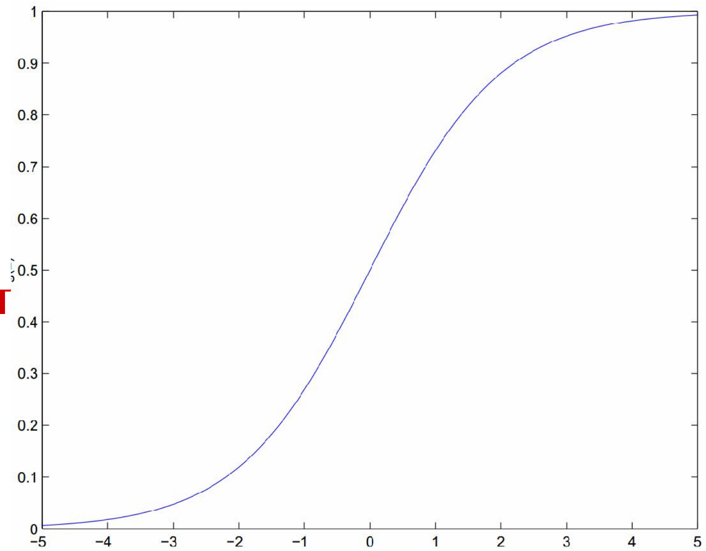
[PRP He] [VBZ reckons] [DT the] [JJ current] [NN account] [NN deficit] [MD will] [VB narrow] [TO to] [RB only] [# #] [CD 1.8] [CD billion] [IN in] [NNP September] [. .]

- NN、NNS、NNP、NNPS、PRP、DT、JJ分别代表普通名词单数形式、普通名词复数形式、专有名词单数形式、专有名词复数形式、代词、限定词、形容词

$$b(\boldsymbol{x}, i) = \begin{cases} 1 & \text{if the observation at position } i \text{ is the word ``September''} \\ 0 & \text{otherwise.} \end{cases}$$

$$t_j(y_{i-1}, y_i, \boldsymbol{x}, i) = \begin{cases} b(\boldsymbol{x}, i) & \text{if } y_{i-1} = \text{IN and } y_i = \text{NNP} \\ 0 & \text{otherwise.} \end{cases}$$

从Logistic回归谈起



- Logistic函数

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

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$$\begin{aligned} g'(z) &= \frac{d}{dz}\frac{1}{1+e^{-z}} \\ &= \frac{1}{(1+e^{-z})^2}\left(e^{-z}\right) \\ &= \frac{1}{(1+e^{-z})}\cdot\left(1-\frac{1}{(1+e^{-z})}\right) \\ &= g(z)(1-g(z)) \end{aligned}$$

Logistic回归参数估计

- 假定：
$$\begin{aligned} P(y=1 \mid x;\theta) &= h_\theta(x) \\ P(y=0 \mid x;\theta) &= 1-h_\theta(x) \end{aligned}$$

$$p(y \mid x;\theta) = \left(h_\theta(x)\right)^y \left(1-h_\theta(x)\right)^{1-y}$$

$$\begin{aligned} L(\theta) &= p(\vec{y} \mid X;\theta) \\ &= \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)};\theta) \\ &= \prod_{i=1}^{m} \left(h_\theta(x^{(i)})\right)^{y^{(i)}} \left(1-h_\theta(x^{(i)})\right)^{1-y^{(i)}} \end{aligned}$$

对数似然函数

$$
\begin{aligned}
\ell(\theta) &= \log L(\theta) \\
&= \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1-y^{(i)}) \log(1-h(x^{(i)}))
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j}\ell(\theta) &= \left(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)}\right)\frac{\partial}{\partial \theta_j} g(\theta^T x) \\
&= \left(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)}\right) g(\theta^T x)(1-g(\theta^T x))\frac{\partial}{\partial \theta_j}\theta^T x \\
&= \left(y(1-g(\theta^T x)) - (1-y)g(\theta^T x)\right) x_j \\
&= \left(y - h_\theta(x)\right) x_j
\end{aligned}
$$

参数的迭代

- Logistic回归参数的学习规则：

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

- 比较上面的结果和线性回归的结论的差别：
  - 它们具有相同的形式！

Repeat until convergence {

$$\theta_j := \theta_j + \alpha \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

Loop {

for i=1 to m, {

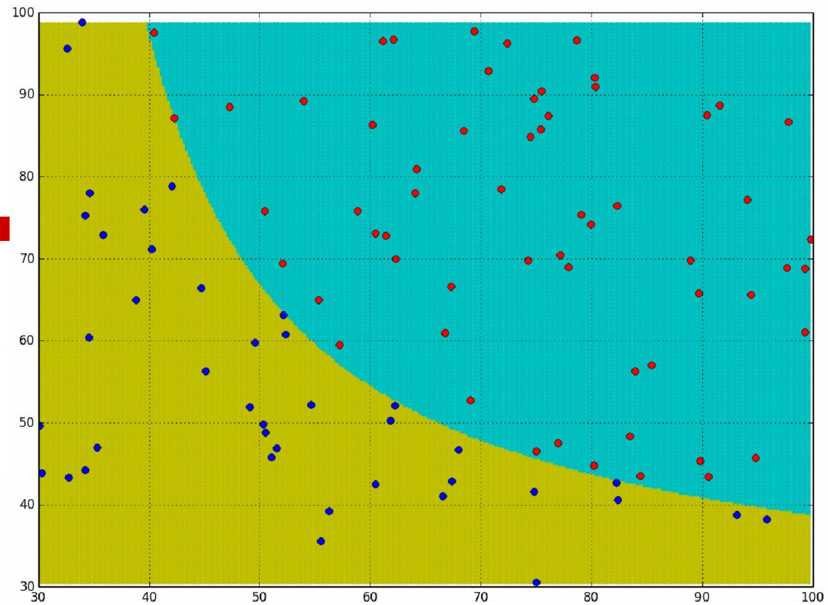
$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

