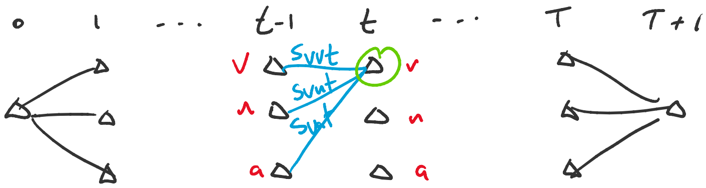
Section A: HMMs + BayesNets

Wednesday, November 9, 2022 10:02 AM

$\boxed{\text{Viterbi Algo (edge weights)}}$



for $t=1,\ldots,T+1$:

for $k=1,\ldots,K$:

$$w_t(k) = \max_{j\in\{1,\ldots,K\}} w_{t-1}(j)\, s_{kjt}$$

$$b_t(k) = \operatorname{argmax}_{j\in\{1,\ldots,K\}} w_{t-1}(j)\, s_{kjt}$$

backpointers

$$w_t(v) = \max\big(w_{t-1}(v)\, s_{vvt},\ w_{t-1}(n)\, s_{vnt},\ w_{t-1}(a)\, s_{vat} \big)$$

For HMM:

$$s_{kjt} = p(y_t = k \mid y_{t-1} = j)\; p(x_t \mid y_t = k)$$

Ex: Senate Race in PA

$F=1 \Rightarrow$ Fetterman beats Oz

$B=1 \Rightarrow$ politics is Boring

$C=1 \Rightarrow$ voters are fiscally Conservative

$L=1 \Rightarrow$ voters are fiscally Liberal

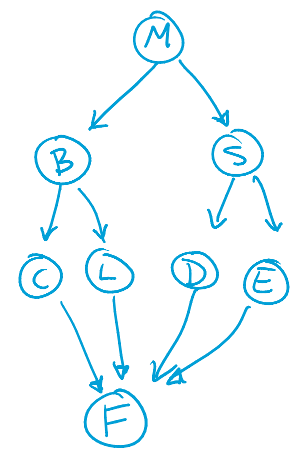
$S=1 \Rightarrow$ politics is Sensationalized

$D=1 \Rightarrow$ voters are concerned about Fetterman's disability

$E=1 \Rightarrow$ voters are concerned about Oz's Exploitation of those w/ disabilities

$M=1 =$ Millionaires are involved

Bayes Net



$$p(F,B,\ldots,E) = p(F\mid C,L,D,E)\; p(C\mid B)\; p(L\mid B)\; p(D\mid S)\; p(E\mid S)\; p(B\mid M)\; p(S\mid M)\; p(M)$$

Q: How to represent $P(F,B,C,L,S,D,E)$?

Idea #1: Chain Rule

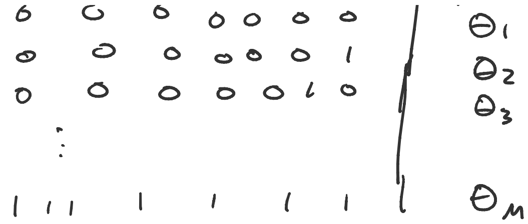
$$P(F,B,\ldots,E) = P(F\mid B,C,\ldots,E)\, P(B\mid C,L,\ldots,E) \cdots P(D\mid E)\, P(E)$$

Cons: not very compact

Idea #2: Full Joint Table

F	B	C	L	S	D	E	$p(\cdot)$
0	0	0	0	0	0	0	θ_1
0	0	0	0	0	0	1	θ_2
0	0	0	0	0	1	0	

cons: not very



cons: not very compact

<u>Idea #3:</u> Complete Independence

$$P(F, B, \dots, E) = P(F)\,P(B) \cdots P(E)$$

pros: compact
cons: not very expressive

<u>Idea #4:</u> Naive Bayes

$$P(F, B, \dots, E) = P(B|F) \cdots P(E|F)\, P(F)$$

"

<u>Idea #5:</u> Bayesian Network

Proof of Cond. Indep.

