10-601B Bayes Nets

Structured Prediction

$$\hat{y} = \underset{y_1, y_2, \ldots, y_J}{\text{argmax}} \; p_\theta(y_1, y_2, \ldots, y_J \mid \vec{x})$$

$\forall j \;\; y_j \in \{0, 1\}$

Joint Distributions

Giant Table

y_1	y_2	$\cdots$	y_J	$p_\theta(\cdot \mid \cdot)$
0	0	$\cdots$	0	π_1
1	0	$\cdots$	0	π_2
0	1	$\cdots$	0	π_3

rows: 2^J
cols: $J+1$

Alt #1 Chain Rule:

$$\begin{aligned} p(y_1, y_2, \ldots, y_J) = \; & p(y_1 \mid y_2, \ldots, y_J) & \leftarrow 2^{J-1} \\ & p(y_2 \mid y_3, \ldots, y_J) & \vdots \\ & \vdots & \leftarrow 2^3 \\ & p(y_{J-1} \mid y_J) & \leftarrow 2^2 \\ & p(y_J) & \leftarrow 2 \end{aligned}$$

(# rows)

Gaussian

$y_j \in \mathbb{R} \qquad \vec{y} \sim \mathcal{N}(\vec{\mu}, \Sigma)$

Alt #2
Indep.

$$p(y_1,\dots,y_J) = \prod_{j=1}^{J} p(y_j)$$ #row: 2

Alt #3
In-Btwn: NB

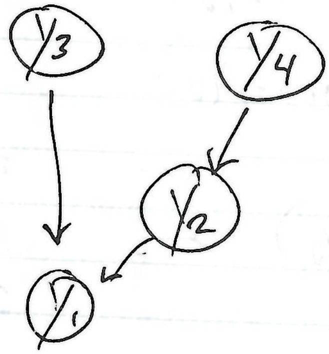
$$p(y_1,\dots,y_J) = p(y_1) \prod_{j=1}^{J} p(y_j \mid y_1)$$

#r: 2 #r: 4

Alt #4
In-Btwn

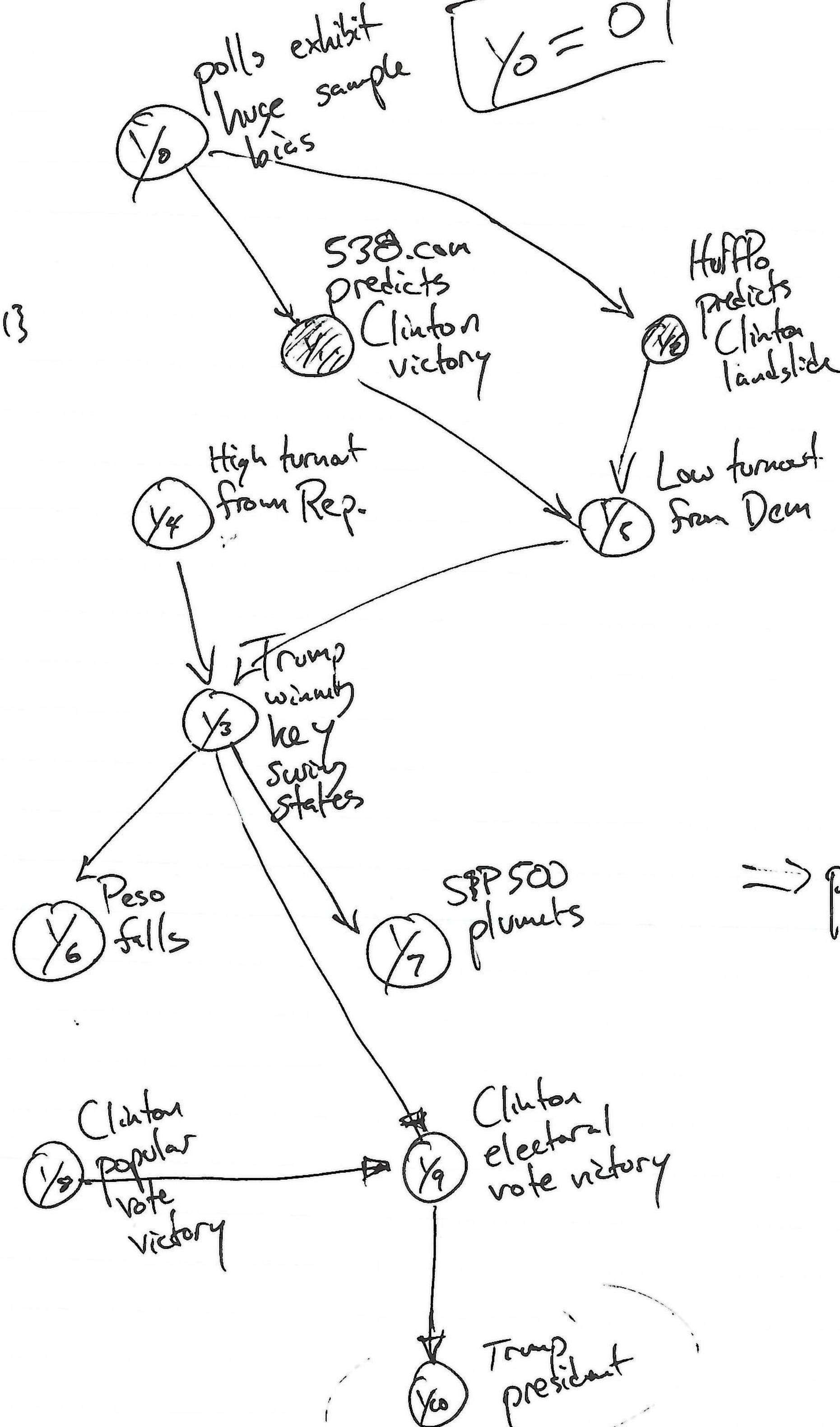
$$\cancel{p(y_1,\dots,y_J) = p(y_1 \mid y_2,\dots,y_J)}$$