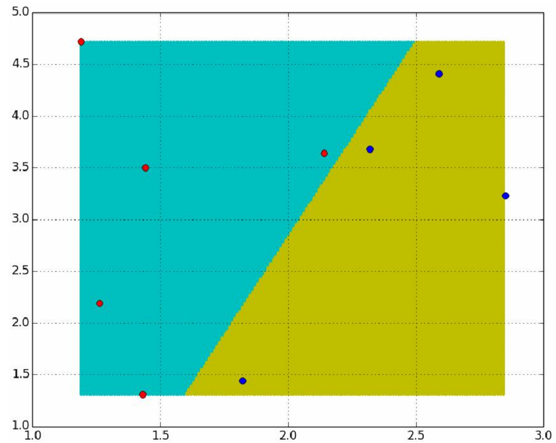
}

分类：Logistic回归

- 沿Logistic函数的负梯度
- 维度提升

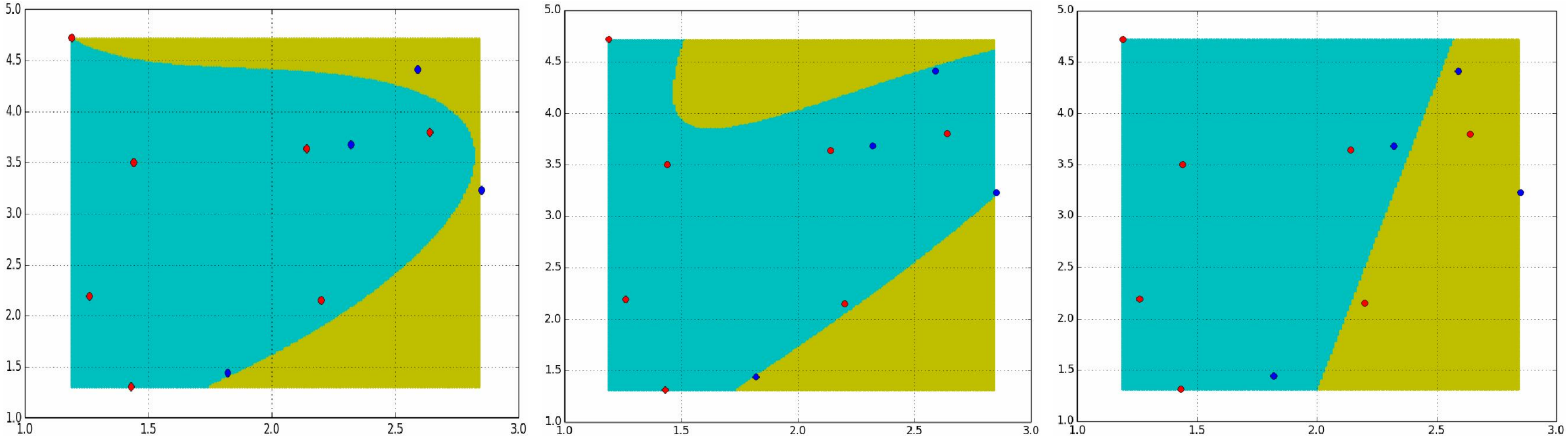


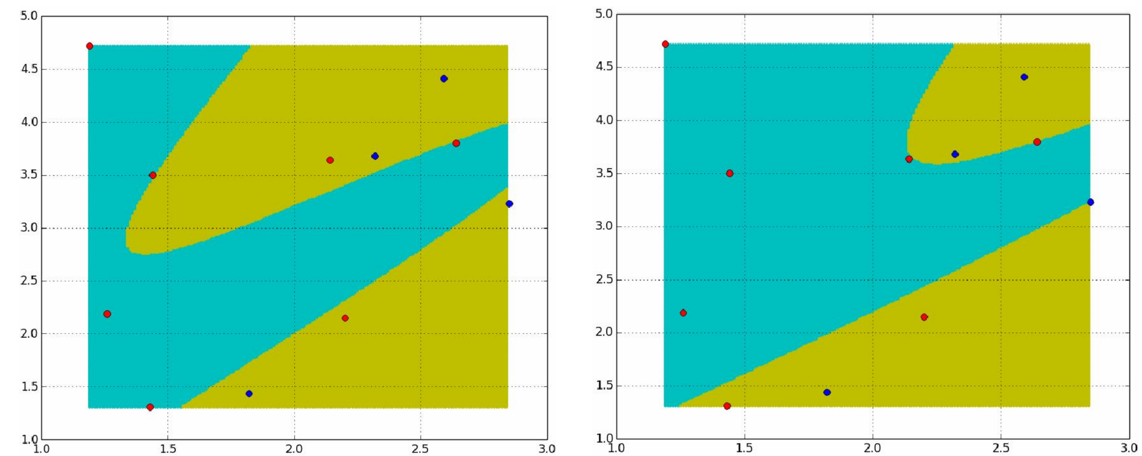

```
def logistic_regression(data, alpha, lamda):
    n = len(data[0]) - 1
    w = [0 for x in range(n)]
    w2 = [0 for x in range(n)]
    for times in range(10000):
        for d in data:
            for i in range(n):
                w2[i] = w[i] + alpha * (d[n] - h(w,d))*d[i] + lamda * w[i]
            for i in range(n):
                w[i] = w2[i]
            print w
    return w
```

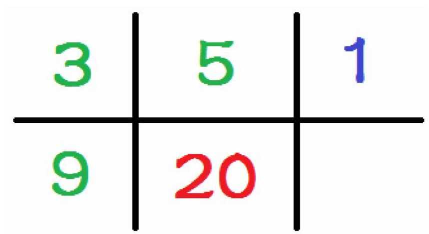



```
def feature_whole(x1, x2, e):
    d = [1]
    for n in range(1,e+1):
        for i in range(n+1):
            d.append(pow(x1, n-i) * pow(x2, i))
    return d
```

数据升维："选取特征"







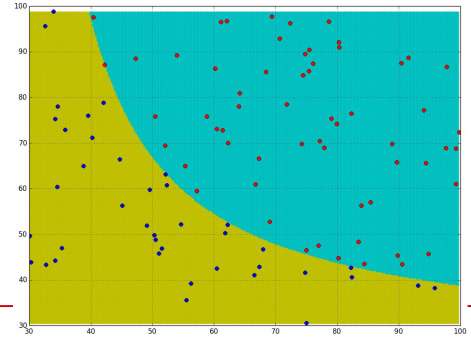
对数线性模型

- 一个事件的几率odds，是指该事件发生的概率与该事件不发生的概率的比值。
- 对数几率：logit函数

$$
\begin{aligned}
P(y=1 \mid x;\theta) &= h_\theta(x) \\
P(y=0 \mid x;\theta) &= 1-h_\theta(x)
\end{aligned}
$$

$$
\operatorname{logit}(p)=\log\frac{p}{1-p}=\log\frac{h_\theta(x)}{1-h_\theta(x)}=\log\left(\frac{\dfrac{1}{1+e^{-\theta^T x}}}{\dfrac{e^{-\theta^T x}}{1+e^{-\theta^T x}}}\right)=\theta^T x
$$

对数线性模型的一般形式



- 令x为某样本，y是x的可能标记，将Logistic/Softmax回归的特征$(x_1, x_2, \cdots, x_n)$记做Fj(x,y)，则对数线性模型的一般形式：

$$p(y \mid x; w) = \frac{1}{Z(x, w)} \exp\left(\sum_j w_j F_j(x, y) \right)$$

- 其中，归一化因子：

$$Z(x, w) = \sum_y \exp \sum_j w_j F_j(x, y)$$

- 给定x的预测标记为：

$$\hat{y} = \arg\max_y p(y \mid x, w) = \arg\max_y \sum_j w_j F_j(x, y)$$

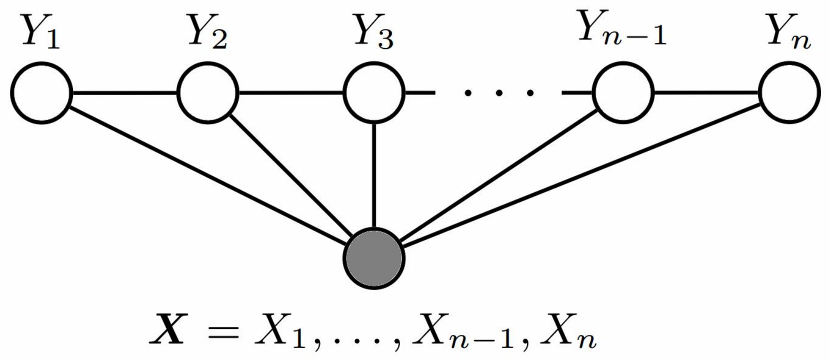
特征函数的选择：自然语言处理为例

- 特征函数几乎可任意选择，甚至特征函数间重叠；
- 假定观测x是单词，则下列都是合理可行的特征：
  - x以大写字母开头
  - x以J开头
  - x等于"JulyEdu"
  - x的长度为7
  - x中包含两个大写字母
- NLP中常用的特征：
  - 前缀、后缀、词典位置、前置/后置标点