<u>Case #1:</u> Cascade

$$p(x, z | y) = \frac{p(x, y, z)}{p(y)}$$

$$= \frac{p(x|y)\,\overbrace{p(y|z)\,p(z)}}{p(y)}$$

$$= \frac{p(x|y)\,p(z|y)\,\cancel{p(y)}}{\cancel{p(y)}}$$

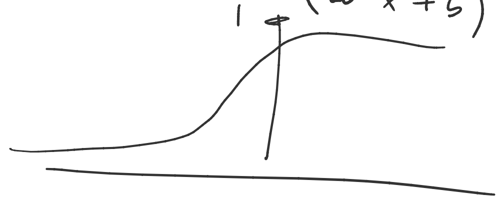
$$= p(x|y)\,p(z|y)$$

$$\Rightarrow \quad X \perp\!\!\!\perp Z \mid Y$$

Case #2: Common Parent
Case #3: V-structure
$\Big\}$ left as exercises

$$w_m \leftarrow w_m - \gamma\left(y^{(i)} - p(y=1 \mid \vec{x}^{(i)})\right) x^{(i)}$$

$$w_m \leftarrow w_m - \delta\left(y^{(i)} - \underbrace{p_{w,b}(y=1\mid \vec{x}^{(i)})}_{\sigma(\omega^T x + b)}\right) x_m^{(i)}$$



likelihood:

prob of data given params

$$p(D\mid\theta) = \prod_{i=1}^{N} p(x^{(i)}\mid\theta)$$

$$L_D(\theta)$$

$$L(\theta, x_1, \dots, x_m)$$

$$L(\theta \mid x_1, \dots, x_n)$$

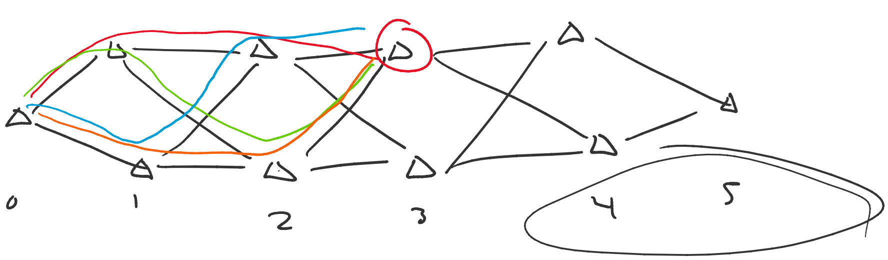
conditional likelihood:

$$\prod_{i=1}^{N} p(y^{(i)}\mid x^{(i)}, \theta) = p(Y \mid X, \theta)$$

$$L_{y,x}(\theta)$$

$$p(y_2 = n \mid x_1, x_2, x_3) = \sum_{y_1} p(y_1, y_2 = n \mid x_1, x_2, x_3)$$

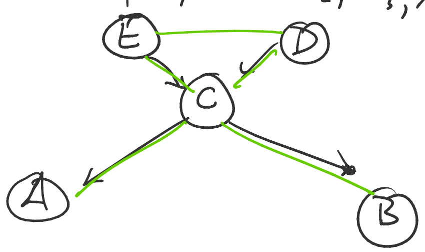
$$= \sum_{y_1} \sum_{y_3} p(y_1, y_2 = n, y_3 \mid x_1, x_2, x_3)$$

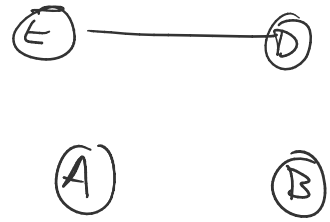


$$\underline{p(y_3 \mid x_1, x_2, x_3)} = \sum_{y_1}\sum_{y_2} p(y_1, y_2, y_3 \mid x_1, x_2, x_3)$$

$$= \sum_{y_1}\sum_{y_2}\sum_{x_4}\sum_{x_5} p(y_1, y_2, y_3, x_4, x_5 \mid x_1, x_2, x_3)$$

$$\neq p(y_3 \mid x_1, x_2, x_3, x_4, x_5)$$

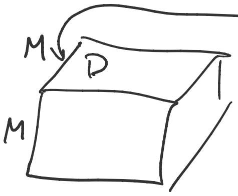




$$A \perp\!\!\!\perp B \mid C$$

$$p(A,B,C,D,E) = p(A \mid C)\, p(B \mid C)\, p(C \mid D,E)\, p(D)\, p(E)$$

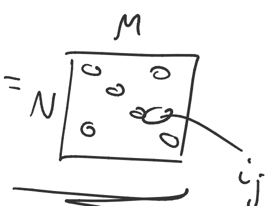
$$\frac{\partial h}{\partial W} \in$$



$h \in \mathbb{R}^{M}$

$W \in \mathbb{R}^{M \times D}$

$$\frac{\partial \vec{h}}{\partial \vec{x}} =$$



$x \in \mathbb{R}^N$

$\vec{h} = \text{sigmoid}(\vec{x}) = \sigma(\vec{x})$

$$\frac{\partial h_i}{\partial x_j}$$

$$\underline{\frac{\partial h_i}{\partial x_i} = h_i(1 - h_i)}$$