$$p(y_1,y_2,y_3,y_4) = p(y_1 \mid y_2, y_3, \cancel{y_4}) \quad 8$$
$$p(y_2 \mid \cancel{y_3}, y_4) \quad 4$$
$$p(y_3 \mid \cancel{y_4}) \quad 2$$
$$p(y_4) \quad 2$$



$\boxed{y_0 = 0}$

* $y_j \in \{0,1\}$



$\Longrightarrow p(y_1, \ldots, y_{10}) =$

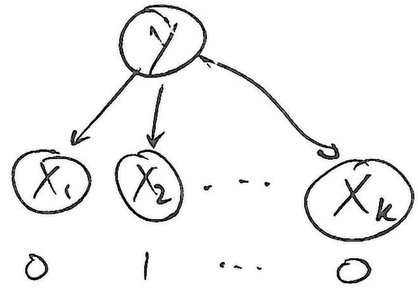
$p(y_0)$
$p(y_1 | y_0)$
$p(y_2 | y_0)$
$p(y_5 | y_1, y_2)$
$\vdots$
$p(y_{10} | y_9)$

$$p(y_{10} = 1 \mid y_1 = 1, y_2 = 1)$$

$$\propto \sum_{y_0, y_3, \ldots, y_9} p(y_1, \ldots, y_{10})$$

10-601B, Nov. 14

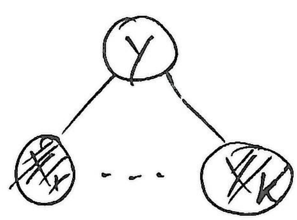
NB as a BN



1. Clustering, given y predict x ← 1-hot vector
2. Regression, given x predict y ← continuous
3. Classification, given x predict y ← binary

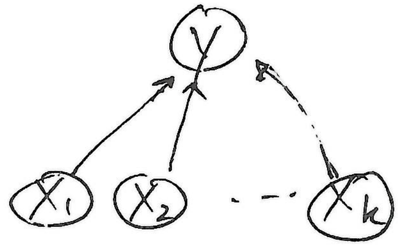
$$p(x_1, \ldots, x_k, y) = p(y)\, p(x_1|y) \cdots p(x_k|y)$$
$$= \underbrace{p(y) \prod_{k=1}^{K} p(x_k|y)}_{\text{NB Assumption}}$$

☆ Shaded nodes are observed



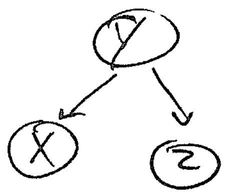
$$p(y|\vec{x}) \propto p(\vec{x}, y)$$

LR. as a BN



$$p(y|\vec{x})$$

Common Parent



Claim: $X \perp\!\!\!\perp Z \mid Y$

$$\frac{p(Y,X,Z)}{p(Y)} = \frac{\cancel{p(Y)}}{\cancel{p(Y)}}\, p(X|Y)\, p(Z|Y)$$

def: $p(Z|Y)\,p(X|Y) = p(X,Z|Y)$

	Classification	Structured Pred.
Disc.	$p(y \mid \vec{x})$	$p(\vec{y} \mid \vec{x})$
Gen.	$p(\vec{x}, y)$	$p(\vec{x}, \vec{y})$