词性标注的特征函数

- 词性标注是指将每个单词标记为名词/动词/形容词/介词等。
  - 词性：POS，Part Of Speech
- 记w为句子s的某个单词，则特征函数可以是：
  - w在句首/句尾(位置相关)
  - w的前缀是anti-/co-/dis-/inter-等(单词本身)
  - w的后缀是-able/-ation/-er/-ing等(单词本身)
  - w前面的单词是a/could/SALUTATION等(单词间)
  - w后面的单词是am/is/are/等(单词间)
  - w前面两个单词是Would like/There be等(单词和句子)
- 高精度的POS会使用超过10万个特征。
  - 每个特征只和当前词性有关，最多只和相邻词的词性有关；
  - 但特征可以和所有词有关。

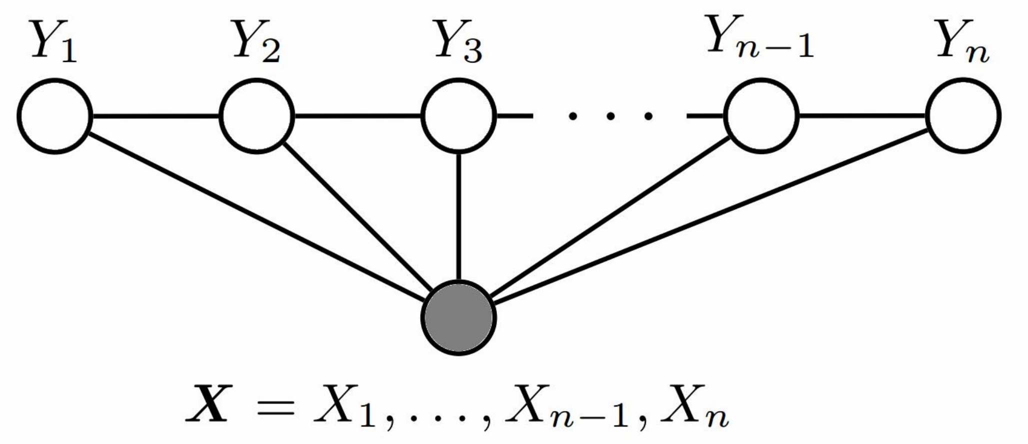
词性标注

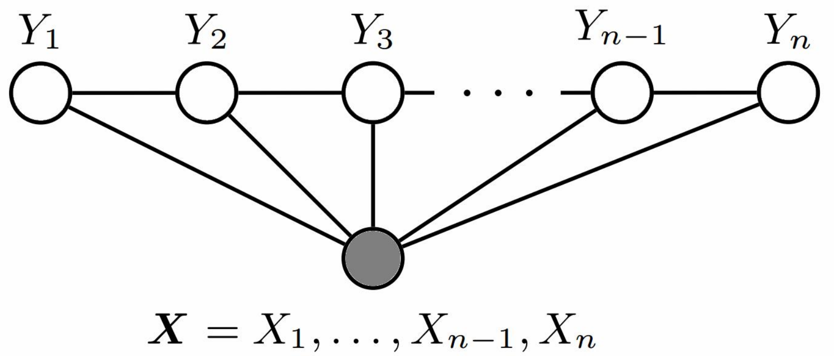


- 词性标注被称为“结构化预测”，该任务与标准的类别学习任务存在巨大不同：
- 如果每个单词分别预测，将丢失众多信息；
  - 相邻单词的标记是相互影响的，非独立。
- 不同的句子有不同长度；
  - 这导致不方便将所有句子统一成同长度向量。
- 标记序列解集与句子长度呈指数级增长。
  - 这使得穷举计算几乎无法实用。

线性链条件随机场

- 线性条件随机场可以使用对数线性模型。
- 使用$\overline{x}$表示n个词的序列；$\overline{y}$表示相应的词性
- 定义：$p(\overline{y} \mid \overline{x}; w) = \frac{1}{Z(\overline{x}, w)} \exp\left(\sum_j w_j F_j(\overline{x}, \overline{y}) \right)$





次特征

- 定义句子$\bar{x}$的第j个特征$F_j(\bar{x},\bar{y})$是由若干次特征$f_j(y_{i-1},y_i,\bar{x},i)$组合而成的，这里的fj依赖或部分依赖于当前整个句子$\bar{x}$、当前词的标记y_i、前一个词的标记y_{i-1}、当前词在句子中的位置i。

$$F_j(\bar{x},\bar{y})=\sum_i f_j(y_{i-1},y_i,\bar{x},i)$$

- 将每一个位置i上的次特征fj相加，即得到特征Fj，从而解决训练样本变长的问题。

参数训练

- 给定一组训练样本(x,y)，找出权向量w，使得下式成立：

$$\bar{y}^* = \arg\max_{\bar{y}} p(\bar{y} \mid \bar{x}, w)$$

- 满足上式的w，即为最终的推断参数。

参数推断的两个难点

- 如果给定x和w，如何计算哪个标记序列y的概率最大？ $\bar{y}^* = \arg\max_{\bar{y}} p(\bar{y} \mid \bar{x}, w)$

- 如果给定x和w， p(y|x,w)本身如何计算？
  - 归一化因子与所有的可行标记y有关，不好计算

$$p(\bar{y} \mid \bar{x}; w) = \frac{1}{Z(\bar{x}, w)} \exp\left(\sum_j w_j F_j(\bar{x}, \bar{y}) \right)$$

$$Z(\bar{x}, w) = \sum_{\bar{y}} \exp \sum_j w_j F_j(\bar{x}, \bar{y})$$

状态关系矩阵

- 根据 $p(\overline{y} \mid \overline{x}; w) = \frac{1}{Z(\overline{x}, w)} \exp\left(\sum_j w_j F_j(\overline{x}, \overline{y})\right)$
- 得

$$\overline{y}^* = \arg\max_{\overline{y}} p(\overline{y} \mid \overline{x}, w) = \arg\max_{\overline{y}} \sum_j w_j F_j(\overline{x}, \overline{y})$$

- 根据 $F_j(\overline{x}, \overline{y}) = \sum_i f_j(y_{i-1}, y_i, \overline{x}, i)$，得：

$$\overline{y}^* = \arg\max_{\overline{y}} \sum_j w_j \sum_i f_j(y_{i-1}, y_i, \overline{x}, i) = \arg\max_{\overline{y}} \sum_j \sum_i w_j f_j(y_{i-1}, y_i, \overline{x}, i)$$

$$= \arg\max_{\overline{y}} \sum_i \sum_j w_j f_j(y_{i-1}, y_i, \overline{x}, i)$$

$$= \arg\max_{\overline{y}} \sum_i g_j(y_{i-1}, y_i) \qquad g_j(y_{i-1}, y_i) = \sum_j w_j f_j(y_{i-1}, y_i, \overline{x}, i)$$

利用前向概率选择最大标记序列

- 称$\alpha_k(v)$为前向概率，表示第k个词的标记为v的最大得分值(该得分值归一化后即为概率)，即：

$$\alpha_k(v) = \max_{y_1, y_2, \cdots, y_{k-1}} \left(\sum_{i=1}^{k-1} g_i(y_{i-1}, y_i) + g_k(y_{k-1}, v) \right)$$

- 得递推公式：

$$\alpha_k(v) = \max_{y_{k-1}} \left(\alpha_{k-1}(y_{k-1}) + g_k(y_{k-1}, v) \right)$$

- 时间复杂度：$O(m^2n)$
  - 标记数目为m，句子包含的单词数目为n

计算概率 $p(\bar{y}\mid\bar{x};w)=\frac{1}{Z(\bar{x},w)}\exp\left(\sum_j w_j F_j(\bar{x},\bar{y})\right)$

- 如果给定x和w， p(y|x,w)本身如何计算？
  - 归一化因子与所有可行标记y有关，不容易计算
- 由于：$g_j(y_{i-1},y_i)=\sum_j w_j f_j(y_{i-1},y_i,\bar{x},i)$
- 得，

$$Z(\bar{x},w)=\sum_{\bar{y}}\exp\sum_j w_j F_j(\bar{x},\bar{y})$$

$$=\sum_{\bar{y}}\exp\sum_i g_j(y_{i-1},y_i)$$

$$=\sum_{\bar{y}}\prod_i \exp\left(g_j(y_{i-1},y_i)\right)$$

状态关系矩阵

- 定义m×m的矩阵 $M_t(u,v)=\exp(g_t(u,v))$
  - 对于$M_1(u,v)$，任选某u=start状态
  - 对于$M_{n+1}(u,v)$，任选某v=stop状态

$$
\begin{aligned}
&M_{12}(start,v) \\
&=\sum_q M_1(start,q)M_2(q,v) \\
&=\sum_q \exp(g_1(start,q))\cdot\exp(g_2(q,v))
\end{aligned}
$$

状态关系矩阵

- 矩阵连乘： $$M_{123}(start,v)=\sum_q M_{12}(start,q)M_3(q,v)$$
$$=\sum_q\left(\sum_r M_1(start,r)M_2(r,q)\right)M_3(q,v)$$
$$=\sum_{q,r} M_1(start,r)M_2(r,q)M_3(q,v)$$
- 从而，

$$M_{1,2,3\cdots,n+1}(start,stop)=\sum_{y_1,y_2,\cdots,y_n} M_1(start,y_1)M_2(y_1,y_2)\cdots M_{n+1}(y_n,stop)$$
$$=\sum_{\bar{y}}\prod_i \exp\left(g_j(y_{i-1},y_i)\right)$$

  - 时间复杂度O(m^3n)

参数训练

- 给定一组训练样本(x,y)，找出权向量w，使得下式成立：$\bar{y}^* = \arg\max_{\bar{y}} p(\bar{y} \mid \bar{x}, w)$
  - 满足上式的w，即为最终的推断参数。
- 方法：求对数目标函数的驻点。

$$p(y \mid x; w) = \frac{1}{Z(x,w)} \exp\left(\sum_j w_j F_j(x,y) \right)$$

$$\boxed{Z(x,w) = \sum_{\bar{y}} \exp \sum_j w_j F_j(x,y)}$$

$$\Rightarrow \log p(y \mid x; w) = \log \frac{1}{Z(x,w)} + \log \exp\left(\sum_j w_j F_j(x,y) \right)$$

$$= -\log Z(x,w) + \sum_j w_j F_j(x,y)$$

$$\log p(y \mid x; w) = -\log Z(x, w) + \sum_j w_j F_j(x, y)$$

参数估计

$$Z(x, w) = \sum_{\bar{y}} \exp \sum_j w_j F_j(x, y)$$

- 计算梯度：

$$\frac{\partial}{\partial w_j} \log p(y|x; w) = F_j(x, y) - \frac{\partial}{\partial w_j} \log Z(x, w)$$

$$= F_j(x, y) - \frac{1}{Z(x, w)} \sum_{y'} \frac{\partial}{\partial w_j} \exp \sum_{j'} w_{j'} F_{j'}(x, y')$$

$$= F_j(x, y) - \frac{1}{Z(x, w)} \sum_{y'} [\exp \sum_{j'} w_{j'} F_{j'}(x, y')] F_j(x, y')$$

$$= F_j(x, y) - \sum_{y'} F_j(x, y') \frac{\exp \sum_{j'} w_{j'} F_{j'}(x, y')}{\sum_{y''} \exp \sum_{j''} w_{j''} F_{j''}(x, y'')}$$

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

$$= F_j(x, y) - E_{y' \sim p(y'|x;w)}[F_j(x, y')]$$

- 梯度上升：

$$w_j := w_j + \alpha(F_j(x, y) - E_{y' \sim p(y'|x;w)}[F_j(x, y')])$$

若干理论问题

- 无向图模型
  - 马尔科夫随机场
- 团、最大团
- Hammersley-Clifford定理

无向图模型

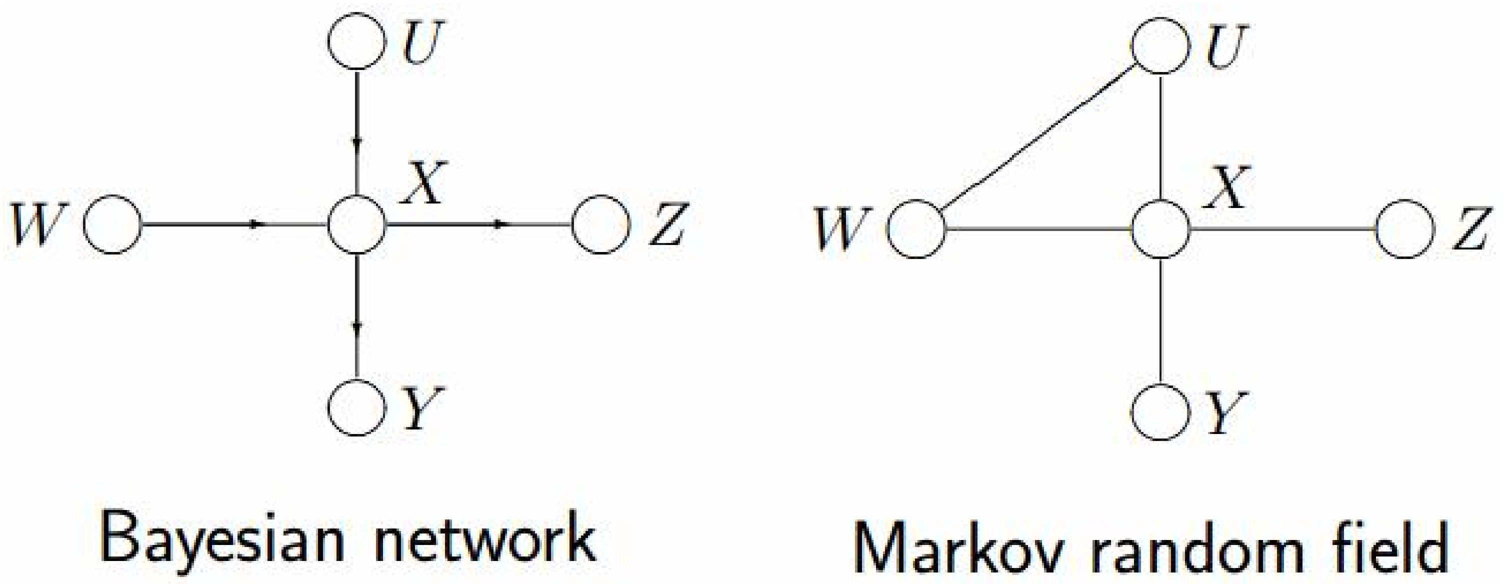
- 有向图模型，又称作贝叶斯网络(Directed Graphical Models, DGM, Bayesian Network)
  - 事实上，在有些情况下，强制对某些结点之间的边增加方向是不合适的。
- 使用没有方向的无向边，形成了无向图模型(Undirected Graphical Model,UGM), 又称马尔科夫随机场或马尔科夫网络(Markov Random Field, MRF or Markov network)

  - 注：概率有向图模型/概率无向图模型

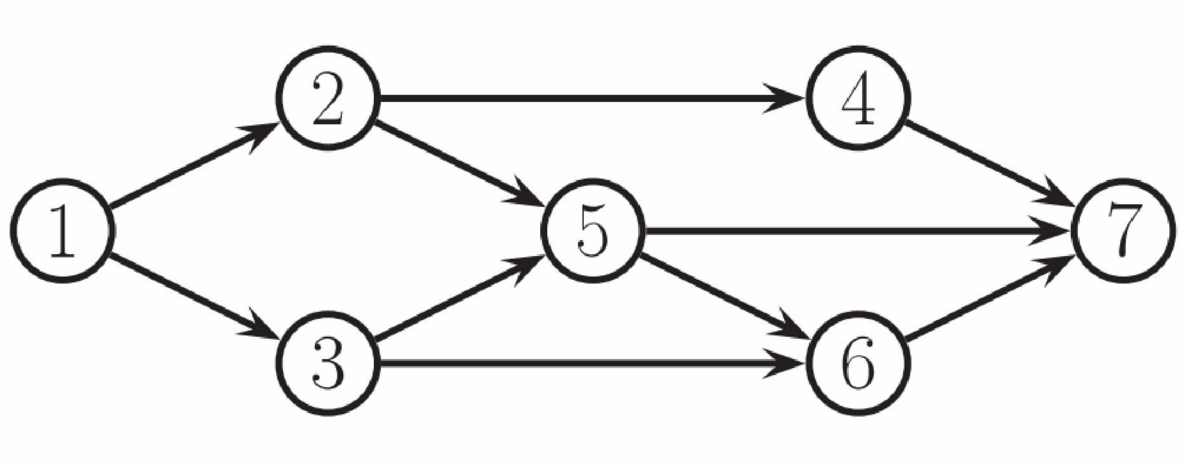
条件随机场

- 设$X=(X_1,X_2...X_n)$和$Y=(Y_1,Y_2...Y_m)$都是联合随机变量，若随机变量Y构成一个无向图G=(V,E)表示的马尔科夫随机场(MRF)，则条件概率分布P(Y|X)称为条件随机场(Conditional Random Field, CRF)
  - X称为输入变量、观测序列
  - Y称为输出序列、标记序列、状态序列
  - 大量文献将MRF和CRF混用，包括经典著作。
  - 一般而言，MRF是关于隐变量(状态变量、标记变量)的图模型，而给定观测变量后考察隐变量的条件概率，即为CRF。
  - 但这种混用，类似较真总理和周恩来的区别。
    - 混用的原因：在计算P(Y|X)时需要将X也纳入MRF中一起考虑

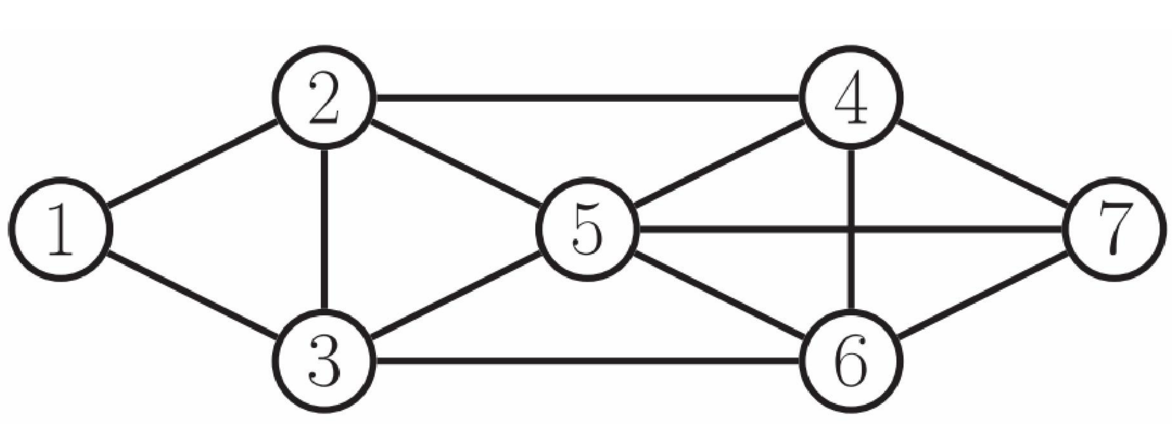
DGM转换成UGM



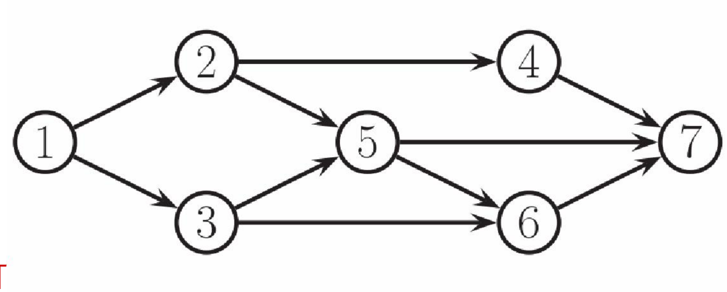
DGM转换成UGM



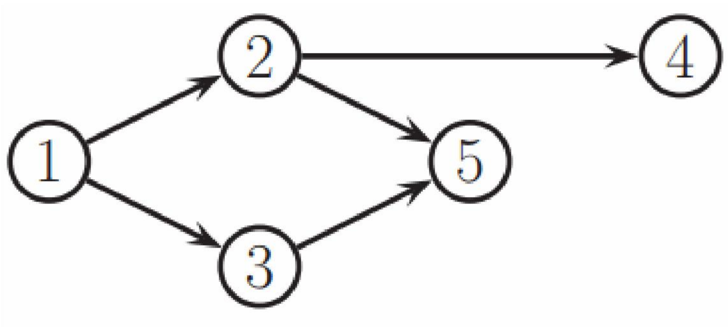
$$4 \perp 5 | 2$$

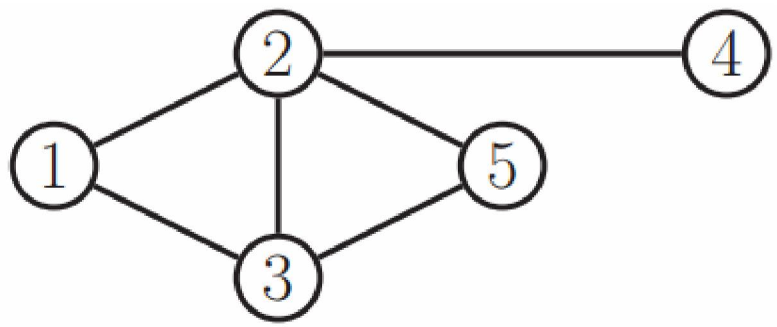


条件独立的破坏



- 靠考察是否有$A \perp B|C$，则计算U的祖先图(ancestral graph)：$U = A \cup B \cup C$

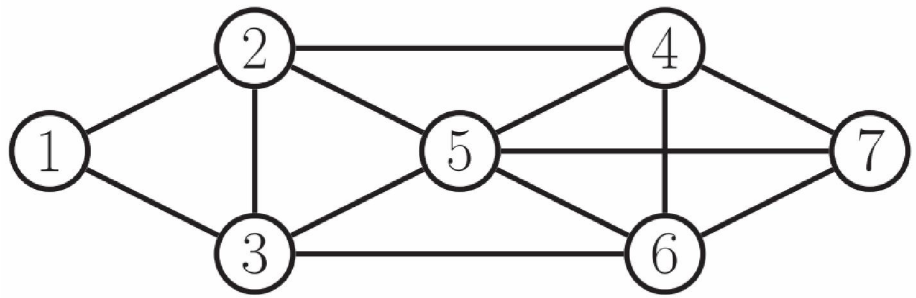




$$4 \perp 5 | 2$$

MRF的性质

- 成对马尔科夫性
  - parewise Markov property
- 局部马尔科夫性
  - local Markov property
- 全局马尔科夫性
  - global Markov property



Pairwise $1 \perp 7 | \text{rest}$

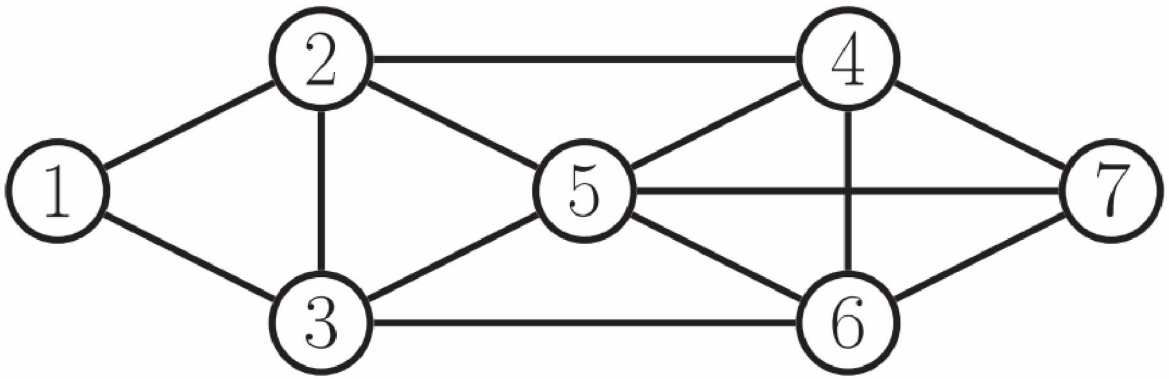
Local $1 \perp \text{rest} | 2, 3$

Global $1, 2 \perp 6, 7 | 3, 4, 5$

- 记号：随机变量$Y=(Y_1,Y_2 \ldots Y_m)$构成无向图$G=(V,E)$，结点(集)v对应的(联合)随机变量是Y_v。

成对马尔科夫性

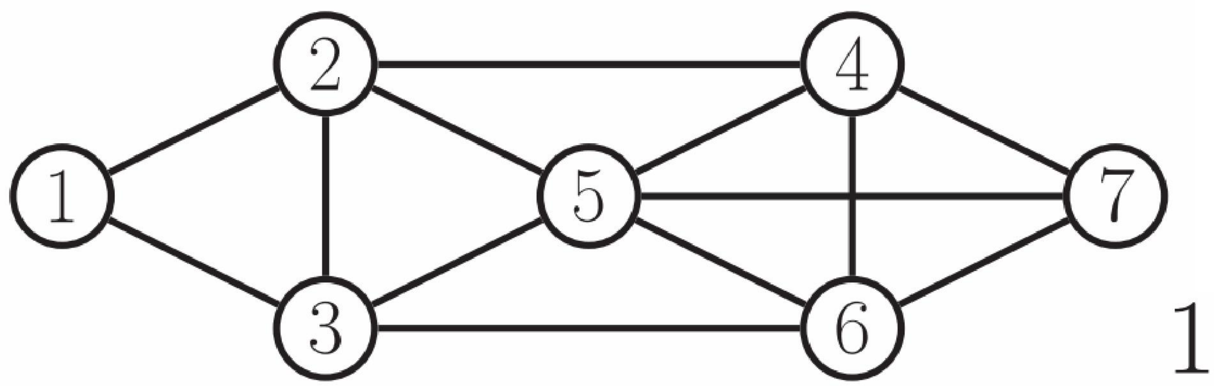
- 设u和v是无向图G中任意两个没有边直接连接的结点，G中其他结点的集合记做O；则在给定随机变量Yo的条件下，随机变量Yu和Yv条件独立。
- 即：P(Yu,Yv|Yo)= P(Yu|Yo)* P(Yv|Yo)



$$1 \perp 7 | \text{rest}$$

局部马尔科夫性

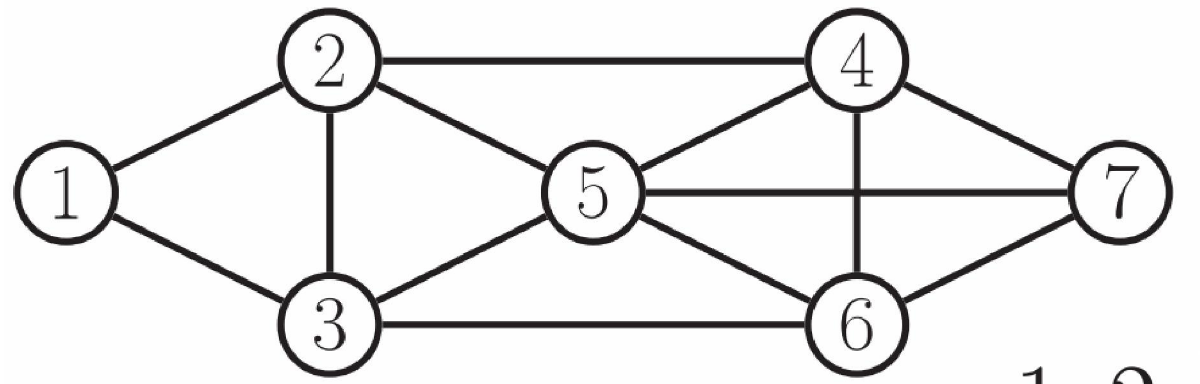
- 设v是无向图G中任意一个结点，W是与v有边相连的所有结点，G中其他结点记做O；则在给定随机变量Yw的条件下，随机变量Yv和Yo条件独立。
- 即：P(Yv,Yo|Yw)= P(Yv|Yw)* P(Yo|Yw)



$$1 \perp \text{rest} | 2, 3$$

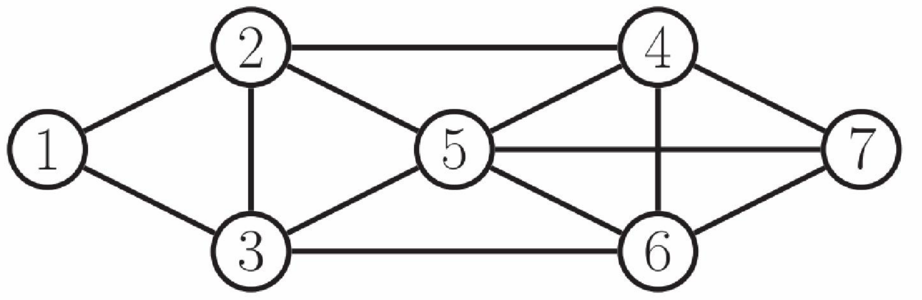
全局马尔科夫性

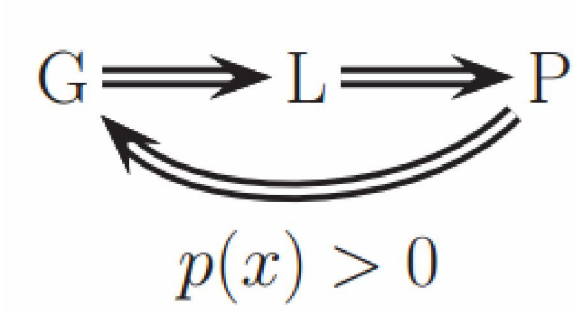
- 设结点集合A，B是在无向图G中被结点集合C分开的任意结点集合，则在给定随机变量Y_C的条件下，随机变量Y_A和Y_B条件独立。
- 即：$P(Y_A, Y_B | Y_C)= P(Y_A | Y_C)* P(Y_B | Y_C)$



$$1, 2 \perp 6, 7 | 3, 4, 5$$

三个性质的等价性



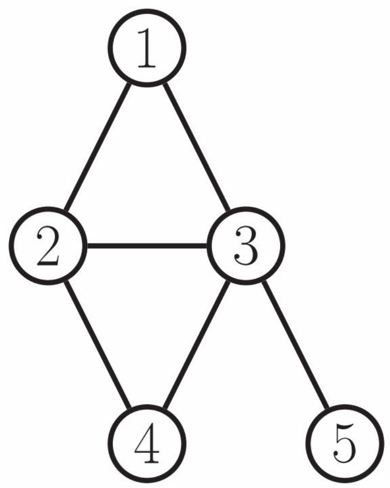


- 根据全局马尔科夫性，能够得到局部马尔科夫性；
- 根据局部马尔科夫性，能够得到成对马尔科夫性；
- 根据成对马尔科夫性，能够得到全局马尔科夫性；

- 事实上，这个性质对MRF具有**决定性**作用：
  - 满足这三个性质(或其一)的无向图，称为MRF。

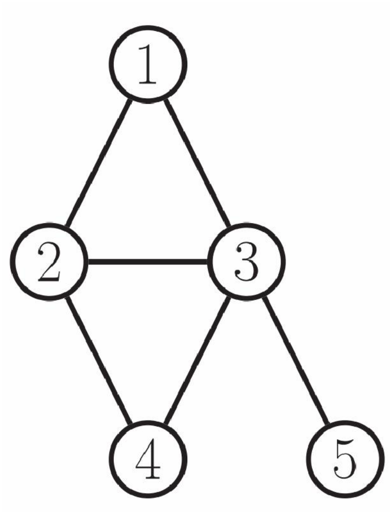
团和最大团

- 无向图G中的某个子图S，若S中任何两个结点均有边，则S称作G的团(Clique)。
  - 若C是G的一个团，并且不能再加入任何一个G的结点使其称为团，则C称作G的最大团(Maximal Clique)。
- 团：{1,2}, {1,3}, {2,3}, {2,4}, {3,4}, {3,5},{1,2,3}, {2,3,4}
- 最大团：{1,2,3}, {2,3,4}, {3,5}



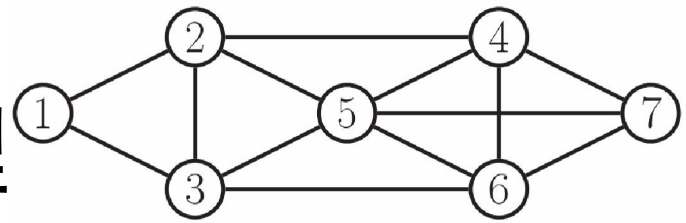
Hammersley-Clifford定理

- UGM的联合分布可以表示成最大团上的随机变量的函数的乘积的形式；这个操作叫做UGM的因子分解(Factorization)。



$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\psi_{123}(y_1, y_2, y_3)\psi_{234}(y_2, y_3, y_4)\psi_{35}(y_3, y_5)$$

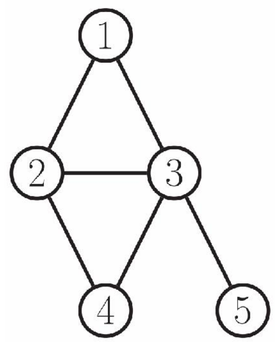
Hammersley-Clifford定理



- UGM的联合概率分布P(Y)可以表示成如下形式：

$$P(Y)=\frac{1}{Z}\prod_{c}\psi_c(Y_c)$$

$$Z=\sum_{Y}\prod_{c}\psi_c(Y_c)$$



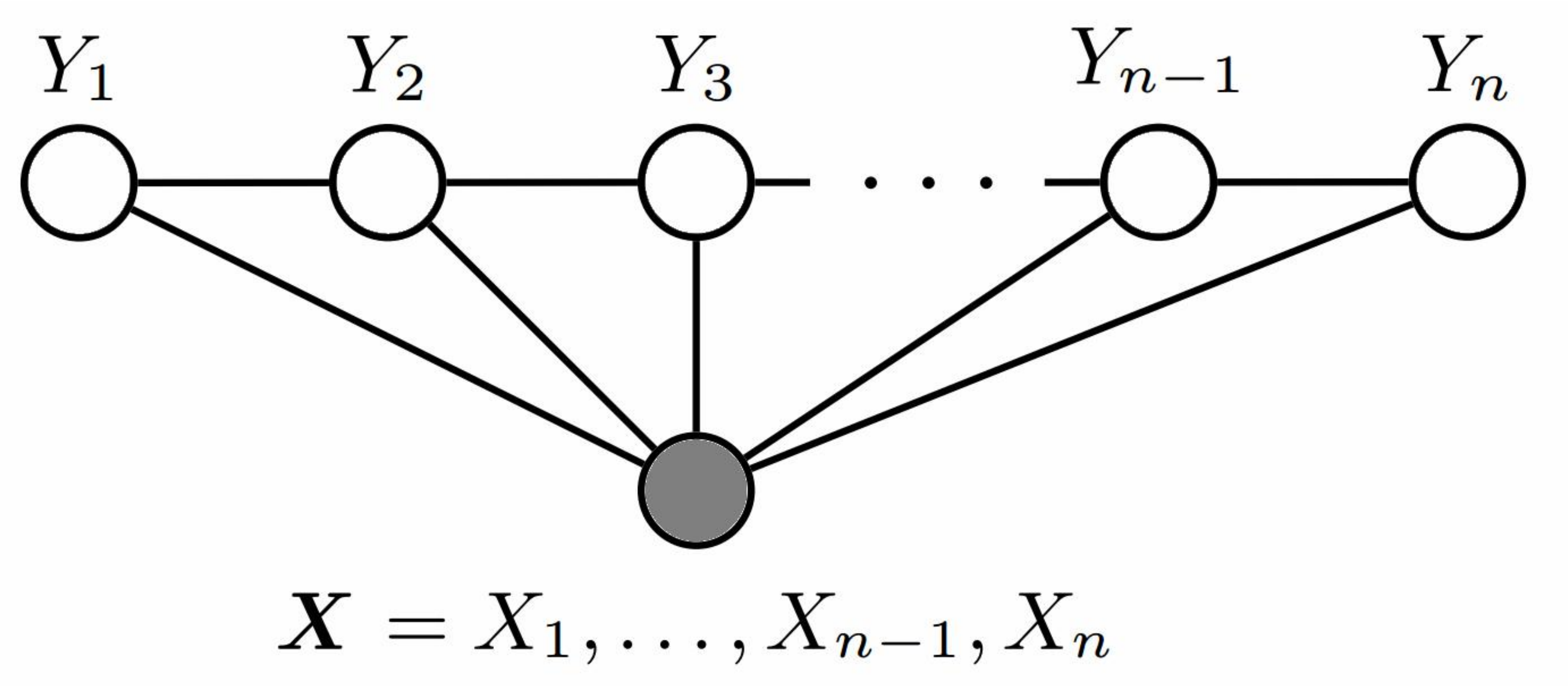
- 其中，C是G的最大团，$\psi_c(Y_c)$是C上定义的严格正函数，被称作势函数(Potential Function)。因子分解是在UGM所有的最大团上进行的。

线性链条件随机场

- 设$X=(X_1,X_2...X_n)$和$Y=(Y_1,Y_2...Y_m)$都是联合随机变量，若随机变量Y构成一个无向图G=(V,E)表示的马尔科夫随机场(MRF)，则条件概率分布P(Y|X)称为条件随机场(Conditional Random Field, CRF)
- 即：$P(Y_v|X,Y_w,w \neq v) = P(Y_v|X,Y_w,w \cong v)$
- 其中，$w \cong v$表示与结点v相连的所有结点w
- 一种重要而特殊的CRF是线性链条件随机场(Linear Chain Conditional Random Field)，可用于标注等问题。这时，条件概率P(Y|X)中，Y表示标记序列(或称状态序列)，X是需要标注的观测序列。

线性链条件随机场

- 线性链条件随机场的无向图模型



线性链条件随机场的定义

- 设$X=(X_1,X_2\ldots X_n)$和$Y=(Y_1,Y_2\ldots Y_n)$均为线性链表示的随机变量序列，若在给定随机变量序列X的条件下，随机变量序列Y的条件概率分布P(Y|X)构成条件随机场，即满足马尔科夫性

$$P\left(Y_i \middle| X, Y_1, Y_2 \cdots Y_n\right) = P\left(Y_i \middle| X, Y_{i-1}, Y_{i+1}\right)$$

- 则称P(Y|X)为**线性链条件随机场**。在标注问题中，X表示输入序列或称观测序列，Y表述对应的输出标记序列或称状态序列。

线性链条件随机场的参数化形式

- 设P(Y|X)为线性链条件随机场，则在随机变量X取值为x的条件下，随机变量Y取值为y的条件概率有以下形式：

$$P(y|x)=\frac{1}{Z(x)}\exp\left(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\right)$$

- 其中，$Z(x)=\sum_y \exp\left(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\right)$
  - 特征函数：$t_k(y_{i-1},y_i,x,i)$、$s_l(y_i,x,i)$
  - 特征函数对应的权值：λ_k、μ_l
  - Z(x)为规范化因子，保证P(Y|X)为概率分布。

参数说明

- t_k是定义在边上的特征函数，称为转移特征，依赖于当前和前一个位置；
- s_l是定义在结点上的特征函数，称为状态特征，依赖于当前位置；
- t_k 和s_l都依赖于位置，是局部特征函数；
- 通常，t_k 和s_l取值为1或者0；满足特征条件时取1，否则取0；
- CRF由特征函数t_k、s_l和对应的权值λ_k μ_l确定。
- 线性链条件随机场模型属于对数线性模型。

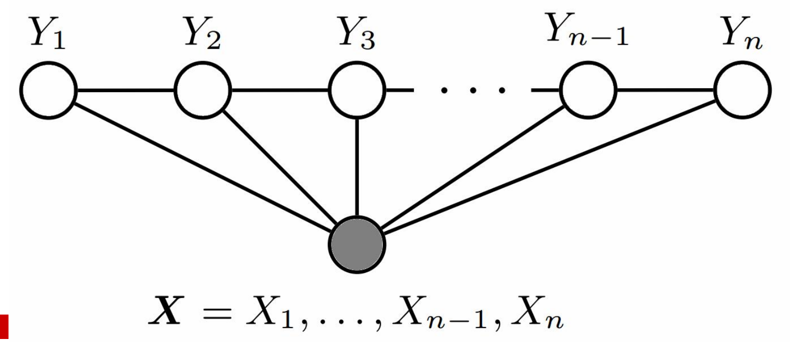
条件随机场举例

[PRP He] [VBZ reckons] [DT the] [JJ current] [NN account] [NN deficit] [MD will] [VB narrow] [TO to] [RB only] [# #] [CD 1.8] [CD billion] [IN in] [NNP September] [. .]

- NN、NNS、NNP、NNPS、PRP、DT、JJ分别代表普通名词单数形式、普通名词复数形式、专有名词单数形式、专有名词复数形式、代词、限定词、形容词

$$b(\boldsymbol{x}, i) = \begin{cases} 1 & \text{if the observation at position } i \text{ is the word “September”} \\ 0 & \text{otherwise.} \end{cases}$$

$$t_j(y_{i-1}, y_i, \boldsymbol{x}, i) = \begin{cases} b(\boldsymbol{x}, i) & \text{if } y_{i-1} = \text{IN and } y_i = \text{NNP} \\ 0 & \text{otherwise.} \end{cases}$$



CRF总结

- 条件随机场可以使用对数线性模型表达。不严格的说，线性链条件随机场可看成是隐马尔科夫模型的推广，隐马尔科夫模型可看成是线性链条件随机场的特殊情况。
  - 概率计算使用前向-后向算法；
  - 参数学习使用梯度上升算法(或IIS)；
  - 应用于标注/分类，在给定参数和观测序列(样本)的前提下，使用Viterbi算法进行标记的预测。
- 标记序列y要求链状，但x无要求，除了一维的词性标注、中文分词，还可以用于离散数据(如用户信息)，或二维数据(如图像)，用途广泛。
- 缺点：有监督学习计算参数、参数估计的速度慢。

参考文献

- https://en.wikipedia.org/wiki/Conditional_random_field
- John Lafferty, Andrew McCallum, Fernando C.N. Pereira, 2001. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data".
- Charles Elkan, 2008. "Log-linear models and conditional random fields".
- Hanna M. Wallach, 2004. "Conditional Random Fields: An Introduction“.
- Charles Sutton, Andrew McCallum, 2012. "An Introduction to Conditional Random Fields".
- Kevin P. Murphy, 2012. "Machine Learning:A Probabilistic Perspective", The MIT Press.
- Bishop M, 2006. "Pattern Recognition and Machine Learning", Chapter 13, Springer-Verlag

我们在这里

7 七月算法 http://www.julyedu.com/

- 视频/课程/社区

□ 七月题库APP：Android/iOS

- http://www.julyapp.com/

□ 微博

- @研究者July
- @七月题库
- @邹博_机器学习

□ 微信公众号

- julyedu

感谢大家！

恳请大家批评指正！